ISBN: 978-1-7329562-1-6

Estd. in 2001

Your Platform to Build a Bright Future

ISO 9001:2015 Certified Institute	NBA Accredited Programs	NAAC Accredited Institute with 'A' Grade	AICTE-CII Survey rating in Platinum category for Industry linkages	78th All India Rank published in Outlook survey on 24th May 2018

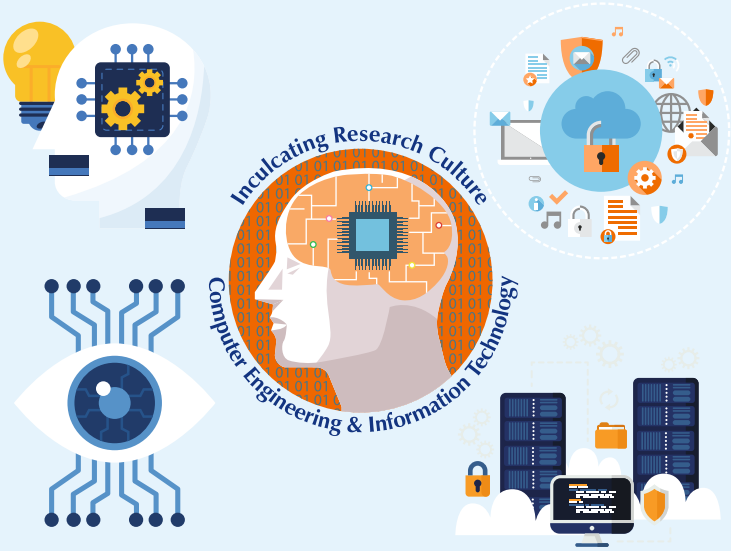
2019 Invention & Innovation

Proceedings of

International Conference on Intelligent Systems and Communication Networks - 2019

IC-ISCN - 2019

Organized by: **Department of Computer Engineering & Information Technology**



Editors

Dr. Rajesh Bansode

Dr. Sheetal Rathi

Mr. Vikas Kaul

Mrs. Shiwani Gupta

Mr. Anand Khandare

Mrs. Pranjali Kasture

Mr. Aditya Desai

Mrs. Vaishali Nirgude

In Association With

IOSR Journals
International Organization of Scientific Research
UGC Approved

Lagdu Singh Charitable Trust's (Regd.)

THAKUR COLLEGE OF ENGINEERING & TECHNOLOGY

(Approved by AICTE, Govt. of Maharashtra & Affiliated to University of Mumbai*)

Institute Accredited by National Assessment and Accreditation Council (NAAC), Bangalore#

• ISO 9001 : 2015 Certified • Accredited Programmes by National Board of Accreditation, New Delhi**

*Permanent Affiliated UG Programmes : • Computer Engineering • Electronics & Telecommunication Engineering • Information Technology (w.e.f. A.Y. 2015-16) • Electronics Engineering (w.e.f. A.Y. 2017-18)

**1st time NBA Accredited UG Programmes : • Computer Engineering • Electronics & Telecommunication Engineering • Information Technology (3 years w.e.f. 16-09-2011)

**2nd time NBA Accredited UG Programmes : • Computer Engineering • Electronics & Telecommunication Engineering • Information Technology • Electronics Engineering (3 years w.e.f. 01-07-2016)

1st cycle of NAAC Accreditation : • "A" Grade for 5 years (w.e.f. 30-10-2017)

Estd. in 2001

ISO 9001:2015 Certified Institute | NBA Accredited Programs | NAAC Accredited Institute with 'A' Grade | AICTE-CII Survey rating in Platinum category for Industry linkages | 78th All India Rank published in Outlook survey on 24th May 2018

IC-ISCN – 2019 Conference Proceedings

MCW 2019 Invention & Innovation

MULTICON-W 2019

- A platform for Multiple Conferences and Workshops

10th International & National Conferences & Workshops

In Association with

IOSR Journals
International Organization of Scientific Research









"International Conference on Intelligent Systems and Communication Networks (IC-ISCN 2019)"

Chief Patron
Mr. V. K. Singh, Chairman, TEG

Patron
Mr. Jitendra R. Singh, Trustee, TEG
Mr. Karan V. Singh, CEO, TCET, TIMSR, TIMSCDR

Program Chair
Dr. B. K. Mishra, Principal, TCET

Technical Program Committee
Dr. Anupam Shukla, Director, IIIT, Pune
Dr. Amol Ambardekar, Senior scientist, Microsoft, USA
Dr. Amitava Mukherjee, Sr. Manager (Retired), IBM Pvt. Ltd., Kolkata
Dr. Pablo Geovanny, Professor, University of Chile, Chile
Dr. D. Datta, HOD Radiology Department, BARC
Dr. B.P. Patil, Professor & Principal AIT, Pune
Dr. Vivek Deshpande, VIT, Pune
Dr. S.M. Gulhane, Professor Jawaharlal Darda College, Yavatmal
Mr. Sanjay Kumar, Joint Director, STPI, Noida

Program Co- Chair (IC-ISCN 2019)
Dr. Deven Shah, Vice-Principal
Dr. R.R. Sedamkar, Dean Academic
Dr. Kamal Shah, Dean R&D

Convenor
Dr. Rajesh Bansode, Professor, HOD-IT

Joint Convenor
Dr. Sheetal Rathi, Associate Professor, HOD-CMPN

Co-convenors
Mr. Vikas Kaul, Deputy HOD-IT
Mrs. Shiwani Gupta, Deputy HOD-CMPN

Treasurer
Mr. Vikas Singh, Assistant Professor, CMPN Dept.
Mrs. Mary Margaret Valentine, Assistant Professor, IT Dept.
Mr. Shailesh Sangle, Assistant Professor, CMPN Dept.

Publication Committee
Dr. Sangeeta Vhatkar, Assistant Professor, IT Dept.
Mr. Anand Khadare, Assistant Professor, CMPN Dept.
Mrs. Vaishali Nirgude, Assistant Professor, CMPN Dept.
Mr. Aaditya Desai, Assistant Professor, IT Dept.

PREFACE

It is a great honour for us to write a message for the INTERNATIONAL CONFERENCE ON INTELLIGENT STSTEMS AND COMMUNICATION NETWORKS (IC-ISCN19) scheduled during Feb.22-23, 2019.

This is the tenth event in the series of international conferences organized by TCET under the umbrella of MULTICON since the first event ICWET 2010. This event is organized in view of strengthening the research culture among its stakeholders namely students, faculty and Industry. It aims to help students graduating from TCET recognition as successful scientists, professionals, and business leaders. Moreover, need of the time is technological development in the areas of Intelligent Systems and Communication Engineering which can simplify our life in the eco-friendly environment with better connectivity and security. The 10th Annual International Conference, IC-ISCN 19 has affiliation with Scopus Indexed journal for Intelligent systems, leading publication house Tata McGraw Hill, UGC recognised journal IOSR and Conference proceeding with ISBN number. It serves as a premier platform that gathers all academicians, researchers and professionals, in the relevant engineering disciplines and domains, to showcase their research contributions and share their thoughts. IC-ISCN 19 has gained wide publicity through website, social media coverage, and visits to various colleges by the team of faculty members and our well-wishers. As a result, this event has got over-whelming response.

TCET has strong belief in quality and relation building. Therefore, a lot of care has been taken for branding the event, identifying the resources, logistic support required for the event, compilation and printing of conference proceeding, souvenir etc. Effort has been taken to make the delegates feel at home though away from home with the confidence of doing something for the development of nation.

TCET has successfully held nine international events in a row. ICWET 2010, 11, 12 and MulticonW 2013, 14, 15,16,17 and 18 in association with International journals of repute and multinationals in the IT industry with an objective to reach a new horizon. We feel that the present endeavour could not have been possible without the moral support, strong belief and time to time motivation from management of Thakur Educational Group.

During the two days there will be about 125 presentations from national as well as International Researchers and Industrial personnel, Idea presentations with deliberation by the delegates and the resource persons. Our best wishes and good luck to all.

We thank to all the members of the organizing and editorial committee for supporting the event and extending their cooperation to make it a grand successful event.

Dr. Rajesh Bansode
Convenor: IC-ISCN 2019
Dr. Sheetal Rathi
Joint Convenor: IC-ISCN 2019

CONTENTS

Preface *v*

INTELLIGENT SYSTEMS

1 Intelligent Tourist System 1-6

Prof. Bharati Ainapure, Anuj Mahajan, Shivam Pawargi, Nidhi Sadany

2 A Technological Solution For Cleaning And Maintenance Of Sewerage System 7-10

Sunil Yadav, Divya Tiwari, Kantilal Vishwakarma

3 Big Data Analysis for Associated Cloud Service using Parallel Trust Computing: A Survey. 11-14

Harshita Bhorshetti1 Shreyas Ghuge2 Athang Kulkarni3 Prof. Sukhada Bhingarkar4 Prof. Dr.Netra Lokhande5

4 Tracking System For Mentally Challenged People 15-18

Jagruti Bhagwan Durgude, Pooja kisan Aher, Pooja Somanath Avhad

5 Big Data Analysis for Associated Cloud Service using Parallel Trust Computing: A Survey. 19-24

1Veda Waikul, 2Onkar Ravgan, 3Prof. Aruna Pavate

6 A Technological Solution For Cleaning And Maintenance Of Sewerage System 25-28

Craig Menezes, Pruthav Singhadia, Ashukalp Srivastava, Prachi Janrao

7 Implement System To Predicttrends From Business Report Using Machine Learning 29-34

Niranjan Abhyankar, Rushikesh Shinde, Shubham Dukare, Dhananjay Gadakh, Rohini A. Patil

8 Big Data Analysis for Associated Cloud Service using Parallel Trust Computing: A Survey. 35-40

Akshay A. Jain1 Rajashree Gadhave2

9 Ensuring Safety of an Individual by using EEG and Machine Learning 41-46

Dr. Nazneen Ansari, Shraddha Pawar

10 Tracking System For Mentally Challenged People 47-52

Bhagyashri N Bachhav, Darshana A Deore, Pranjali B Deore, Sarika M. Jagtap

11 Intelligent Tourist System 53-58

Atanu Roy, Swapnil Kothari, Kaustubh Chakkarwar

12 Driverless system to enhance safety with voice assistance 59-64

Yugantari Kailas Mahajan, Pooja Pramod Saner, Mansi Sanjay Bihani, Dr. V. M. Birari

13 Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic 65-70

Jaineel Parmar , Pranit Patil , Ritwik Singh, Ms.Meghrani Patil

14	Implement System To Predicttrends From Business Report Using Machine Learning ***1Aruna Pavate, 2Pranav Nerurkar,3 Mansi Shah, 4 Rajesh Bansode, 5 Anita Chaudhary***	71-76
15	Invisible Video Watermarking for Data Integrity and Security based on Discrete Wavelet Transform – A Review ***Nitin A. Dhawas, Sambhaji A. Patil***	77-80
16	Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic ***Dhaval Desai, LakhanKabra, Mohit Jain, Mr. Shailesh Sangle***	81-84
17	Implement System To Predicttrends From Business Report Using Machine Learning ***P. G. Magdum, Sheetal Rathi***	85-88
18	Big Data Analysis for Associated Cloud Service using Parallel Trust Computing: A Survey. ***Prasad Pol, Vijay Borkar, Mahesh Bakve, Manil Kapadia, Mr. Vikas Singh (Guide)***	89-94
19	Ensuring Safety of an Individual by using EEG and Machine Learning ***1Shraddha More, 2Dr.Rajesh Bansode, 3Anita Chaudhari, 4Aruna Pavate, 5Ms.Brinzel Rodrigues***	95-98
20	Prediction of Rainfall in Pune City Region Using Deep Neural Networks ***Sneha L.Gaikwad, Pragati K.Bachhav, Shraddha B.Gangurde, Tejaswini S.Deshmukh***	99-102
21	Design and Develop Sentiment Classifier Model for a Customer using Support Vector Machine ***Kshtiij Poojary, Prashant Sharma, Varun Toshniwal, Mrs.Lydia Suganya Mrs.Lydia Suganya***	103-106
22	Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic ***Shivani Patel, Pooja Songadkar, Mr. Nitin Harane***	107-110
23	Implement System To Predicttrends From Business Report Using Machine Learning ***Nayana Agwan, Vaishali Nirgude***	111-114
24	Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic ***Mrs. Neha Kamble, Mrs. Rashmi Thakur***	115-118
25	Survey on MRI Image Segmentation for Cancer Detection ***1Mrs. Ujwala V. Gaikwad, 2Mr. Vishwajit B. Gaikwad, 3Dr. Kamal Shah***	119-122
26	Column-Oriented NoSQL Database : An Optimized Alternative for Analytical Environment ***Er. Swati Ganar,Gulhasan Siddiquee, Attaullah Khan,Soyab Anwar***	123-126
27	Plagiarism Checker Using Natural Language Processing ***Bhagyashree Mahajan, Prathamesh More, Prashant Gupta, Guide- Dr. Rekha Sharma***	127-130
28	Implementation of Stock Market Prediction Using Regression ***1Dhekale Prasad,2Jamdade Pradip,3Rahinj Swapnil 4 Shantanu Rai 5S V Athawal***	131-134

29	Intelligent Tourist System ***Yogesh Ghadigaonkar, Prof. Sunil Somani, Prof. Kalpana Bapat***	135-138
30	Tracking System For Mentally Challenged People ***Vijay M. Birari, Jayashree L. Gadhave, Kaveri B. Handore, Devashree M. Jadhav***	139-142
31	Web Application on Result Analysis ***Akshada D. Avhad, Priti R. Gojare, Santosh S. Nagpure, Sarika M. Jagtap***	143-148
32	Tracking System For Mentally Challenged People ***Lalit Nitin Chopade,Suwarna Prasad Rao,Rani Sham Shinde,Prof.Balasaheb Jagannath Pawar***	149-152
33	Driverless system to enhance safety with voice assistance ***Aishwarya Ramnath Shelar, Ashlesha Bhagwat Sable,Minal Gangaram Lokhande,Viraj R. Sonawane***	153-156
34	Driverless system to enhance safety with voice assistance ***Tejaswini C. Janjale,Shubham K. Khodade,Vishal A. Murkute,Atul R. Chaudhari***	157-162
35	Ensuring Safety of an Individual by using EEG and Machine Learning ***Akshay Pakhle, Ashvin Khairnar, Animesh Barmukh, Namita Pulgam, Mohammed Waqar Ali***	163-166
36	Plagiarism Checker Using Natural Language Processing ***Supriya Mandlik, Sakshi Mudrale, Ravita Mishra, Apurva Rawal, Prachi Mirkute***	167-170
37	Big Data Analysis for Associated Cloud Service using Parallel Trust Computing: A Survey. ***Suraj R. Dokhale, Nayan V. Nikam, Akshaykumar G. Ushir, Arvind P. Meshram***	171-174
38	Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic ***Priyanka Bolinjkar, Niki Modi, Harshali Patil***	175-178
39	Column-Oriented NoSQL Database : An Optimized Alternative for Analytical Environment ***Shrinidhi Gindi, Saad Naik, Safa Ansari, Sufiyan Shah***	179-182
40	Column-Oriented NoSQL Database : An Optimized Alternative for Analytical Environment ***1Lallan Pandit, 2Sankalp Mukadam, 3Samant Sahani, 4Rishabh Pathak, 5Guide- Prof. Satish M. Ranbhise***	183-186
41	Driverless system to enhance safety with voice assistance ***Pooja Yadav, Kamini Nawale, Isha Raval, Lydia Suganya***	187-188
42	A Technological Solution For Cleaning And Maintenance Of Sewerage System ***Ribah Shaikh, Sushmita Shinde, Rishabh Singh, Mrs. Purvi Sankhe***	189-192
43	Crop Prediction by Soil Analysis using Enhanced Random Forest Algorithm for Vidarbha Regions ***Nikunj Sartan, Varun Jasani, Akshit Singh, Mrs. Ashwini Patil***	193-194
44	Android Based Complaint Management System ***Ravishankar S. Singh1, Pratik S. Singh2, Swaraj S. Kasliwal3, Mrs. Veena Kulkarni4***	195-198

45	Machine Learning: Review of Algorithms	199-202
	Saurabh Singh, Adesh Borchate, Gaurav Shukla, Guide-Mrs. Harshali Patil	
46	An Intelligent College Enquiry Voice Chatbot	203-208
	Ms. Vani Singala, Ms. Nidhi Mishra, Ms. Aneri Moradia, Ms. Sonu Mehta, Mrs. MegharaniPatil	
47	Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic	209-214
	Mr. Prakhar Agrawal, Mr. Dhruvil Parekh, Mr. Rohan Gupta, Mrs.Megharani Patil	
48	UI based Content Management System	215-218
	Saurabh Singh, Shubham Vishwakarma, Ajaykumar Verma, Rahul Singh, Dr. Sheetal Rathi	
49	Webpage ranking Search Engine with SEO Suggester	219-222
	Mr. TejasSutone, Mr. VikasKaul, Mr. Harsh Bharucha, Mr. S. K. Narayankhedkar	
50	Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic	223-228
	Isheta Tilara,Purvi Sankhe	
51	Android Based Complaint Management System	229-234
	Sangeeta Vhatkar, Mohitraj Shah, ShashiKumar Thakur, Shubham Singh	
52	Android Based Voting Application	235-240
	Mary Margarat Valentine,Rupesh Vishwakarma, Abhishek Singh	
53	College Management System And Data Analysis With Machine Learning	241-244
	Meet Gadoya, Raj Asha, Saurabh Maydeo, Mary Maragarat	
54	An Intelligent College Enquiry Voice Chatbot	245-248
	Vivek Kumar, Bharat Paliwal, Vedprakash Jha, Neha Patwari	
55	Web Application on Result Analysis	249-254
	Kameshwar Gupta, Pratikkumar Mishra, Akanksha Das, Bhushan Nemade	
56	Android Based Car Tracking System	255-258
	Karan Pawar, Aditya Gupta, Dr. Radhika Kotecha (Professor)	

57	Intelligent Tourist System	259-264
	Rohan Shetty, Priyanka Verma, Dharti Singh, Radhika Kotecha	
58	Smart Warehouse – Reduction in postharvest losses	265-270
	Aishwarya Nair, Anand Ranka, Vikram Deshmukh, Dr. Kamal Shah	
59	Tracking System For Mentally Challenged People	271-274
	Chirag Rai, Mr. Namdeo Badhe, Aditya rao Nuty, Ayush Kedia	
60	Design and Development of Language Translation and Detection System using Machine Learning	275-280
	Ravi Sitapara, Pradeep Kotian, Prashant Chaudhary, Shridhar Kamble	
61	Web Application on Result Analysis	281-284
	Kirtan Patel, Maulik Shah, Rishab Tiwari, Ashish Soni	
62	Design and develop algorithm to translate audio into multiple Indian languages scripts using ML	285-288
	Ms.Akshata Hirve, Mr.Amey Patil, Mr. Anand Khandare	
63	Prediction of Rainfall in Pune City Region Using Deep Neural Networks	289-292
	Mr.Sanat Moharir, Dr. Nitin Pise	
64	Driverless system to enhance safety with voice assistance	293-296
	Pooja B.Bhamare, Manjusha A. Adsare,, Bhushankumar N. Shinde.	
65	Implementation of Effective Learning for Adolescents by Digital Solution and Evaluation Process	297-302
	Shriya Sundriyal, Vaibhav Singh, Amit Sharma, Mr. Shailesh Sangle	
66	Design and Development of Language Translation and Detection System using Machine Learning	303-306
	Aishwarya Rajeshkumar Verma	
67	Improved System For Appointment Scheduling In Hospitals	307-308
	Rohit Ragde , Disha Yadav , Sudeet Mhatre , Mrs. Rashmi Thakur(Guide)	
68	Smart Dustbin for Segregation of Wet and Dry Waste	309-312
	Sagar Pathare, Annanya Sah, Saketsingh Sisodiya, Shiwani Gupta	
69	Heart Disease Classification Using Optimized Machine Learning Algorithm	313-318
	Ms. Aishwarya Manjalkar PG Student , Mrs. Shiwani Gupta Assistant Professor	

70	Classification of Image Segmentation for Brain Abnormality	319-322
	Mr Shreyash Sharma PG Student, Mrs. Megharani Patil Assistant Professor	
71	Crop Prediction by Soil Analysis using Enhanced Random Forest Algorithm for Vidarbha Regions	323-326
	Priyankar Ravindra Tiwari, Prof. Anand Khandare	
72	Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic	327-328
	Divij Shah, Paridhi Somani, Siddhita Tawde	
73	Design and Develop Sentiment Classifier Model for a Customer using Support Vector Machine	329-332
	Rishabh Singh, Shreya Singh, Akshay Tripathi	
74	Visually Aware Fashion Recommendation System	333-336
	Anushi Doshi , Jaimish Ashar, Shridhar Kamble Assistant Professor	
75	Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic	337-340
	Ms. Ishita Gupta, Mr. Anay Awasthi, Mr. Sagar Gour, Dr. Sheetal Rathi	
76	Ensuring Safety of an Individual by using EEG and Machine Learning	341-344
	Prof. Dr. Sharmishtha Desai, Prof. Shamla Mantri, Adesh Raut, Kshitij Nirgude, Mrunali Ruikar, Aarohi Mali	
77	Plagiarism Checker Using Natural Language Processing	345-350
	Gautamkumar Ojha, Mihir Rane, Vivek Saroj, Dr. Rekha Sharma	
78	Big Data Analysis for Associated Cloud Service using Parallel Trust Computing: A Survey.	351-356
	Monika Gedam, Bharati Ainapure	
79	Livestock Monitoring in Agriculture Using IOT	357-360
	Khushbu Shah, Kinjal Shah , Bhavin Thakkar	
80	Search Ranking Factors using Organic and Paid Marketing	361-368
	Nazneen Pendhari, Arwa Katib, Javeria Ansari, Uzma Sayed	
81	Implement System To Predicttrends From Business Report Using Machine Learning	369-370
	Vishwa Solanki, Raj Shah	
82	Emotion Classification in Speech Signals Using Feature Space Reduction and SVM	371-374
	Mr. Vishwajit B. Gaikwad, Mrs. Ujwala V. Gaikwad	
83	A Technological Solution For Cleaning And Maintenance Of Sewerage System	375-378
	Heena Ganesh Shirke Prathamesh Vijay Rawool Akshata Anand Salvi Information Technology Department Information	
84	Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic	379-382
	Er .Zainab Mirza(HOD), Mr. Taha Bikanerwala, Mr. Mohd. Shibli Qureshi, Mr. Husain Amreliwala	
85	Column-Oriented NoSQL Database : An Optimized Alternative for Analytical Environment	383-388
	Er. Zainab Mirza (HOD), Pradeepkumar Vishwakarma, Arbaaz Tamboli, Wajeed Sharief	

86	Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic	389-392
	Roshakumar R. Maurya, DhiajYadav, Anand Khandare	
87	Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic	393-398
	Kajal Jaiswal, Jay Pal, Dr .Sheetal Rathi	
88	Forecast of Foreign Exchange Prices with techniques of Regression.	399-402
	Yadnesh Mohan Zagade, Shivani Gupta, Siddhesh Vinay Vartak, Dr. Sheetal Rathi	
89	Integration of C# with CUDA for performance improvement of Stand alone applications	403-404
	Abhinav Watve, Dr. Vinayak Bharadi HOD, Siddhesh Salvi, Sarvesh Horambe	
90	Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic	405-406
	Nikhil Pandey, Amit Tiwari, Ekta Tiwari	
91	Design of Efficient URL based Web Page Classification using NLP	407-410
	Karan Risbud, Rohit Singh, Vivek Vishwakarma, Dr. R. R. Sedamkar	
92	Development of a Knowledge Management Tool for Calculating Attainments of Course and Program Outcome	411-414
	Md. Awes Shaikh, Ashutosh Singh, Saurabh Singh, Jigarkumar Soni	
93	Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic	415-416
	Chinmay Pethe, Anagha Kiran Mundale, Priyanka Prakash Parab, Prof. Deven Ketkar	
94	Look Based Media Player	417-420
	Vandana Munde, Atulkumar Mishra, Abhijit Chaudhary, Rahul Vishwakarma	
95	Estimation and Simulation of Distribution of Liability Insurance	421-224
	Ganesh Mishra, Smit Supeda, Abhishal Yadav, Kaushal Yadav, Gayatri Dantal, Sunil Yadav	
96	Estimation and Simulation of Distribution of Liability Insurance	452-428
	Parvaneh Rafieisangari Chongqi Zhang	
97	Capstone Project-Big Data	429-430
	Anurag Singh, Nishant Singh, PranaySurana	
98	Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic	431-434
	Mrs. Veena Kulkarni , Mrs. Lydia Suganya	

COMMUNICATION NETWORKS

99	Guest Tagging System Analysis for Surveillance using IoT	435-442
	Prachi Jain, Drishti Kadam, Dr. Deven Shah	
100	To Study and Implement IEEE 802.11g Security Aspects Using Index Policy Method	443-448
	Yamisha Savalia, Ashwini Nakil, Jasmina Ambani, Dr. Rajesh Bansode	
101	Predicting Outbreak of a Disease through Epidemiology using Data from Twitter	449-454
	Sneha Milind Yadav, Dr. Brijit Marakarkandya	
102	Emotion Detection and Feedback Generation	455-458
	Akshdeep Sharma, Rishabh Gupta, Shivangi Singh, Parth Khurana	
103	Face Recognition using Raspberry PI	459-464
	Shruti AnantAmbre Mamata VijayMasurekar Shreya Shivaji Gaikwad	
104	Transport Vehicle Selection Predictor	465-470
	Shubham Matale, Apurva Taru, Pratik Mahamuni,Sukhada Bhingarkar	
105	Investigating Enterprise Software System Integration Prototypes Involved in Integrating Disparate Software Systems	471-474
	Nazneen Ansari, Sahil Raut	
106	A Comparative Study on Recent Trends and Applications in Cloud Computing	475-480
	Jaishree Tawade, Vikas Kaul	
107	Improvement In Productivity And Income Using Farming Assistance	481-482
	Mr. Jugal Parulekar,Ms. Shraddha Pathare ,Ms. Kajal Dhakate,Ms. Chaitali Waghulde, Mrs. Surekha Janrao	
108	Smart Water Management System and IOT	483-486
	Shraddha Shrikrishna Adgaonkar, Rutuja Pradip Ambekar, Bhavana Dyaneshwar Deshmukh, Arvind Pandhari Meshram	
109	Heart Attack Detection using sensors and IOT	487-490
	Miloni Nisar, Purvesh Dave, Shridhar Kamble	
110	Career Counselling using Machine Learning Algorithms	491-494
	Pranav Shah, Shohan Shetty, Lakshit Ukani,Priyank Jethva	
111	Privacy-Preserving Social Media Data Publishing for Personalized Ranking-Based Recommendation	495-498
	Snehal Tukaram Patil	
112	Smart 3-Phase Water Pump Controller	499-502
	Gauri D. Gaikwad, Nikita S.. Labhade,, Shraddha B. Chavan, Balasaheb J.Pawar4	
113	Implementing secure system to visualize business report using data analytics.	503-504
	Abhishek Anand, Aakash Patel, Abhijeet Saha	

114	Development of Next Generation Encryption Algorithm for Cloud Security	505-508
	Siddhant Deshmukh, Aviral Gupta, Sagar Gupta, Viaks Kaul	
115	Study of Tubestile	509-512
	Akash J. Joshi,Shekhar V. Lokare,Urjita R. Jagtap,Dipali V. Patil	
116	IoT Based Weather Station	513-516
	Sunny Sall, Abhishek Yadav, Akash Rai, Sagar More, Ashish Jaiswar	
117	First Person Perspective (FPP) Facade for purchase of Estates	517-518
	Nuruddin Warsi, Gulamwaris Warsi, Komal Bhatnagar, Hetansh Shah, Anil Chaturvedi	
118	Handover Security Solution for HetNet In Millimeter Range Communication	519-522
	Devidas Chikhale, Shankar Deosarkar, Mahesh Munde	
119	Ad-hoc Network and Database Object Analytics	523-528
	Aniket Pathak, Sagar Parmar, Pankti Parikh, Vikas Kaul, Ranganathan Iyer, Manoj Nambiar	
120	Vehicle Tracking/Parking System	529-534
	Vandana Munde, Shreyas Snail, Yash Mujumdar, Shweta Panchal	
121	Authentication and Access Control	535-538
	Akanksha Jain, Shivani Rege, Pooja Jain, Mr. Sudhir Dhekane	
122	Gabor-Canny Algorithm for Under Vehicle Object Detection Application	539-542
	P.G.Salunke, Dr.S.T.Gandhe, J.S.Soni	
123	Smart Signalling and Emergency Vehicle Management System	543-548
	Vidya Prashant More, Sharmishtha desai	
124	Blockchain Based Automated Trading Platform	549-552
	Neil Naidu, Sidddhesh Bhatgaonkar, Abhishek Delwadia, Megharani Patil	
125	Efficient System to Prevent Session Hijacking Attack on Servers Using Proxy Server	553-558
	Anuradha Maurya, Veena Kulkarni	
126	Embedded System for Remote Controlled Automobile using RFID	559-564
	Keyur Pandya, Pareen Harsora, Kruten Patel	
127	Cloud Based Automated Malware Analysis for Advanced Persistent Threats For Government organization	565-568
	Nikhil Pandey Amit Tiwari Ekta Tiwari	
128	Design of Smart Ambulance System with an added feature of Traffic Overriding	569-574
	Lydia Suganya	
129	IOT Based Fire Fire Fighting Robot using Arduino	575-578
	Chinmay Verlekar, Pratik Patel, Manoj Dixit, Aakash Kanojiya, Vaishali Nirgude	

130	Efficient Recommendation of IT Problems using Image Recognition	579-582
	Divij Shah, Paridhi Somani, Siddhita Tawde	
131	Online shopping Engine to search commodities	583-586
	Saurabh Rai,Varun Shah, Drishti Sisodiya, Riddhi Vaidya, Shiwani Gupta	
132	Wireless Sensor Network For Environmental Monitoring And Control	587-592
	Pankaj Malviya, Deep Modi, Kautilya Joshi, Zahir Aalam	
133	RFID Based Electronic Payment System	593-600
	Shreya Mistry, Nishita Desai, Tanay Sanghavi, Shruti Mathur	
134	Raspberry Pi Based Cost Effective Interactive Device for Blind, Visually Impaired, Deaf and Dumb People	601-606
	Vijay Karotra, Shubham Maheshwari, Tanishq Mehra, Shridhar Kamble	
135	Comparison of time series forecasting model for predictive analysis	607-612
	Munaf Patel, Zahir Aalam	
136	Solving Communication Networking Problems with Machine Learning Approach: A Case Study	613-616
	Nidhi Kamothi, Himani Tiwari, Megharani Patil	
137	Determining Level of Cognition Of Mild Cognitive Impaired Person Using Screening Test and EEG Signals	617-620
	Amruta Koli, Prof. Shredevi Karande	
138	Evaluation Of Aurdino Based Soil Moisture Measurement Using IOT	621-626
	Renuka Sahani,Manu Nair,Smriti Saxena, Ms Vaishali Nirgude	
139	Disease Prediction System Using Data Mining	627-630
	Surabhi Pandey, Gauraja Parab, Srushti Sharma, Rashmi Thakur	
140	Implementation Of Fake Fingerprint Image Detection System	631-636
	Akshay Kumbhar, Akshay Jaiswal, Saurabh Yadav, Jesalkumari Varolia	
141	Business Intelligence using Internet of Things (IoT) : A Case Study	637-640
	Avani Agrawal, Pooja Patil, Rajesh Bansode	
142	Smartphone Sensors based Rubik's Cube Timer	641-646
	Sanam Kazi, Aliza Shaikh, R. Ihsaan Muhiyadheen, Tazeen Ansari	
143	Biometric Identification And Real Time Control and Security System in Smart Car (BIRCS3)	647-652
	Riya Parikh, Manasi Kadam,	
144	Security in Hadoop	653-656
	Omkar Shinde, Sweta Mishra	
145	Event Management and Analysis System for College Fests	657-660
	Mariyam Jawadwala, Priyank Patel, Hinal Patel, Priyanka Patil	

146	The Alternate Method of Agriculture: Aquaponics Using IOT	661-664
	Maryam. Jawadwala, Yogesh. Pingle, Soumya. Kulal, Bhumika Mate	
147	Svara – The Music Therapy App	665-668
	Swapnil H. Mistry, Ashok R. Patel, Sridhar Subramanian, Prof. Yogesh P Pingle	
148	Best Out Of Waste(Recycling Waste)	669-672
	Milind Purohit, Jaspreet Kaur, Hitakshi Patel, Aditi Pandit, Prof. Yogesh Pingle	
149	Flora Export	673-674
	Jaspreet Kaur Thethi, Hitakshi Patel, YogeshPingle	
150	Blockchain For Electronic Health Records(EHR)	675-678
	Sarthak Taneja, Yashvardhan Tewari, Gaurang Tyagi	
151	IOT Based Solution for Road Surveillance and Profiling	679-684
	Shrinidhi Gindi, Purva Surve, Rukhsar Khan	

Student Performance Analysis and Counselling System (SPACS) using Soft Computing

Prof. Bharati Ainapure, Anuj Mahajan, Shivam Pawargi, Nidhi Sadany

Computer Engineering, MIT College of Engineering, Pune, India

bharati.ainapure@mitcoe.edu.in, anujsmahajan98@gmail.com, shivampawargi@gmail.com, nidhisadany@gmail.com

***Abstract*— In this modern world, students are vulnerable to many distractions which affects their education and in turn their career. Career counselling is a methodology that will help the students to recognize themselves and the world of work in order to build their career, education and make life decisions. Taking this social aspect into consideration, this paper discusses the proposed system, named Student Performance Analysis and Counselling System (SPACS), which will play a crucial role in analyzing the students' overall academic performance and identify the factors affecting it and will help counsel the student to improve the performance. Therefore, in this paper, analysis of various soft computing techniques for rule formation and decision making are discussed. This paper also provides directions to build such a system that will take input in the form of student answers related to a number of factors like emotional stability, economic stability, academics, etc. and generate guidelines to improve the performance. The student will be able to use these guidelines to assess his performance and get a clearer view on the line of action for his future endeavors.**

Keywords*— *Data Mining, Soft Computing, Artificial Intelligence, Students' Performance Analysis and Counselling System (SPACS), Machine Learning.

I. INTRODUCTION

Career is a much larger and significant part of our lives today than ever before. As there are numerous opportunities, jobs, institutions, specializations, courses; day by day it is becoming difficult to narrow down a career path for a candidate and get access to personalized counselling to choose and make the most appropriate decision as per his/her inclination. As education plays an integral part in a students' development, it is necessary that the student receives the right education. To be able to make use of the right education, the student needs to be in the right atmosphere and the right mind set. Various factors like behavioural science, economic stability, exposure to the world, participation in co-curricular and extra-curricular activities, etc. will determine the kind of atmosphere the student is in. With so many factors deciding the ideal career path of a student, to have a balance between his/her strengths and weaknesses plays a major role. In accordance to that, not every candidate feels free and comfortable enough to discuss their weaknesses. Hence, focusing on the serious social aspect of career counselling this proposed system aims to simulate a career advisor and provide counselling on a software application in isolation [1]. The difference between sharing information with a human counsellor and a software system is that the human counsellor might get biased towards decision making for the candidate just by looking at the patterns of his/her past on a superficial level. Whereas, the proposed system will make complete use of data to analyse, make a proper report and generate guidelines for the candidate to help make a correct decision of his/her career. The student will feel more comfortable knowing that the results are unbiased and the judgment is given without any human intervention. The guidelines generated by the system will give the student more clarity and further help him choose a relevant career path. Also, the desired career path engages an individual making them more informed, confident and dedicated. A right career path, therefore, can mould the entire way of life.

This paper has five sections. The section on literature survey discusses five papers with their methodologies in brief and their architectural diagrams. After literature survey, is the problem definition. The next section discusses the architectural diagram of the proposed system explaining the use of different modules. The section on methodologies discusses various algorithms that can be used to implement the proposed system. The important steps are mentioned in brief. The last section discusses the conclusion and future scope.

II. LITERATURE SURVEY

In 2016, Alibek Barlybayev, Altynbek Sharipbay, Gulden Ulyukova, Talgat Sabyrov and Batyrkhan Kuzenbayeva proposed a fuzzy logic based students' performance evaluation system. The proposed system forms rules from the academic factors and evaluates the weight of each factor by calculating arithmetic mean. This method is not efficient enough. The proposed system also tries to establish a relationship among the factors [2].

In 2016, Ritu Bansal and Vishu Madaan proposed a fuzzy logic and association rule mining method for student performance analysis and counselling system. They have considered five major factors namely personal factor, family factor, university factor, academic factor and teaching factor. On the basis of these, apriori algorithm is applied to identify most relevant factors and then fuzzy rule based expert system is designed. Using association rule mining algorithm reduces the complexity of the system. This system does not take into account the emotional stability of the student [3].

Figure 1. Shows the workflow of the proposed system in the paper.

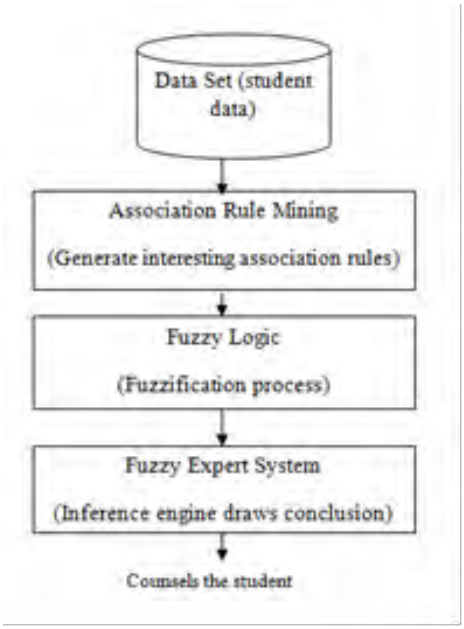


Figure 1. Workflow of the System [3].

In 2017, Archana Parab, Siddhesh Palkar, Satish Maurya and Sonal Balpande integrated Natural Language Processing (NLP) with Artificial Intelligence (AI) to build an intelligent career counselling bot. The answers given by the user are interpreted using NLP and relevant keywords are extracted from the knowledge base to interpret the results. The system determines the most influential factors that affect the performance. However, the proposed system takes into consideration a predefined knowledge base which makes the system robust and rigid [4].

Figure 2. shows the workflow of the system proposed in above paper.

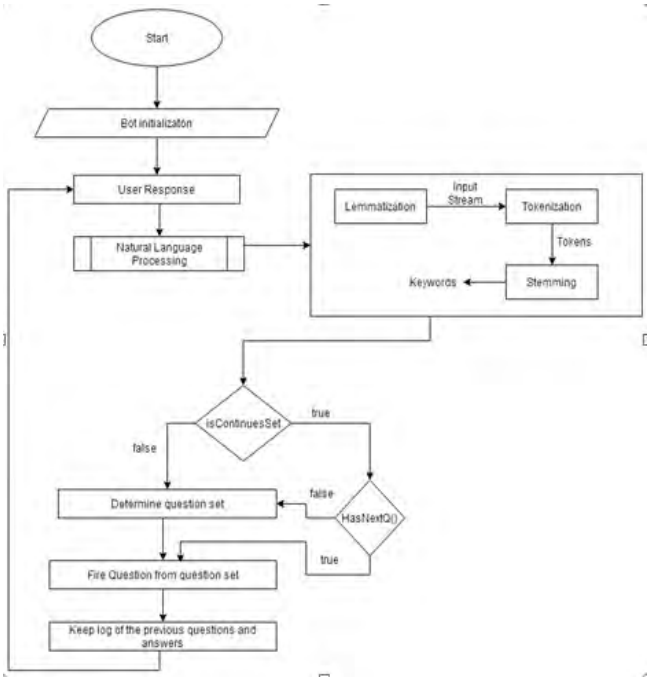


Figure 2. Workflow of the System [4].

In 2017, Kuboye Bamidele Moses and Emmanuel Onwuka Ibam have proposed a system that uses forward chaining to evaluate students' performance based on students' age, courses selected, favourite subjects and by calculating the intelligent quotient (IQ). IQ is calculated using Mental Age (Level of Intellectual Performance) and Chronological Age. The proposed system does not incorporate the non-academic factors like extra-curricular activities or the economic stability of the student or the emotional factors of the student [1].

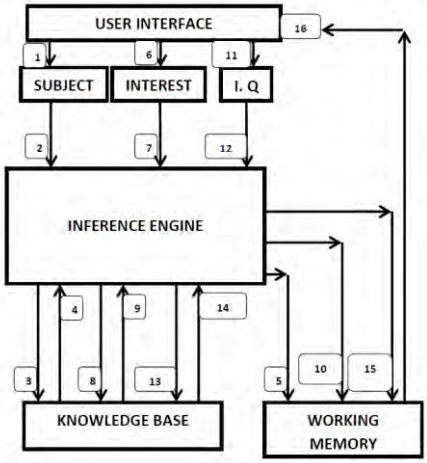


Figure 3 Architecture of the System [1].

Figure 3. shows the workflow of the system proposed in above paper.

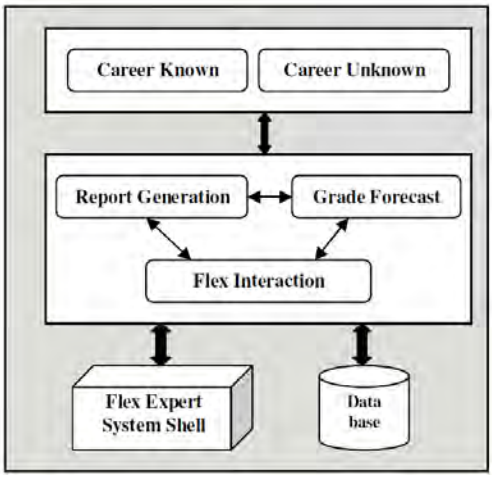


Figure 4. Workflow of the System [5].

In 2006, Chathra Hendahewa, Maheshika Dissanayake, Savindhi Samaraweera, Narmada Wijayawickrama, Anusha Ruwanpathirana and Asoka S. Karunananda used the forward chaining and backward chaining methodology in their proposed system for artificial intelligence approach to effective career guidance. The system uses flex expert system shell. The system calculates the GPA (Grade Point Average) from students' past performance and suggests career options based on students' interests as well as

GPA. The disadvantage of the system is it does not account for non-academic factors [5].

Table 1 shows the comparative study of various algorithms studied in Literature Survey.

TABLE 1. LITERATURE SURVEY

Sr. No	Title	Author	Methodology	Advantages	Disadvantages
1	Students performance evaluation by fuzzy logic	Alibek Barlybayev, Altynbek Sharipbay, Gulden Ulyukova, Talgat Sabyrov,Batyrkhan Kuzenbayeva	Fuzzy Logic	Relation between the factors	Not efficient enough
2	Students Performance Analysis and Counselling System	Ritu Banswal, Vishu Madaan	Fuzzy logic and Association Rule Mining	Less Time Complexity	Emotional stability not considered
3	An Intelligent Career Counselling Bot	Archana Parab, Siddhesh Palkar, Satish Maurya, Sonal Balpande	NLP and AI	Determines the most influential factors	Pre-defined Knowledge Base
4	Development of Web based Intelligent Career Guidance System	Kuboye Bamidele Moses, Emmanuel Onwuka Ibam	Backward Chaining and Forward Chaining	The system is intelligent.	Emotional stability not considered
5	Artificial Intelligence Approach to Effective Career Guidance	Chathra Hendahewa, Maheshika Dissanayake, Savindhi Samaraweera, Narmada Wijayawickrama, Anusha Ruwanpathirana, Asoka S. Karunananda	Backward Chaining and Forward Chaining, Flex intelligence server	The system is intelligent.	Emotional stability not considered

III. PROBLEM DEFINITION

With so many career options to choose from, a candidate is prone to confusion. Various factors like students' economic stability, exposure to the world, participation in co-curricular and extra-curricular activities, emotional stability, etc. determine his choice. In accordance to that, not every student feels free and comfortable enough to discuss his decision with others. That is why there is a need of an unbiased system that makes complete use of data to analyse, make a proper report and generate guidelines for the candidate to help make a decision of his/her career. This paper gives directions to building such a system specifying various algorithms that can be used.

IV. ARCHITECTURAL VIEW OF THE SYSTEM

Figure. 5 shows the basic architecture of the proposed system. It describes and gives an insight on what are the different modules of the system and how they interact with each other.

User will fill up all the details which is provided by the user interface. The answers given by the user to the questionnaire provided by the system will be fed to the inference engine. There is interaction between Inference Engine and Knowledge Base for the rule formation and decision making. The basic data required is extracted from database. The modules are explained in detail below.

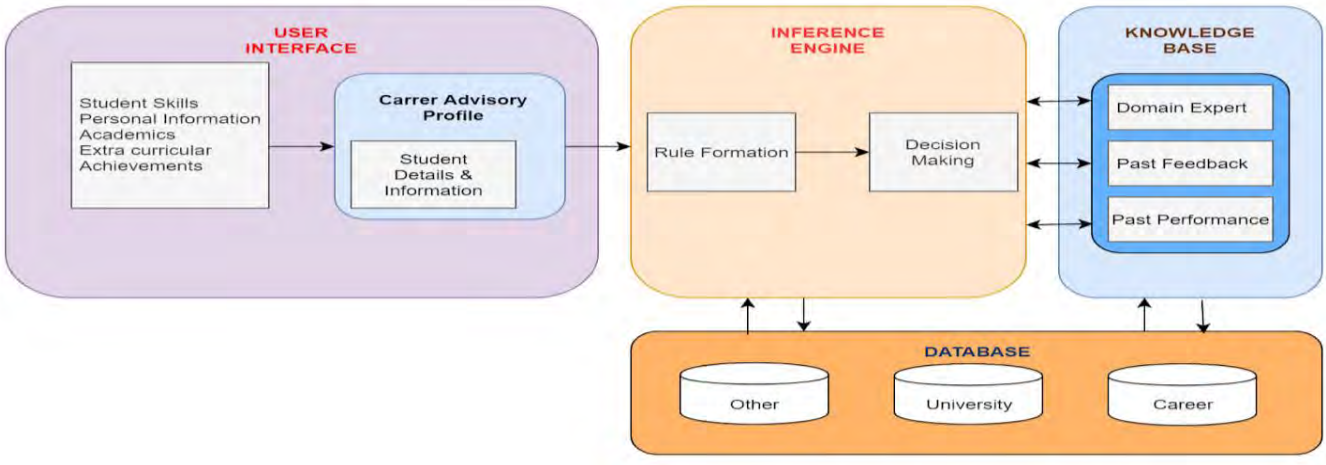


Figure 2. Architectural diagram of the proposed system

The system architecture consists of following modules:

- User Interface: This will allow interaction between the user and the system. After successfully logging in, the user will be able to answer a set of relevant questions designed to understand his interests, economic stability, etc. The answers given by the user will be extracted and fed to the inference engine for further processing.
- Inference engine: This module will apply logical rules to the knowledge base coupled with answers given by user to deduce information and take decision accordingly. Various algorithms that can be used to apply rules are discussed in Section V.
- Knowledge Base: This module will hold the domain knowledge and other structured and unstructured knowledge vital for decision making. The inference engine will use this knowledge and map them with the user responses to take further decisions. These decisions taken will be shown to the user in the form of career guidelines.
- Database: This will hold other information, mostly static information, like user credentials, university details, etc.

IV. Methodolodies

The first proposed methodology deals with the chaining of rules. The methods of chaining are discussed.

A. Rule Based Expert System

The rule based expert system is made up of knowledge represented as a set of rules. Any rule consists of two parts: The IF part, also called the antecedent (premise or condition) and the THEN part, also called the consequent (conclusion or action).
The basic syntax of this rule structure is:
IF antecedent
THEN Consequent

The hierarchies of IF condition proceeds forward by checking every IF condition and activating the relevant then part of that rule.

If 'A' is true then check condition 'B'
If 'B' is true then answer is 'C'

The answer 'C' is activated if condition 'A' and 'B' is satisfied.

For Example:
If grades > A then
If Family Income > 1000000 then
Student should opt for MS

If Family Income < 1000000 then
Student should opt for M.Tech

Multiple antecedents can be joined by the keywords AND, OR, or a combination of both.
These IF-THEN rules are chained and used by the inference engine to manipulate knowledge in the knowledge base to form a line of reasoning.

Types of chaining are:

- Forward Chaining
- Backward Chaining

a) Forward Chaining

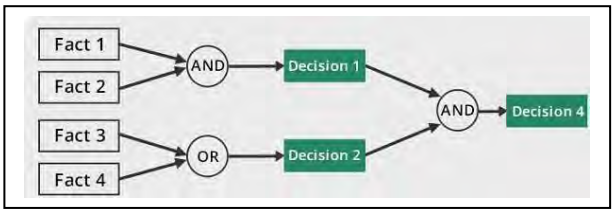


Figure 3. Forward Chaining [9].

Forward chaining is called data driven reasoning because the reasoning starts from the known data and moves forward to evaluate the solution. It searches for the rules that move from the current state of the problem to the final solution. Whenever a rule is fired, the new facts are added to the database.

When there are thousands of rules, the problem arises in forward chaining is that the unnecessary rules can also get fired. This increases the execution time and we get the facts might be far from what is required. Backward Chaining is used as a solution [6].

b) Backward Chaining

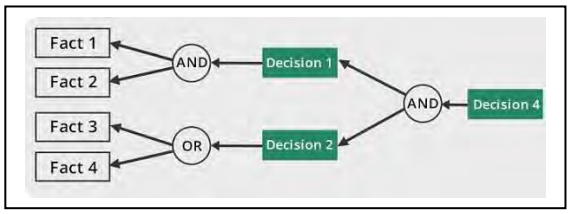


Figure 4. Backward Chaining [9].

Backward chaining is called goal driven reasoning. In this process, the system already knows what the goal is, the inference engine tries to find the facts which result into the goal. First, the inference engine will search the knowledge

base to find rules that have required a solution and such rule will have goal in their action (THEN) parts. If such case is found and if its IF part matches the data in database then rule is fired and goal is proven to be correct [7].

The next proposed methodology deals with uncertain knowledge which cannot be given Boolean values. This algorithm works in majorly four steps.

B. *Fuzzy Expert System*

To take care of the uncertain, incomplete knowledge, a rule can be associated with a weight or a confidence factor. Using uncertain knowledge with uncertain data is called reasoning with uncertainty. Such systems are called Fuzzy systems and the logic used is called fuzzy logic. It is also known as multi valued logic. In classical model there are only two truth values that is True (1) and False (0). In fuzzy logic all the values are described on the basis of real number system from the interval between 0 to 1. This method answers the question "How true is the value?" It gives the measure of truthfulness.

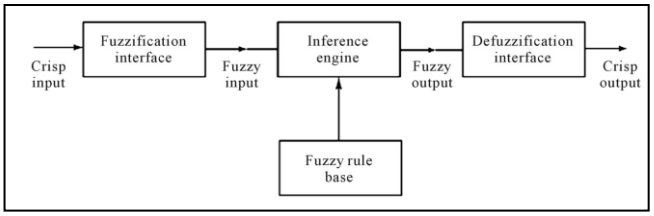


Figure 5. Steps in Fuzzy Expert System [8].

a) Step 1: Fuzzification

Crisp inputs A and B are taken into consideration in this step and degree is determined to which these inputs belong to each of the appropriate fuzzy sets.

b) Step 2: Rule evaluation

Fuzzified inputs are taken and they are applied to antecedents in this step.

c) Step 3: Aggregation of the rule outputs

Unification of output of all rules take place in this process, thus the clipped or scaled consequent membership functions act as a input to this process and one fuzzy set for each output variable is the output of this process.

d) Step 4: Defuzzification

The output of fuzzy system is represented by crisp number. Defuzzification is the process of converting output to crisp value. Thus aggregate output fuzzy set acts as a input to defuzzification process and output is a single number.

For example, a simple fuzzy rule can be written as:

If GPA=A and Interested Subject 1=Commerce and interested subject2=Law then recommend=Law

The next proposed methodology is based on how relevant is the new information given that the existing information is true. It involves majorly two steps of evaluation.

C. *Fuzzy K - Nearest Neighbour*

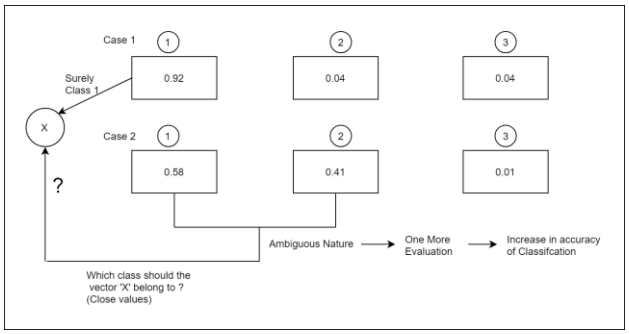


Figure 6. Fuzzy K - Nearest Neighbour.

The fuzzy K-nearest neighbour (KNN) algorithm allocates the membership function to a sample vector and does not assign any particular vector to a membership function. The advantage is, no random assignments to specific classifier groups are made. Also, the values accompanying the vector should justify in which membership function it should belong to. For example, the vector is assigned 0.92 in class one and 0.04 in class two and also class three then the system can be definitely sure that the vector will belong to the class having 0.92 membership. But if a vector is assigned 0.58 memberships in class one, 0.41 in class two and 0.01 in class three, then one more evaluation is needed for class one and two to be done to confirm the classification procedure as the system can be confident that it will not belong to class three. Clearly the process of membership allocation in the later part will help the system gain more accuracy and improve the level of classification [11].

The intension of the algorithm is to allocate membership as a function to vector's distance from its K-nearest neighbours and those neighbours membership in the possible classes. The advantage of Fuzzy KNN over KNN is that it takes into consideration the ambiguity of assignment in specific membership classes where it takes case that such ambiguous nature does not play a vital role in the classification technique. The second advantage being residues are assigned a membership value in each class rather than binary decision of "belongs to" or "does not belong to". Such type of assignment of membership values to different classes act as strength or confidence with which the current vector belongs to a particular class [11].

CONCLUSION AND FUTURE SCOPE

After studying various algorithms, this paper found that different data mining algorithms like K Nearest Neighbor (KNN) coupled with fuzzy used for pattern formation and classification of the candidate can be used for the implementation of the stated systems. It will help the system to predict more precise and accurate results. Further implementation of the stated systems can include more number of parameters which will help to give results from a broader perspective. The stated system can also be used by companies to be able to recruit the right candidate for their firm.

REFERENCES

[1] Kuboye Bamidele Moses, Emmanuel Onwuka Ibam, "Development of Web based Intelligent Career Guidance System for Pre-Tertiary Science Studnets in Nigeria," Circulation in Computer Science, Vol.2, No.8, pp: (4-17), September 2017, DOI: 10.22632/ccs-2017-252-51

[2] J. Alibek Barlybayev, Altynbek Sharipbay, Gulden Ulyukova, Talgat Sabyrov, Batyrkhan Kuzenbayeva, "Students performance evaluation by fuzzy logic," Procedia Computer Science 102 (2016) 98-105, doi: 10.1016/j.procs.2016.09.375

[3] Ritu Banswal, Vishu Madaan, "SPACS: Students Performance Analysis and Counselling System using Fuzzy logic and Association Rule Mining," International Journal of Computer Applications (0975- 8887) Volume 134 No.3, January 2016

[4] Archana Parab, Siddhesh Palkar, Satish Maurya, Sonal Balpande, "An Intelligent Career Counselling Bot-A System for counselling," International Research Journal of Engineering and Technology (IRJET)Volume: 04 Issue: 03, March 2017

[5] Chathra Hendahewa, Maheshika Dissanayake, Savindhi Samaraweera, Narmada Wijayawickrama, Anusha Ruwanpathirana, Asoka S. Karunananda, "Artificial Intelligence Approach to Effective Career Guidance," Sri Lanka Association for Artificial Intelligence (SLAAI), (September 2016)

[6] http://www.myreaders.info/07_Expert_Systems.pdf

[7] http://www.academia.dk/BiologiskAntropologi/Epidemiologi/DataMining/Artificial_Intelligence-A_Guide_to_Intelligent_Systems.pdf

[8] E. H. Mamdani and S. Assilian, "An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller," International Journal of Man-Machine Studies, Vol. 7, No. 1, 1975, pp. 1-13. doi:10.1016/S0020-7373(75)80002-2

[9] https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_expert_systems.htm

[10] Gufran Ahmad, "Career Guidance through Multilevel Expert System Using Data Mining Technique," I.J International Journal of Information Technology and Computer Science (August 2017) 8, 22-29, DOI: 10.5815/ijitcs.2017.08.03

[11] Akhilesh Latoria, Alok Singh Chauhan, Dr.Anand Swarup Saxena, "Fuzzy Approach for pattern recognition using Classification Algorithms," International Journal of Computers & Technology, Volume 3 No. 3, Nov-Dec, 2012

Detection and Classification of Brain Tumor Using Artificial Intelligence

Sunil Yadav, Divya Tiwari, Kantilal Vishwakarma

Asst. Prof, Information Technology, SLRTCE, Thane, India

sunvan.15@gmail.com, divyatiwari0301@gmail.com. kanti14314@gmail.com

***Abstract*— Brain Tumor is an abnormal intracranial growth caused by cells reproducing themselves in an uncontrolled manner. Most researches in the developed countries shows that main cause of death of people having Brain Tumor is incorrect detection of Brain Tumor. It is one of the most dangerous diseases and therefore it should be detected quickly and accurately. Generally, MRI or CT scan that is directed into the intracranial cavity produces the complete image of the Brain Tumor. Magnetic Resonance Imaging (MRI), a highly developed technique of medical imaging is used to visualize internal structure of human body without any surgery. For the accurate detection of Brain Tumor segmentation of MRI image is important. Classification of Brain Tumor through segmented MR images, is a difficult task due to complexity and alteration in Tumor tissue characteristics like its location, size, gray level intensities and shape. Nowadays due to increasing Tumor cases it is difficult to examine all the reports manually also it sometimes becomes hazardous for a patient due to delay in the detection of the Tumor or the right time required for its surgery.**

In order to eradicate this problem an intelligent system is required for the detection and classification of brain Tumor automatically. Intelligence can be generally described as the ability to perceive information and retain it as knowledge to be applied towards adaptive behaviors within an environment or context. We propose a method which can automatically consider the MR images as an input and further analyze and process the input image and classify according to its presence and absence along with the type of Tumor detected. This can be done using two machine learning based techniques like neural networks and deep learning which can be used to solve many real-world problems.

Keywords—Brain Tumor, MR Images, Neural Networks, Deep Learning, Intelligent System.

I. INTRODUCTION

Human Body comprises of several types of cells with each cell having a precise function. These cells grow and divide in an orderly manner which forms new cells to keep the body in good physical condition. Sometimes few cells cease their capability to control their growth and they grow in an improper fashion which leads to growth of extra cells forming a mass of tissue which is called a Tumor. Tumor when formed within a brain is called "Brain Tumor." Brain Tumor is one of the most difficult types of Tumor to fight because of its vast variety of types. In total, there are more than 120 different types of Brain Tumors. Brain Tumors are classified as either Primary or Secondary. A Primary Brain Tumor means the cancerous cells originate in the Brain. A Secondary Brain Tumor is a mass of cells that originates elsewhere in the body, then it spreads through the bloodstream to the brain in a process known as Metastasis. The most common secondary brain tumors arise from the lung, breast, kidney, colon and blood cells. Once a diagnosis is made, the medical team attempts to determine whether the Tumor is intra-axial (within the central nervous system) or extra-axial (located outside the central nervous system). Brain Tumors are classified as either Benign or Malignant, although this classification is not always precise. [14]

Everything in our surrounding is evolving rapidly whether it is case of humans or machines. In this rage of evolution another concept emerged, having both the efficiency of machine and thinking ability of humans, we recognize it as "Artificial Intelligence". AI can be applied to any field which requires the ability of thinking and analysis. Medical domain is one of such fields. Mostly doctors spend much of their time in the analysis of the medical reports. There are various machines available which can be used to find flaws in human body but it requires human intervention to reach conclusion. In order to reduce time of detecting issues and utilize that valuable time in treatment of the patient an Intelligent system is needed. So, we decided to develop an Intelligent system which will assist Neurological Surgeons in Detecting and Classifying Brain Tumor using Artificial Intelligence.

II. LITERATURE SURVEY

Brain Tumor Segmentation based on a Hybrid Clustering Technique, Eman Abdel-Maksoud et.al (2015) presented their methodology at Egyptian Informatics Journal, describing image segmentation as a process of partitioning an image into mutually exclusive regions. It can be considered as the most crucial process for facilitating the delineation, characterization and visualization of regions of interest in any medical image. A methodology of integrating K-mean clustering technique with Fuzzy C-mean algorithm for image segmentation is proposed. This process partitioned the image into mutually exclusive regions. The proposed technique can get benefits of the K-means clustering for image segmentation in the aspects of minimal computational time. In addition, it also gets advantages of the Fuzzy C-means in the aspects of accuracy. [1]

An Improved Brain MR Image Binarization Method as a pre-processing for abnormality detection and feature extraction, Sudipta Roy et.al (2017) proposed methodology in a Research Article, introducing a computerized method of MRI of Brain binarization for the uses of pre-processing of feature extraction and brain abnormality identification. One of the main problems faced in binarization is that many pixels of brain part cannot be correctly binarized due to extensive black background or large variation in contrast between background and foreground of MRI. In order to overcome such problems Binarization that uses mean, variance, standard deviation and entropy to determine a threshold

value followed by a non-gamut enhancement is used. The proposed binarization technique after testing with a variety of MRI generated good binarization with improved accuracy and reduced error. Here experimentation was performed on two steps and proved that the proposed method is capable of working in MR images and their application domains. This method overcomes the problem of large intensity difference of foreground and background of MR images and it does not have any over binarization or under binarization problem for different kinds of images. [12]

Classification using Deep Learning Neural Networks for Brain Tumors, Heba Mohsen et.al (2017) proposed an automated technique at Future Computing and Informatics Journal, introducing Deep Learning which is a new Machine Learning field that gained a lot of interest over the past few years. It is widely used and applied in several applications and has been proven to be a powerful machine learning tool for solving complex problems. The technique proposed used Deep Neural Network classifier which is one of the Deep Learning architectures for classifying a dataset of Brain MRIs into various classes like normal, glioblastoma, sarcoma and metastatic bronchogenic carcinoma Tumors etc. It was observed that the classifier was combined with the discrete wavelet Transform a powerful feature extraction tool and principal components analysis which proved to be performing better over all the performance measures. The advantage of using this technique is that it requires less hardware specifications and takes a convenient time of processing for large size images. In addition, using Deep Neural Network classifier accuracy rate was higher as compared to other traditional classifiers. In future it is possible that more better results can be achieved if Discrete Wavelet Transform is used with Convolutional Neural Network. [7]

III. Proposed Method

The purpose of this paper is to detect and classify the Tumorous region of the Brain which will be further used for detailed diagnosis of Tumor essential for the treatment of the patient. The details about the proposed method is given below.

Neural Networks and deep learning provide the best solutions to many problems in fields of image recognition, speech recognition, natural language processing etc. Earlier technical methods used for the detection of Brain Tumor were efficient but in some cases the methods failed to notice the Tumorous part correctly from MR images because the image consists of non-Tumor tissue. (various types of tissues which makes it difficult to identify or differentiate between tumor affected and unaffected region) For this reason we propose method which uses Convolutional Neural Networks which is best for image segmentation and classification of objects. Segmented image will help to detect the tumorous region (Region of Interest) and based on the detected ROI we will classify the Tumor whether it is Cancerous or Non-Cancerous.

A. DATASET

The Dataset which we will use for experimentation is obtained from BraTS 2018 Challenge. This data is basically provided for segmentation challenge. It has four modalities T1, T1c, T2 and Flair. It consists of training datasets, validation datasets and test data. The dataset received is pre-processed and is used for training the model along with comparing the result with ground truth which is already provided in the dataset.

B. FILTRATION

Filtration is a process of removing unwanted signals or noisy data that gets introduced in MR images during MRI scan due to electro-magnetic interferences. The MR images that we use for detecting Tumor thus consists of noisy data which is essential to be removed. Noise filtering techniques are basically of two types namely linear and non-linear. Linear filtering techniques apply the algorithm linearly to all the pixels of the image which defines the image as corrupted or uncorrupted pixel. Since, algorithm applies to all the pixels of image it causes the uncorrupted pixels to be filtered and hence becomes ineffective. The non-linear filtering technique is considered to be a two-phase process. In the first phase the pixels are identified as corrupted or uncorrupted and then further in the second phase corrupted pixel is filtered using specified algorithm while uncorrupted pixel value is retained. Most widely used non-linear filtration technique is median filter which uses median value to replace the corrupted pixel with the capability to remove impulsive noise while preserving the edges. Further on the basis of technique followed, median filters too are classified into many types out of which the Adaptive Median filter is found to be best. Adaptive Median Filter changes the size of the filtration window based on the conditions hence helps in retaining all the uncorrupted pixel values efficiently. It basically choses the median value of the pixel for filtration.

ALGORITHM

Adaptive Median Filter applies the noise detection and filtering algorithms to remove impulsive noise.
Let, $\boldsymbol{Iij}$ be the pixel of the corrupted image
$\boldsymbol{Imin}$ = minimum pixel value
$\boldsymbol{Imax}$ = maximum pixel value in the window
W = current window size applied $\boldsymbol{Wmax}$ = maximum window size that can be reached
$\boldsymbol{Imed}$ = median of the window assigned.
Then, the algorithm of this filtering technique completes in two levels as described:

- Level A:

a) If $I < I < I$, then the median value is not an impulse, so the algorithm goes to Level B to check if the current pixel is an impulse.
b) Else the size of the window is increased and Level A is repeated until the median value is not an impulse so the algorithm goes to Level B; or the maximum window size is reached, in which case the median value is assigned as the filtered image pixel value.

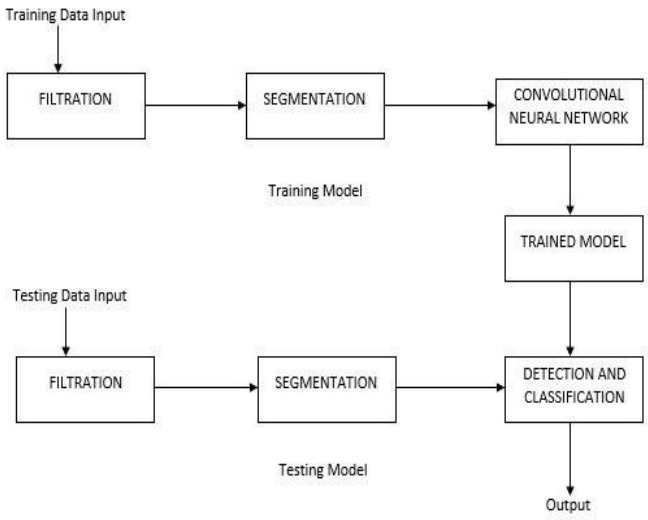
- Level B:

a) If $I < I < I$, then the current pixel value is not an impulse, so the filtered image pixel is unchanged.
b) Else the image pixel is either equal to I or I (corrupted), then the filtered imaged pixel is assigned the median value from Level A.

C. SEGMENTATION

Segmentation is the process of partitioning an image into multiple segments (set of pixels also known as super-pixels.) Segmentation is done to simplify and change the representation of image into something which is easier and more meaningful to analyze. Human Brain is made of various soft tissues (White matter, Gray Matter and Cerebos Spinal Fluid) tumor is also a kind of tissue which is not the part of brain. Image segmentation is one of the most important steps in preprocessing as it will

provide the area of interest on basis of which it will perform feature extraction. Segmentation is the most important process since the further training and testing of the model for its efficiency depends on it. Overview of the methodology is depicted below.



D. CONVOLUTIONAL NEURAL NETWORK

In mathematical terms convolutional is a function derived from two given functions by integration which expresses how the shape of one is modified by the other.

$$(f * g)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau)\, g(t-\tau)\, d\tau$$

Convolutional Neural Networks like Neural Networks are made up of neurons with learnable weights as well as biases. Each neuron receives multiple inputs and takes weighted sum over them passing it through an activation function and responds with an output. Convolutional Neural Networks consists of many convolutional layers and each layer have its activation function (Rectified Linear Unit). After Convolutional layers there are fully connected layers with activation function at the output. These fully connected layers have various parameters due to which it is prone to overfitting.

CNNs are basically used for the automation of models since it is capable of extracting features from datasets as well as training the model accordingly. On the basis of the pixel values the activation function provide output as an input to the next convolutional layer and in this manner all the values of the tumorous region get detected and it provides output as the detected Brain Tumor part. In case the Brain MR image is normal it will provide output as Tumor absent. Similarly, if the Tumor is detected it will use convolutional neural network to classify whether the Tumor is cancerous or non-cancerous. This model will work in two parts. First part is the training part in which large amount of dataset Tumorous Brain dataset is provided so that the model can understand and learn the pattern of the Tumor pixels present in the MR image. Similarly, for the classification we will provide dataset with two types of Tumor that is cancerous and non-cancerous. So that the model understands and learns the patterns of the cancerous and non-cancerous Tumor. After successful training we perform validation in which the obtained outcome will be compared with the ground truth received with the dataset to understand the accuracy rate of the model. The second part of the model consists of testing in which our model will be checked against large number of datasets for its efficiency. Testing part is done to ensure the reliability rate, accuracy and the efficiency of the model.

IV. CONCLUSION

The successful implementation of the project will accurately detect as well as classify the type of Tumor identified from MR image provided to the model. The accurate data received from the model can be counted as one of the success achieved in the field of medical against Tumor as it will help in detailed diagnosis of the Tumor without wasting much of the crucial time essential in the treatment of the patient.

REFERENCES

[1] Eman M, Mohammed E, Rashid A. Brain Tumor Segmentation based on a hybrid clustering technique. Egyptian Informatics Journal 2015 Production and hosting by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University; 71-81.

[2] Suman S. Image Denoising Using New Adaptive Based Median Filter. Signal & Image processing: An International Journal (SIPIJ) Vol.5, No.4, August 2014.

[3] Geethu M, M. Monica S. MRI Based medical image analysis: Survey on brain tumor grade classification. Published by Elsevier Biomedical Signal Processing and Control 39 (2018) 139–161.

[4] Mohammadreza S, Guang Y, Tryphon L, Nigel A, Timothy J, Thommas B, Franklyn H, Xujiong Y. Supervised Learning based Multimodal MRI Brain Tumour Segmentation using Texture Features from Supervoxels. Accepted Manual. To appear in: Computer Methods and Programs in Biomedicine.

[5] Sanjeev K, Chetna D, Sunila G. Classification of Brain MRI Tumor Images: A Hybrid Approach. Published by Elsevier B.V. Peer-review under responsibility of the scientific committee of the 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.

[6] Javeria A, Muhammad S, Mussarat Y, Steven F, A Distinctive Approach to Brain Tumor Detection and Classification using MRI, Pattern Recognition Letters (2017), doi: 10.1016/j.patrec.2017.10.036.

[7] Heba M, El-Sayed A. E, El-Sayed M. E, Abdel-Badeeh M. S. Classification using deep learning neural networks for brain tumors. Future Computing and Informatics Journal xx (2017) 1-4.

[8] Ali I, Cem D, Melike S. Review of MRI-based brain tumor image segmentation using deep learning methods. 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria. Published by Elsevier B.V.

[9] Rajesh C, Dr. Kosalani R. Tumor Detection in Brain using Genetic Algorithm. 7th International Conference on Communication, Computing and Virtualization 2016. Published by Elsevier B.V.

[10] Ivan C, Iker G. MRI Segmentation Fusion for Brain Tumor Detection, Information Fusion (2016), doi: 10.1016/j.inffus.2016.10.003.

[11] Swapnil T, Amit P, Ankita K. Detection of Brain Tumor from MRI Images by using Segmentation and SVM. 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (WCFTR'16).

[12] Sudipta R, Debnath B, Samir K, Tai-Hoon KIM. An Improved brain MR image binarization method as a preprocessing for abnormality detection and features extraction. Research Article. Accepted April 1, 2016.

[13] Shaswati R, Pradipta M. A Simple Skull Stripping Algorithm for Brain MRI. 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR).

[14] Manish Rana and Sunil Yadav. Protein: An Input for Medical Diagnosis of Brain Tumour. International Conference & Workshop on Emerging Trend in Technology (2014).

[15] SobriMuda and M. Mokji. Brain Lesion Segmentation from Diffusion Weighted MRI Based on Adaptive Thresholding and

Gray Level Co-occurrence Matrix. J. Telecommunication Electronic and Computer Engineering, Vol. 3, No. 2,2011.

[16] Reza P, Ying Z, Holly N, Robert M. Brain Tumor Segmentation in MRI Scans Using Deeply-Supervised Neural Networks. Radiation Oncology Branch, National Cancer Institute, National Institute of Health, Bethesda, MD 20184, USA. Springer International Publishing AG, part of Springer Nature 2018.

[17] Sara S. Brain Tumor Segmentation Using a Multi-path CNN Based Method. Springer International Publishing AG, part of Springer Nature 2018.

[18] M. Saritha and K. Paul J. Classification of MRI Brain Images using Combined Wavelet Entropy Based Spider Web Plot and Probablistic Neural Network. Elsevier, Pattern Recognition Letters 34, pp. 2151-2159, 2013.

[19] R. Muthukrishnan and M.Radha. Edge Detection Techniques for Image Segmentation. International Journal of Computer Science & Information Technology. Vol 3, No. 6. Dec 2011.

[20] K. Thapaliya and G.Kwon. Extraction of Brain Tumor based on Morphological Operations. In Proceedings in IEEE-8th International Conference on Computing Technology and Information Management. Pp. 515-520, 2012.

[21] S.H.Teoh et.al. Median Filtering Frameworks for Reducing Impulsive Noise from Grayscale Digital Images: A Literature Survey in Interantional Journal of Future Computer and Communication. Vol. 1, No.4, pp. 323-333 December 2012.

[22] J.Vijay and J.Subhashini. An Efficient Brain Tumor Detection Methodology Using K-Means Clustering Algorithm in International conference on Communication and Signal Processing. April 3-5, pp.653-658, 2013.

[23] M.U.Akram and A.Usman. Computer Aided System for Brain Tumor Detection and Segmentation in Proceedings in IEEE-International Conference on Computer Networks and Information Technology. Vol. 1. pp.299-302, 2011.

[24] Kim E.Y, Park S.H, Kim H.J. A genetic Algorithm-based segmentation of Markov random field modeled images. Signal Processing Letters. IEEE Digital Object Identifier 10. 1109/ 97.873564. Vol. 7, No. 11, pp.301-315, 2000.

[25] Philippe Andrey. Genetic Algorithms Applied to Image Segmentation. Image and Vision Computing. Vol. 17, pp.175-187, 1999.

[26] Esther A, Giles T, Stefano T, Marie B, Alexander V, Benedikt W, Claus Z, Bjoern H. Multi-modal Image Classification Uisng Low-Dimensional Texture Features for Genomic Brain Tumor Recognition. Springer International Publishing AG 2017.

[27] Navigating Life with a Brain Tumor. Lynne P. Taylor, By Alyx B. Porter Umphrey and With Diane Richard.

[28] Brain Tumors an Encyclopedia Approach 3rd Edition. By Andrew H. Kaye, Edward R. Laws. Jr.

[29] Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Γ, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694

[30] Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117

Low Budget Smart Glasses for Visually Impaired People

Harshita Bhorshetti[1] Shreyas Ghuge[2] Athang Kulkarni[3] Prof. Sukhada Bhingarkar[4] Prof. Dr.Netra Lokhande[5]
[1,3,4]*Computer Engineering*
[2,5]*IT Engineering*
MIT College of Engineering
Pune, India
[1]hbhorshetti@gmail.com [2]ghugeshreyas2@gmail.com[3]athang.kulkarni03@gmail.com [4]sukhada.bhingarkar@mitcoe.edu.in
[5]netra.lokhande@mitcoe.edu.in

***Abstract*—Smart glasses have wide applications including industrial, medical and healthcare, consumer, logistics, security, and government, etc. In practice, blind people identify objects or find out walkable paths using a white cane. Inspite of a variety of such canes (long cane, guide cane, support cane) available, it doesn't recognize what the actual object is. The guiding sticks give an idea about the objects that lie only in a limited area. Having a camera vision alongwith support of audio messages for the visually impaired gives them a better and more precise sense of their surroundings, especially in an outdoor environment. However, because of the high cost of such devices, the dependency of blind people on other people or traditional canes is consistent. This paper provides a list of existing glasses which are expensive because of their processor. This paperalso gives a brief idea about a new low cost design of Smart Glassesthat a user can afford and wear as normal glasses. This proposed design ofoptical head-mounted device integratescamera and allows hands-free operations. The camera placed on the device obtains the video, equivalent to the field of view which processes it for assisting the visually impaired user.**

Keywords—Computer Vision, Deep Learning, Image Processing, Machine Learning, Object Recognition, Virtual Reality

I. INTRODUCTION

The number of visually impaired people is constantly at a rise over the past few decades. According to the World Health Organization (WHO) [1] around 285 million people are visually impaired, out of which 39 million are completely blind. Irrespective of the cause, the outcome for these visually impaired people turns out to be a constantdependency on other sensory information [2] or people around them in order to avoid obstacles and to navigate. Taking into consideration the variations in their nature, some of them hesitate to ask for helpeventothepeopleclosetothem.Also,therearesituations when they are away from their trusted people and surrounded withstrangerswhichgivesrisetotrustissues.

To overcome these issues, recent advancements in technology have been channelized, leading to the development of wearable smart glasses [3]. These smart glasses have a wide range of applications and provide personal assistance along with guidance along obstacles to the visually impaired people, hence minimizing their dependency on others. The wearable device also facilitates hands-free experience enabling the user to perform multiple tasks simultaneously.

This paper aims at developing such smart guiding glasses but in an extremely cost-effective manner, making itaffordable to low-income people as well. It focuses on smoothing the outdoor experience of the visually impaired by efficiently detecting objects and hurdles from the surrounding. The us ability can be expanded by adding on more module.

II. .EXISTING ECHNOLOGIES

Realizing the need of the hour, various companies and organizations have also contributed by developing different kinds of smart glasses for these visually impaired people [3]. Some of the existing smart glasses are OrCam, NuEyes, Esight,etc.

OrCam [4], an Israeli start-up company whose mission was to develop a portable, wearable, visual system using artificial intelligence and augmented reality, has designed a device which recognizes text and objects pointed towards it. It also provides assistance through an earpiece.

NuEyes e2 [5] is the most recently developed lightweight, wearable electronic magnifier for low vision. Its main purpose is magnifying the objects it points at, thus enabling a better vision.

Esight eyewear [6] uses technology that helps visuallyimpaired people to actually see the world around them, in real time. Advanced, medically-validated algorithms optimize and enhance the footage captured by high speed, high definition camera.

AIRA [7]smart glasses also use a camera and connectivity with their agents to bring assistance to people with a visual impairment. In this case, however, what you're connected to is a trained assistant who provides spoken feedback about what you are looking at.

The following table gives a better understanding of functionalities and drawbacks of Existing Smart Glasses and compares them with the proposed system mentioned in this paper.

TABLE I. EXISTING SMART GLASSES

Sr. No.	Product Name	Functionality	Drawbacks

1.	OrCam	Text Recognition &Object Recognition	i. High Price ($2,500) ii. Output only in English
2.	NuEyes	Magnifying objects for better vision	High Price ($2,750)
3.	Esight	Enables real time vision of surroundings	High Price ($9,995)
4.	AIRA	Assistance through busy roads or indoor environment, reading and identifying objects	i. Monthly Pack of $89 for 100 minutes. ii. Dependency on connectivity with AIRA agents.
5.	Proposed Design	Object recognition and assistance in outdoor environment at low price($100)	Limited applications in initial stage.

III. III.PROPOSEDSYSTEM

In this paper, we provide a solution to overcome the deficits of the above existing methodologies or devices to help the visually impaired people. Smart Glasses not only detect daily objects dynamically but also give out audio outputs to the visually impaired people so as to notify them about their surroundings in real time.

The Smart Glasses developed by the design proposed in this paper is a wearable device which works on the basis of computer vision, machine learning and image processing. These glasses are to be worn by the user, which are connected to a credit card sized processor — Raspberry Pi Zero, speaker and a camera module driven by rechargeable batteries for power requirements. The object detection model must be trained accurately so as to reap the proper benefits of a small microprocessor. Using a fast object detection algorithm such as YOLO v3 [8] makes the model work at a greater speed and thus gives the real-time object or surrounding information to the user. For effective training of model on a large dataset, MicrosoftCOCOdataset[9]wasused.

Training a machine learning model is a time-consuming and tedious job when it comes to such large datasets to train on. As a result, the training part of the model was extended to a GPU with higher capacity to run through the algorithm quicker than the conventional CPUs and then transferring the trained model files to the memory of a micro-controller. YOLO algorithm is then run on the dataset with means of Jetson TK1 GPU which takes much less time to go through the training dataset.

Once the model is trained, the required files are then transferred on to a Raspberry Pi Zero micro-controller which has a handy micro SD card slot capable of holding the memory card. This processor runs at 1 GHz with the ARM instruction set. After the software model is ready and installed on Raspberry Pi then it is connected with a camera module compatible with the processor which is a 5MP shooter with fixed focus. This camera module is mounted on the top middle of the glasses. It gives a better and accurate video of the view the user is trying to get. The camera module operates with 720p @60fps which degraded to about 10-15fps because of the light weighted processor and simultaneous algorithm execution. The video captured by the camera is processed by the Raspberry Pi Zero which runs the YOLO v3 algorithm for every 5thframe itreceives.

Themodeltrainedandtestedfordetectingobjectsinrealtime i.e. YOLO v3, output the textual format of the intended objects found in that particular frame of the image. As this device is built for the visually impaired people they can be assisted by giving an audio output rather than a non-useful textual format. The recognized objectsare thus conveyedout by a speaker which is attached to Raspberry PiZero. The audio signals needed are generated using gTTs library compatible with Python2 and above to convert text to speech. Thewholedeviceeasilyworksonrechargeablebatteries which can provide sufficient power to run a day long. The Raspberry Pi Zero requires 80mA (0.4W) power [10] which is far lessthan other processors will need. This makes Smart Glasses cost effective and at the same time more accurate andefficient.

A. SystemArchitecture

System Architecture of the system describes the important building blocks of the system, one tries to build. Smart Glasses basically has 4 blocks which should be considered while in the architecture. The user interacts with Smart Glasses which in turn triggers the processor which controls the camera, speaker, and memory/database.

Major building blocks of the device's architecture include:

- Processor
- Speaker
- Camera
- Database Memory

The following figure represents the architecture of the smart glasses proposed in this paper. It shows the interaction between the major building blocks and the functions each block performs.

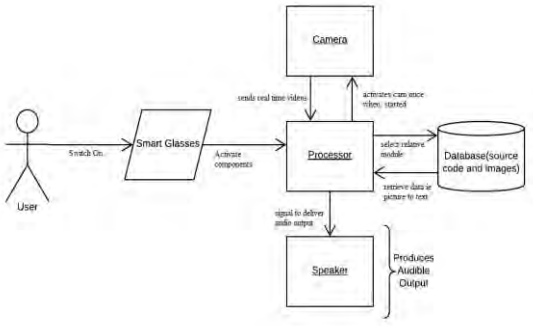


Fig 1. Architecture Diagram

B. Working

Themainfunctionscarriedoutbythesystemarehighlighted as we break into its sub processes. The below figure shows the entities and processes they need to follow in order to makethe device work according to its functionality. Rectangular boxes define the entities, the rectangular boxes with curved edges represent theprocesses.

- Turning on the device Switch On
- Scanvideointheangleofvision
- Detect obstaclesObjects
- VideoAnalysis

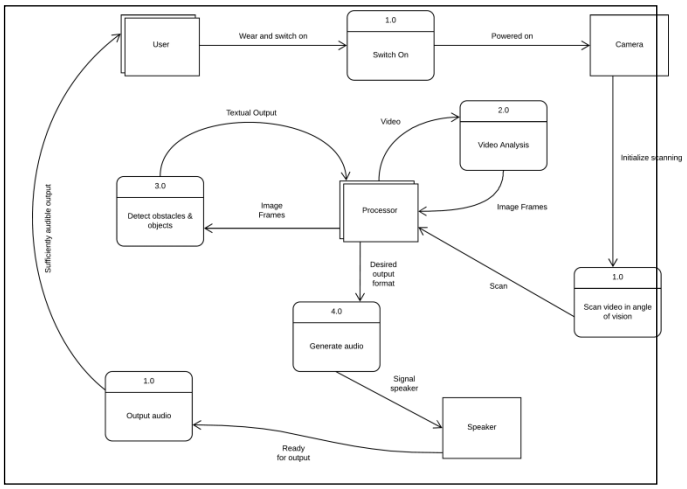


Fig 2. DFD - Working of System

The overall working of the device is shown easily by using data flow diagram given above. Working starts as soon as the userwearandswitchesonthepowersupplyforSmartGlasses. The first component that initiates in the device system with its job, is the Raspberry Pi Zero processor. It further initiates the camera module attached along with the speaker. The circuit is thusfullyrunningonceallthepartareoperative.

The scanning part is done by the camera module of all the surrounding in its angle of vision. Scanning is possible in only sufficientlywellilluminatedenvironmentswhereinthecamera easily captures the area circumambient to the user. The video is continuously sent to the processor for processing. Processor performs video analysis by selecting roughly every 3rdto 5thframe from the video so as to not output the same surrounding again.

Subsequently, every select image frame goes through the trained YOLO model in order to determine the objects at that particular instance. The textual output produced by YOLO algorithm is then converted into speech and speakers are signaled to manufacture its audio version.

Figure 3 describes the process of selecting the frames from live videos and develop the basis for the algorithm to run on.

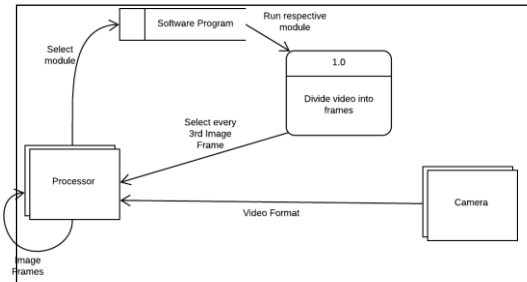


Fig 3. Frame Extraction

Figure 4 sheds light on how the selected frames run through the YOLO algorithm and how finally audio signals are produced for the speakers attached to the device.

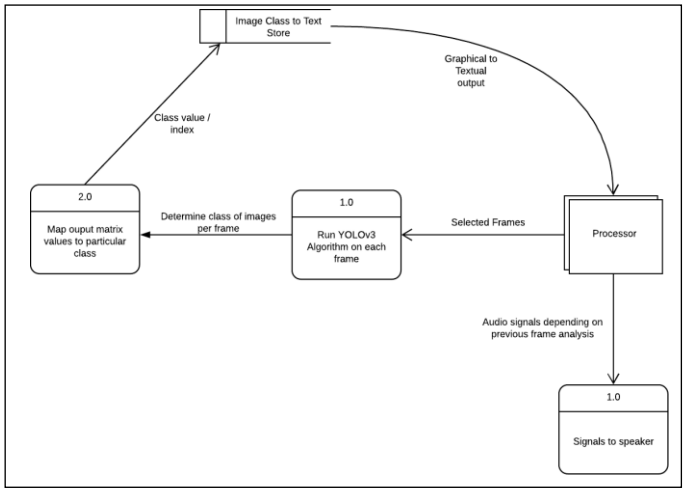


Fig 4. Generation of output signal

TABLE II. HARDWARE REQUIREMENTS

Hardware Requirements	Specification	Functionalities
Processor 1 : Jetson TK1	DRAM 2GB DDR3L Storage - 16 GB Ports - USB 3.0, USB 2.0, HDMI, RS 232 Camera Ports: 2 fast CSI-2 MIPI camera ports (one 4-lane and one 1-lane) APIs Supported: CUDA 6.0, OpenGL 4.4, OpenCV4Tegra	Being a GPU, this processor has tremendous power for embedded system applications. In this project, Jetson TK1 was used to train the object detection model and to make the computations faster than using a conventional CPU.
Processor 2 : Raspberry Pi Zero	1GHz Single-core CPU 12MB RAM, Mini HDMI and USB On-The-Go ports	To replace Jetson TK1 once the model is trained and to run real time analysis using the trained object detection model.

	CSI camera connector Supported Video Formats: 1080p @ 30fps, 720p @ 60fps and 640x480p 60/90 video	
Camera : 5MP Raspberry Pi Zero W Cam- era Module	Fully Compatible with Raspberry Pi Zero W. Small and lightweight camera module.	Continuously capture real time videos once the Smart glasses are switched on.
Speaker : Pimoroni pHAT Automatic	3 x 12-bit ADC @ 0-24V (ADS1015) (2% accuracy), 3 x 24V tolerant buffered inputs, 3 x 24V tolerant sinking outputs, Python library available. library	To output the audio corresponding to the textual data generated for objects.

IV. EXPERIMENTAL SETUP

The proposed design of smart glasses has a set of hardware and software requirements for the purpose of smooth and efficient working. Table 2below represents these requirements along with otherspecification.

V. CONCLUSION AND FUTUREWORK

Smart Glasses for visually impaired people will provide the required assistance to both totally blind as well as partially blind people. Providing real-time object detection capability with audio output will allow users to get an idea of their surroundings with greater precision than using blind man's cane. People will also be able to find free path along the directions they are walking towards. Using cheap but effective processor such as Raspberry pi Zero will, in turn, reduce the total cost.

The functionality of these glasses can be widened using additional modules. They can also be used for personal navigation by training them over the user's daily routes. They can also be trained over the user's family and friend's faces to help the user recognize who he is talking to or come across.

ACKNOWLEDGMENT

Special thanks goes to our guide, Prof. Sukhada Bhingarkar and Prof.Dr. Netra Lokhande for their constant guidance and support in our project. Their keen observation and experiences, along with valuable guidelines and suggestions helped us in accomplishing this particular task.Their advices really played a significant role in making our knowledge and this project presentable.

REFERENCES

[1]. J. Bai, "ShiguoLian, Smart Guiding Glasses for the Visually Impaired People in Indoor Environment from," IEEE Transactions on Consumer Electronics Vol, vol. 63, no. 3, 8 2017.

[2]. P. U. Thakare, "Kote Shubham, Pawale Ankit, Rajguru Ajinkya, Shelke Om, "Smart Assistance System for the Visually Impaired " from Inter- national," Journal of Scientific and Research Publications, vol. Volume, 12 2017, 7, Issue 12.

[3]. E. A. H. T. B. Tang, Smart Glasses for the Visually Impaired People from book Computers Helping People with Special Needs, 7 2016, 13 15. (pp 579 - 582)

[4]. [Online].Available: https://www.visionaware.org/blog/visionaware-blog/ the-orcam-a-portable-wearable-visual-system-for-blind-and-visually- impaired-persons/12

[5]. [Online] Available: https://nueyes.com/products/nueyes-e2

[6]. [Online].Available: https://www.esighteyewear.com/technology

[7]. [Online] Available: https://aira.io

[8]. J. Redmon and A. Farhadi, YOLOv3: An Incremental Improvement. University of Washington.

[9]. "Microsoft COCO dataset," http://www.cocodataset.org.

[10]. [Online].Available: https://www.jeffgeerling.com/blogs/jeffgeerling/ raspberry-pi

Fire Extinguisher Using Acoustic Waves

Jagruti Bhagwan Durgude, Pooja kisan Aher, Pooja Somanath Avhad

Electronics & Telecommunication, NDMVPS'S KBTCOE, NASHIK
j.durgude28@gmail.com, poojaaher137102@gmail.com, poojavhad@gmail.com

***Abstract*— Existing fire extinguishers contains different chemicals, depending upon their applications. Generally they are pressurized with nitrogen or carbon dioxide (CO2), when this pressure released on fire, it will extinguish the fire. As we know there is many such firefighting agents such as water, potassium bicarbonate etc. all these agents have same property of leaving an unproductive system behind it. To deal with fire we need to have complete information on fire and its working in space. The formation of fire requires three elements in proper mixture they are fuel, oxygen and heating element. The project aims to develop the potable fire extinguisher study and analyzes the effect of different frequency of sound wave on flames. This project utilized the scientific principle of physics and the engineering aspects of electronics to successfully suppress a flame. The goal of this project is to use acoustic sound waves to extinguish and suppress a flame. Traditional fire extinguishers, such as chemical foam or water, are used successfully but pose the threat of severely damaging indoor equipment, whereas an acoustic wave fire extinguisher would protect them from further damage caused by the fire. The extinguishing device developed is intended for typical residential and commercial use.**

Key words — Acoustic waves, Atmega328, HC05, Power supply, key.

I. INTRODUCTION

Current method of firefighting technology has short come such as toxic to humans and leaves residues (for dry chemical-based fire extinguisher) while water-based fire extinguishing technique freezes in cold climates and conduct electricity. Sound wave could be one of the potential alternatives as fire extinguisher. The present paper aims to develop the portable fire extinguishing robot using sound waves. Current method of firefighting technology has short come such as toxic to humans and leaves residues (for dry chemical-based fire extinguisher) while water-based fire extinguishing technique freezes in cold climates and conduct electricity. Sound wave could be one of the potential alternatives as fire extinguisher. The present project aims to develop the portable fire extinguishing robot using sound waves. The need of fire extinguishing techniques is vital as fire accidents are catastrophic in nature, leads to unrecoverable loss. The present existing techniques are not eco-friendly. Sound wave could be one of the potential alternatives as fire extinguishers. The low frequency acoustic waves spilled from a speaker tend to extinguish the flames. The project aims to develop the potable fire extinguisher study and analyzes the effect of different frequency of sound wave on flames. Experiments are conducted to study suitable sound wave frequency range to extinguish flame. Sound is pressure wave and displacement caused in the medium through with particles will move in a random direction, and transferring the pressure from one particle to another, hence this how sound travel in any medium. Sound can be travel in two forms they are

1) Longitudinal waves: Longitudinal waves, also known as "l waves", are waves in which the displacement of the medium is in the same direction as, or the opposite direction to, the direction of travel of the wave. Mechanical longitudinal waves are also called compression waves, because they produce compression and rarefaction when traveling through a medium.

2) Transverse Wave: A transverse wave is a moving wave that consists of oscillations occurring perpendicular (or right angled) to the direction of energy transfer. If a transverse wave is moving in the positive x-direction, its oscillations are in up and down directions that lie in the

Y–z plane. Light is an example of a transverse wave. With regard to transverse waves in matter, the displacement of the medium is perpendicular to the direction of propagation of the wave. A ripple in a pond and a wave on a string are easily visualized as transverse waves.

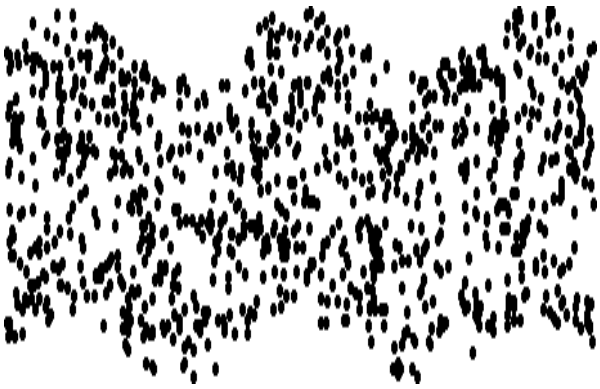


Fig.1 Image of transverse wave particles

II. MOTIVATION

The project demonstrates idea of a developing device which can extinguish fire using sound. Most of the things when compared to space are different from earth one of such is fire. This is major threat in an International space station or in any other space ship. If a fire breaks out in space station astronauts would fight the flames in slightly different way than they would on earth. First, they will turn off the ventilation system to slow the spread of fire, next they will shut off the power supply to the effected unit. Finally, astronauts will seal the area and fill the area with a chemical extinguisher. Extinguisher used in this process may extinguish the fire and will also ravage the total system in that particular area. So the purposed system deals with fire without affecting any other system around it and avoids all messy chemical, therefore resulting in an electronic-extinguisher.

III. ANALYSIS & DESIGN APPROACH

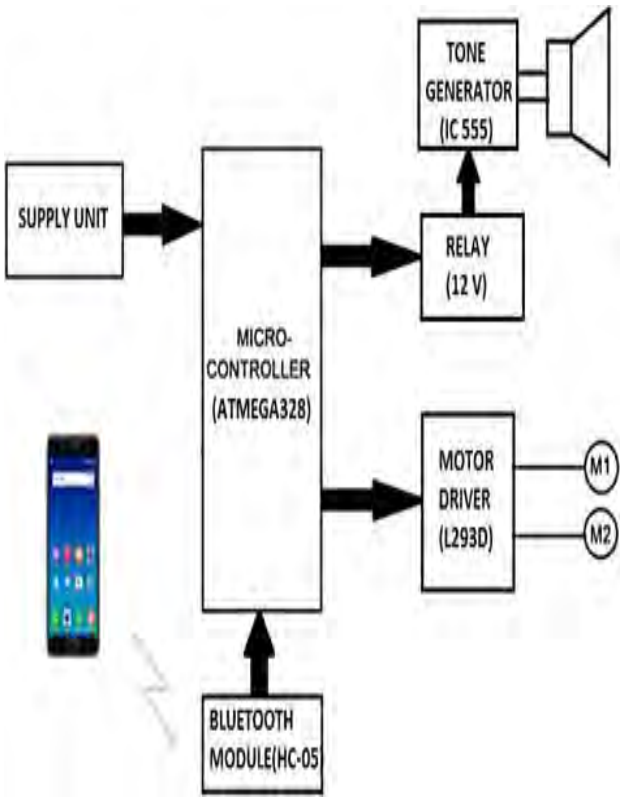


Fig 2. Block Diagram of System

A. POWER UNIT

Power unit is used to provide supply voltage to microcontroller and other circuit. Here we are using 12V battery as a supply and as microcontroller needs 5V to operate so we are connected regulated integrated circuit 7805.

a. 7805 Regulator Features

1) 5V Positive Voltage Regulator
2) Minimum Input Voltage is 7V
3) Maximum Input Voltage is 25V
4) Operating current(IQ) is 5mA
5) Internal Thermal Overload and Short circuit current limiting protection is available.
6) Junction Temperature maximum 125 degree Celsius
7) Available in TO-220 and KTE package

B. RELAY (12V)

Relay is an electromagnetic device which is used to isolate two circuits electrically and connect them magnetically. They are very useful devices and allow one circuit to switch another one while they are completely separate. They are often used to interface an electronic circuit (working at a low voltage) to an electrical circuit which works at very high voltage. For example, a relay can make a 5V DC battery circuit to switch a 230V AC mains circuit. Thus a small sensor circuit can drive, say, a fan or an electric bulb.

a. Relays and transistors

Like relays, transistors can be used as an electrically operated switch. For switching small DC currents (< 1A) at low voltage they are usually a better choice than a relay. However transistors cannot switch AC or high voltages (such as mains electricity) and they are not usually a good choice for switching large currents (> 5A). In these cases a relay will be needed, but note that a low power transistor may still be needed to switch the current for the relay's coil.

Advantages of relays are

1) Relays can switch AC and DC, transistors can only switch DC.
2) Relays can switch high voltages, transistors cannot.
3) Relays are a better choice for switching large currents (> 5A).
4) Relays can switch many contacts at once.
5)

C. FREQUENCY GENERATOR (IC 555)

Used to generate the acoustic waves in order to extinguish the fire. Here for generation of frequency we are using IC 555. The 555 timer is a chip that can be used to create pulses of various durations, to output a continuous pulse waveform of adjustable pulse width and frequency, and to toggle between high and low states in response to inputs. By wiring the 555 timer with resistors and capacitors in various ways, you can get it to operate in three different modes. In this project it is operated in astable mode. In astable mode, the output from the 555 timer is a continuous pulse waveform of a specific frequency that depends on the values of the two resistors (R1 and R2) and capacitor (C1) used in the circuit (fig 1) according to the equation below. Astable mode is closely related to monostable mode, you can see that the schematic is nearly the same. The important difference is that in astable mode, the trigger pin is connected to the threshold pin; this causes the output to continuously toggle between the high and low states.

Output Frequency of IC 555 = $1/[0.7*(R_1+2*R_2)*C_1]$

a. 555 Timer IC Features

1) Typical operating voltage is +5V can withstand a maximum of +18V.
2) The source/sink current of the output pin is 200mA.
3) Consumes up to 3mA when operating at +5V.
4) Trigger Voltage is 1.67 when operating at +5V.
5) Operating Temperature is 70 degree Celsius.
6) Available in 8-pin PDIP, SOIC, and VSSOP packages.

D. MOTOR DRIVER (L293D)

L293D is a typical Motor driver or Motor Driver IC which allows DC motor to drive on either direction. L293D is a 16-pin IC which can control a set of two DC motors simultaneously in any direction. It means that you can control two DC motor with a single L293D IC. Dual H-bridge Motor Driver integrated circuit (IC).The L293D can drive small and quiet big motors as well, check the Voltage Specification at the end of this page for more info. It works on the concept of H-bridge. H-bridge is a circuit which allows the voltage to be flown in either direction. As you know voltage need to change its direction for being able to rotate the motor in clockwise or anticlockwise direction, hence H-bridge IC are ideal for driving a DC motor.

a. Features

7) Can be used to run Two DC motors with the same IC.
8) Speed and Direction control is possible.
9) Motor voltage Vcc2 (Vs): 4.5V to 36V.
10) Maximum Peak motor current: 1.2A.
11) Maximum Continuous Motor Current: 600mA.
12) Supply Voltage to Vcc1 (vss): 4.5V to 7V.
13) Transition time: 300ns (at 5Vand 24V).
14) Automatic Thermal shutdown is available.
15) Available in 16-pin DIP, TSSOP, SOIC packages.

E. BLUETOOTH-MODULE (HC-05)

HC05 module is an easy to use Bluetooth SPP (Serial Port Protocol) module, designed for transparent wireless serial connection setup. Serial port Bluetooth module is fully qualified Bluetooth V2.0+EDR (Enhanced Data Rate) 3Mbps Modulation with complete 2.4GHz radio transceiver and baseband. It uses CSR Blue core 04 External single chip Bluetooth systems with CMOS technology and with AFH. It has the footprint as small as 12.7mmx27mm.In this paper we are using it for controlling our fire extinguisher by mobile phone. The HC-05 has two operating modes, one is the Data mode in which it can send and receive data from other Bluetooth devices and the other is the AT Command mode where the default device settings can be changed. We can operate the device in either of these two modes by using the key pin. HC05 Bluetooth Module used in many applications like Wireless communication between two microcontrollers, Communicate with Laptop, Desktops and mobile phones, Data Logging application, Consumer applications, Wireless Robots, Home Automation etc.

a. HC-05 Technical Specifications

1) Serial Bluetooth module for Arduino and other microcontrollers.
2) Operating Voltage: 4V to 6V (Typically +5V).
3) Operating Current: 30mA.
4) Range: <100m.
5) Works with Serial communication (USART) and TTL compatible.
6) Follows IEEE 802.15.1 standardized protocol.
7) Uses Frequency-Hopping Spread spectrum (FHSS).
8) Can operate in Master, Slave or Master/Slave mode.
9) Can be easily interfaced with Laptop or Mobile phones with Bluetooth. Supported baud rate: 9600,19200,38400,57600,115200,230400,460800

F. MICROCONTROLLER (ATMEGA328)

ATmega-328 is basically an Advanced Virtual RISC (AVR) micro-controller. It supports the data up to eight bits. ATmega-328 has 32KB internal built-in memory. It is a programmable device which takes decision and provides command to further circuitry according to the input condition.

a. Features

1) ATmega328 is an eight (8) bit Microcontroller
2) Built in internal memory is around 32KB
3) ATmega 328 has 1KB Electrically Erasable Programmable Read Only Memory (EEPROM)
4) ATmega-328 has 2KB Static Random Access Memory (SRAM).
5) It operates ranging from 3.3V to 5V
6) The cost efficiency
7) Low power dissipation
8) Programming lock for security purposes
9) Real timer counter with separate oscillator

G. AMPLIFICATION BLOCK

These blocks constrain the main operation of the device by controlling the gain and intensity of sound. This block constructs the wave which can be fed to the speaker as output. In electronics, signal conditioning means manipulating an analog signal in such a way that it meets the requirements of the next stage for further processing. Signal inputs accepted by signal conditioners include DC voltage and current, AC voltage and current, frequency and electric charge. Sensor inputs can be accelerometer, thermocouple, thermistor, resistance thermometer, strain gauge or bridge, and LVDT or RVDT. Specialized inputs include encoder, counter or tachometer, timer or clock, relay or switch, and other specialized inputs. Outputs for signal conditioning equipment can be voltage, current, frequency, timer or counter, relay, resistance or potentiometer, and other specialized outputs. Signal conditioning can include amplification, filtering, converting, range matching, isolation and any other processes required to make sensor output suitable for processing after conditioning. In this paper we are going to perform amplification in order to make signal strong enough to extinguish the fire.

Signal amplification performs two important functions: increases the resolution of the input signal, and increases its signal-to-noise ratio. For example, the output of an electronic temperature sensor, which is probably in the millivolts range is probably too low for an analog-to-digital converter (ADC) to process directly. In this case it is necessary to bring the voltage level up to that required by the ADC. Commonly used amplifiers used for signal conditioning include sample and hold amplifiers, peak detectors, log amplifiers, antilog amplifiers, instrumentation amplifiers and programmable gain amplifiers.

H. SPEAKER

A device containing a transducer that converts electrical signal (electric current) into sound waves (acoustic energy) for the production of sound. Speaker works on the Flemings left hand rule. The magnet in the speaker is surrounded by a voice coil which is wounded copper, both in a perpendicular direction. The electric signal is passed through the voice coil where it is induced by a magnetic field which creates a force to happen in the third direction.

a. Specification

1) Woofer Unit 30Hz – 30kHz Effective Frequency Range
2) 250 / 50 WRMS Power
3) 90dB Sensitivity
4) Normal black cone
5) 86dB ± 2dB
6) Copper VC wire type
7) Ferrite Magnet
8) Tweeter Unit 5 – 20 KHz Effective Frequency Range
9) PEI diaphragm Material
10) ND-Fe-B Magnet material

IV. CIRCUIT DIAGRAM OF FREQUENCY GENERATOR

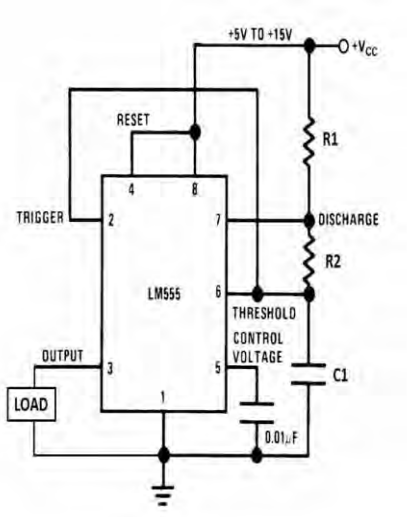


Fig 3. Frequency generator using IC555

FLOW CHART

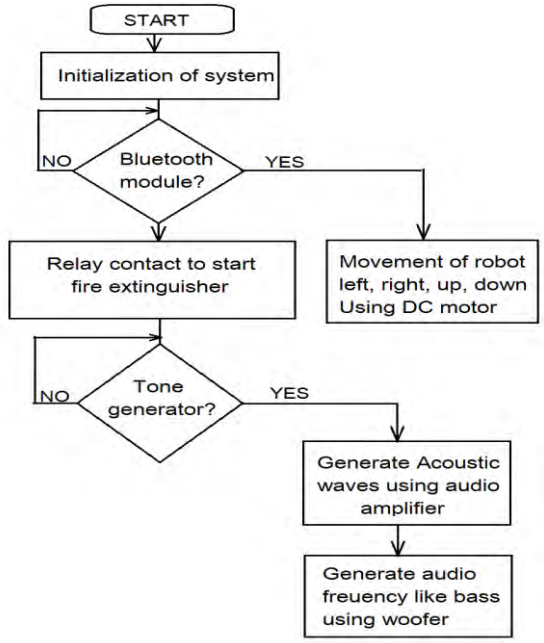


Fig.4 Flow chart

A. ADVANTAGES

1) Avoid wastage of water.
2) Useful in hazardous conditions without human harm.
3) Use of carbon dioxide is avoided.

B. APPLICATION

1) Used in small fires and can be used in home caused fires.
2) Generally, when fire is caused in an electrical panel circuit, using water is not possible as water conducts electricity, thus using sound waves to extinguish fire is one of its main application.
3) It can be used in place of chemical fire extinguishers and water as sound wave fire extinguisher leaves no residue and it is nontoxic to the person and the environment.

V. CONCLUSION

The idea of Sound Fire Extinguisher Robot controlled by Bluetooth potentially serves as an alternative to traditional fire extinguishers. For instance, install in every electrical panel, can be mounted subwoofer on a dedicated circuit, designed to turn on whenever fire is detected. It can be programmed to alternate the frequency based on a width of the flame. So it can be possibly used for a larger area. Without risking human life or limb, robots can replace humans in some hazardously duty service. Robots can work in all types of polluted environments, chemical as well as nuclear. They can work in environments so hazardous that an unprotected human would quickly die.

ACKNOWLEDGMENT

We thankfully acknowledge Electrowill Technologies, Nashik for providing various resources such as laboratories with all needed software platforms for our project and special thanks to Prof.V.P.Gawai for giving us all the help and guidance we needed. Vikas P. Gawai presently working as a Assistant professor at N.D.M.V.P.S's, College Of Engineering, Nashik in the Department of Electronics and Telecommunication Engineering. Jagruti B. Durgude B.E (Electronics and Telecommunication) from N.D.M.V.P.S's, College Of Engineering, Nashik. Pooja s. Avhad B.E (Electronics and Telecommunication) from N.D.M.V.P.S's, College Of Engineering, Nashik. Pooja K. Aher B.E (Electronics and Telecommunication) from N.D.M.V.P.S's, College Of Engineering, Nashik

REFERENCES

[1] To Extinguish a Hot Flame, DARPA Studied Cold Plasma. (2012, July 12). (DARPA) Retrieved March 12, 2015, from http://www.darpa.mil/newsevents/releases/2012/07/12.aspx
[2] Acoustic Fundamentals. (n.d.). ANSYS Mechanical APDL Theory Reference, Chapter 8: Acoustics. Retrieved November 30, 2014, from https://support.ansys.com/portal/site/AnsysCustomerPortal/ 9.
[3] Thomas D. Rossing and Neville H. Fletcher, Principles of Vibration and Sound Second Edition. Springer-Verlag New York, 2004
[4] T. S. Call and D. B. Schwartz, "Electric Fields For Flame Extinguisher," HQAFCESA/RACF, no. ESL-TR-91-54, 1991.
[5] National Semiconductor, "LM386 Low Voltage Audio Power Amplifier," no. August, pp. 1–9, 2000.
[6] Hood, C., & Frendi, A. (2005, June). On the Interaction of a Premixed Flame with an Acoustic Disturbance. In 41 st AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit (pp. 1-10).
[7] http://www.extremetech.com/extreme/132859-DARPA-Creates-sound-based-fire-extinguisher.

Restaurant Review Classification and Recommendation System using SVM

[1]Veda Waikul, [2]Onkar Ravgan, [3]Prof. Aruna Pavate
Computer Engineering Atharva College of Engineering Mumbai, India
[1]vedawaikul21@gmail.com, [2] onkar.10r@gmail.com, [3]arunaapavate@gmail.com

***Abstract*— Whenever people go to any restaurant, approximately 1-2 hours are spent by them at the restaurant depending on their order. They have to order the food and then wait for the 15-20 minutes for the arrival of the food. Wrong order can be taken by the waiter, which can waste the customer's time and money as well as the customer doesn't get satisfied by such food. This process becomes tedious if someone is in great hurry and can't waste so much of time. It becomes worse if there is a waiting queue. The high failure rate of new restaurants- estimates including two out of three restaurants close in the first year, or three out of four close in the first two years, and so on and so on.**

I. INTRODUCTION

In today's world technology and automation in every sector is rapidly increasing. People rely more on mobile devices for almost every task in their day to day lives. Restaurant Business is a sector which has a very large scope in automation and use of technology. At such times waiting for the waiter to take orders, delivering food, lengthy queues, etc. can be very displeasing for the customers of the restaurant.

To overcome these problems a concept of automated restaurants using a system which uses LCD displays, mobile/tablet devices and a system for the chef to interact with customers is proposed. Using Machine Learning and Data Science predictions are made based on the reviews and other data of the customer can help make the dining experience better as well as it will help the restaurant to manage and make the restaurant grow.

In a restaurant while placing order, the customer has to ask the waiter whether a particular food item is available or not and after that he/she has to give the order. As well as several times it happens that customers have to wait for the waiter to come to their table which is sometimes frustrating.

Storing the statistical data of the restaurant is a very tedious task. There is need of managing the data of inventory, customer orders and reviews, staff, payroll.

I. THEORY

(i) Research Foundation

The research began with the idea of automating the entire process of restaurant dining and help the customers to enjoy the dining and also to make the restaurants a better and user friendly place to eat and have fun with everyone. The restaurant should also be benefitted from this, which will be done using generating weekly statistics and data which will help the restaurant grow.

(i) Research Approach

The Research will be performed with people interacting with the system. The reviews given by the customers of the restaurant will help to train the model and give better results with every training process. By analyzing and calculating different performance measures and statistics the system will be made better and it will be more useful.

(ii) Technology and Methodology to be adopted

The System is implemented using SDLC Spiral Model as shown in Figure 1. The Spiral Model is selected because the system to be implemented will require flexibility and it will be implemented in different phases. The SDLC spiral models provides with all the necessary phases and processes to implement the system.

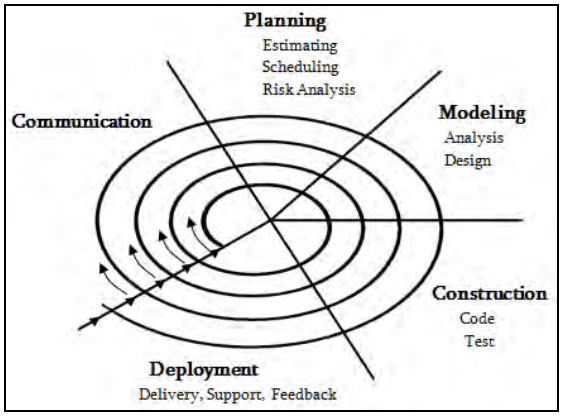


Figure 1: SDLC Spiral Model

A. Communication

From the customers and the users, who will be using the system in a restaurant information was collected from them. The information collected includes the limitations of traditional system, how people want the system to be designed along with different features for the system.

Communication was done with staff of the restaurant and asked about the details of the system they would like to use.

B. Planning

a. Estimating
- The hardware required for the system include mobile / tablet device, a Desktop PC, LCD displays. The total cost estimation for all these will be Rs.20000.
- The software required for the system are java jdk/jre, Android Studio, Scikit-Learn, Python. Softwares which will be used are freeware so no software cost will be required.

b. Scheduling
- The detailed schedule of all the phases is shown in figure 2.
- Initial planning, literature review, requirement gathering and analysis will be done till mid-September 2018. Post planning, the design will be completed till October 2018. Deployment of the project will be done in next two months. Testing and implementation will

be done in February 2019. The final report will be prepared in the next month.

c. Risk Analysis
- One of the major risks is software is handled in a wrong way. Deletion of the customer's data will create a major problem as the system will not have any data to process and the wrong or no preference will be given to the customer and every time system will ask for the likes to the customers. Thus, the proper knowledge of how to manage the system will be given to the restaurant owner/manager.
- Failure of any of the interface may lead to the whole system to crash. As well as the customer's data should be secured. Thus, regular security checks and maintaining the system is necessary.
- Though having the preferred food items at a single click is comfortable but due to the similar orders the sales of the restaurant may flow only in a particular direction.

C. Modelling

a. Analysis

Analysis is done based on how to make user interface more attractive and user-friendly. Also study and analysis of SVM algorithm is done on how to make the model more efficient. The data through which the model will be trained will be analysed.

b. Design

Design of the UI of Android and Web Interfaces will be done. Parameters required to train the machine learning model using SVM will be decided and experimented upon.

D. Construction

a. Code

Coding of Android application for the customer interface and web application for the staff interface will be done. The model will be created using SVM algorithm. Training data will be provided the algorithm to tune the model which will be used for recommending food and classification of review.

b. Test

Testing will be done by making the users use the system and then based on their reviews necessary changes will be made and also error checking will be done.

E. Deployment

a. Delivery

After implementing the changes from the testing phase the system will be deployed and will be implemented in a restaurant or café.

b. Support

If any issues or technical difficulties arise proper on-site support will be provided. Also regular maintenance will be provided.

c. Feedback

Feedback from the customers and staff of the restaurant will be taken frequently and then implemented upon.

II. RESEARCH/ TECHNICAL/ PROJECT WORK

A. Survey

The data used for the Machine Learning model to train was taken from "www.kaggle.com restaurant reviews data". The data had eighty thousand reviews given by the customers which consisted of the degree of positivity and negativity of the review along with review in text format.

B. Modeling/ Simulation

The Machine learning model will use SVM (Support Vector Machine) as the primary algorithm for classifying the restaurant reviews. This model will be deployed on a server. A central Database will be maintained which will store all the data of the restaurant on it.

The full system will be tested and simulated on a restaurant. Where each table of the restaurant will have a tablet computer with a custom interface to place orders. The incoming orders will be displayed on the chef's interface with real time information of each and every order. The customer will also be updated on their orders on their tablet interface. The tablet computer will also have entertainment options for the customers.

There will be admin panel which will have all the data about the restaurant and their employees, menus and the statistical data of the restaurant. The admin will be able to manage all the restaurant's working and transactions through the admin panel.

III. PROPOSED SYSTEM

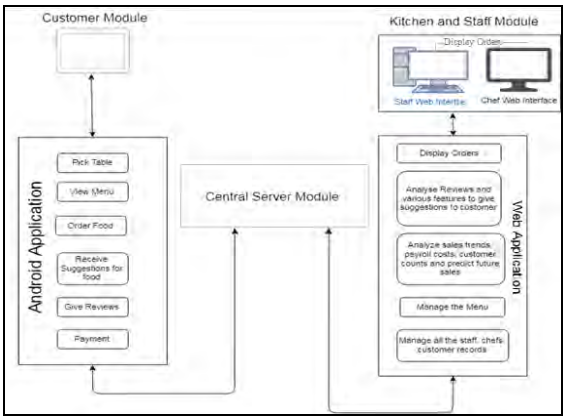


Figure 2: Block Diagram of Proposed System

Figure 2 shows the block diagram of the Proposed System. The proposed system consists of Customer Module, Central Server Module and Kitchen and Staff Module. The functionalities and details of these modules are described below.

A. Customer Module (CM)

- Customer will be given an interface which will be either a mobile/tablet device or a LCD screen interface.
- Predictions and suggestions about the food which a person might want to eat according to various features will be made so that customer can get what he/she wants easily.
- Using Time-Series predictions application will suggest approximate time at which the customer can book the table and availability of the table.
- This makes sure that the customer is sitting on the table and then automatically the device will send the information to the CSM.
- After that customer can place the order through the app on their mobile device.
- Customer will get step by step details about the order.
- After the food is eaten the payment can be done via the phone itself through digital currency like Paytm, NetBanking and Paypal

B. Central Server Module (CSM)

- The wireless connection between the CM and the KM will be routed via a central cloud server which will handle all the operations.
- A cloud database will be created and menu items, specials, information about the staff and customers, reviews, etc. will be kept updating on the server and stored on the database.
- Reviews of the food, review of the restaurant and details and information will all be accessed by the administrator which will help in overall development of the restaurant.
- The SVM model after training will be deployed on the server which will classify the reviews given by the customers into two classes positive and negative. It will also help to give recommendations of the food to the customers according to the preferences set by them.
- All the necessary statistical calculation will be done on the server.

C. Kitchen Module (KM)

- Interface will be provided to chefs and kitchen staff so that they will know the orders given by the customers.
- They will also be able to set availability of the food accordingly which will be displayed on the CM
- Using the reviews given by the customers positive/negative reviews will be classified and hence will help to improve the restaurant. By tracking food and beverage sales on a daily basis.
- It will help to analyze sales trends, payroll costs, customer counts and predict future sales.

The research has proposed a Machine Learning model which will help in classification of reviews of the restaurant as well as classification of reviews of food served by the restaurant. This model is trained using SVM (Support Vector Machine) Algorithm to classify the reviews.

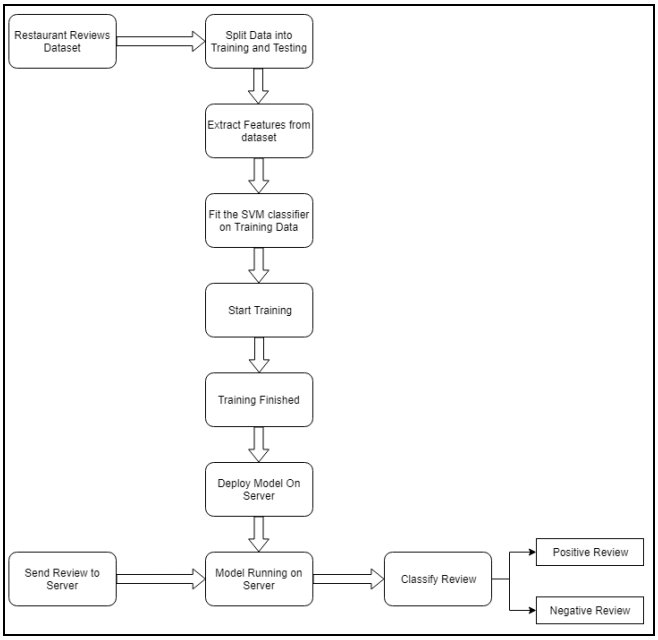


Figure 3: Flow of Working of SVM Classifier

Consider a set T of t training feature vectors $x_i \in \mathbb{R}^{\mathcal{D}}$, i = 1,….t, and the corresponding class labels $y_i \in \{+1,-1\}$ (for the binary classification). Vectors with the class label +1 (Positive Review) are the positive ones (class C+), whereas the others (Negative Reviews) belong to the negative class C−.

Linear SVMs separate data in the D-dimensional input space with the use of the decision hyperplane defined as

$$f(x): w^T x + b = 0, \quad (1)$$

Where w is the hyperplane normal vector, $x_i \in \mathbb{R}^{\mathcal{D}}$, and $b/||w||$ is the perpendicular distance between the hyperplane and the origin ($\|\cdot\|$ is the 2-norm), b $\in \mathbb{R}$. This hyperplane is positioned such that the distance between the closest vectors of the opposite classes to the hyperplane is maximal.

For two linearly separable classes (as already mentioned, with the class labels $y_i \in \{+1,-1\}$, the training data must satisfy the following conditions:

$$w^T x_i + b \quad \geq 1 y_i = +1 \quad (2)$$

$$w^T x_i + b \quad \leq -1 y_i = -1 \quad (3)$$

which can be re-written as:

$$y_i(w^T x_i + b) - 1 \geq 0 y_i \in \{+1,-1\}. \quad (4)$$

The equalities from Eq. (4) hold for the vectors positioned on two parallel hyperplanes, with the distance to the origin given as $|1-b|/||w||$ and $|-1-b|/||w||$, respectively. There are no vectors between these two planes, and the distance between the separating hyperplane and each of these planes is $1/||w||$. Hence, the maximal theoretical margin possible to generate by the decision hyperplane is

$$\varphi(w) = \frac{2}{||w||}. \quad (5)$$

Since we intend to maximize the separating margin, the value of $||w|| = \sqrt{w^T w}$ should be minimized:

$$\min_{w,b} ||w||. \quad (6)$$

To simplify the calculations, it can be given as the quadratic term:

$$\min_{w,b} \frac{||w||^2}{2}. \quad (7)$$

The optimization is performed with respect to the constraints in Eq. (4)—it becomes a quadratic programming (QP) problem. This formulation of the problem is called the primal form. The resulting hyperplane is exploited to classify the incoming data based on the decision function

$$f(a) = \text{sgn}(w^T a + b), \quad (8)$$

where a is a feature vector to be classified.

If we re-write Eqs. (4) and (7) to get the Lagrangian in its primal form, we have

$$\mathcal{L}(w,b,\alpha) = \frac{||w||^2}{2} - \sum_{i=1}^{t} \alpha_i y_i (w^T x_i + b) + \sum_{i=1}^{t} \alpha_i, \quad (9)$$

where α_i are the Lagrange multipliers. This transformation allows for representing the constraints given in Eq. (4) as the constraints on the Lagrange multipliers. In this formulation, the data in both training and test sets will appear in the form of the dot product between the vectors.

Since retrieving the SVM hyperplane is a convex optimization problem, determining the hyperplane is equivalent to finding a solution to the Karush–Kuhn–Tucker (KKT) conditions. The KKT conditions for Eq. (9) are:

$$\begin{cases} \frac{\partial}{\partial w}\mathcal{L}(w,b,\alpha) = w - \sum_{i=1}^{t} \alpha_i y_i x_i = 0 \\ \frac{\partial}{\partial b}\mathcal{L}(w,b,\alpha) = -\sum_{i=1}^{t} \alpha_i y_i = 0 \end{cases} \quad (10)$$

such that

$$y_i(w^T x_i + b) - 1 \geq 0 i = 1,2,\dots,t \quad (11)$$

$$\alpha_i \geq 0 \forall_i \quad (12)$$

$$\alpha_i(y_i(w^T x_i + b) - 1) = 0 \forall_i \quad (13)$$

Incorporating the equation for w from Eq. (10) into Eq. (9)

$$w = \sum_{i=1}^{t} \alpha_i y_i x_i \quad (14)$$

and knowing that

$$\sum_{i=1}^{t} \alpha_i y_i = 0, \quad (15)$$

we have

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{t} \alpha_i - \frac{1}{2}\sum_{i=1}^{t} \sum_{j=1}^{t} \alpha_i \alpha_j y_i y_j x_i^T x_j, \quad (16)$$

where LD denotes the dual form of the Lagrangian. The dual problem may be solved by maximizing LD with respect to α, subject to the constraints given in Eqs. (11)–(13) (this is the Wolfe dual of the problem). Only a small subset (containing s vectors) of the entire T (i.e., SVs) contributes to the position of the hyperplane. The Lagrange multipliers αi corresponding to the SVs are greater than zero. Finally, the decision function becomes:

$$f(a) = \text{sgn}(\sum_{i=1}^{t} \alpha_i y_i x_i^T a + b). \quad (17)$$

In order to apply the above reasoning for non-separable cases, it is necessary to relax the constraints given in Eqs. (2) and (3), and to introduce an additional cost of this operation:

$$w^T x_i + b \geq 1 - \xi_i y_i = +1 \quad (18)$$

$$w^T x_i + b \leq -1 + \xi_i y_i = -1 \quad (19)$$

$$\xi_i \geq 0 \forall i \quad (20)$$

where ξi denotes a positive slack variable. The objective function should be modified to take into account the classification errors:

$$\min_{w,b,\xi} \frac{||w||^2}{2} + C\sum_{i=1}^{t} \xi_i \quad (21)$$

such that

$$y_i(w^T x_i + b) \geq 1 - \xi_i i = 1,\dots,t \quad (22)$$

$$\xi_i \geq 0 i = 1,\dots,t \quad (23)$$

where C is the parameter that controls the trade-off between the margin and the slack penalty (the larger the value of C, the higher penalty to the errors). Considering this trade-off allows for introducing the soft-margin SVMs. As in the separable case, Eq. (21) can be easily transformed into its Wolfe's dual form:

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{t} \alpha_i - \frac{1}{2}\sum_{i=1}^{t} \sum_{j=1}^{t} \alpha_i \alpha_j y_i y_j x_i^T x_j. \quad (24)$$

It is to be maximized, subject to

$$0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^{t} \alpha_i y_i = 0. \quad (25)$$

Finally, we have

$$w = \sum_{i=1}^{s} \alpha_i y_i x_i. \quad (26)$$

As in the separable case, we can retrieve the Lagrangian in its primal form:

$$\mathcal{L}(w,b,\alpha) = \frac{||w||^2}{2} + C\sum_{i=1}^{t} \xi_i - \sum_{i=1}^{t} \alpha_i[y_i(w^T x_i + b) - 1 + \xi_i] - \sum_{i=1}^{t} \mu_i \xi_i, \quad (27)$$

where μ_i enforces the positivity of ξ_i. The KKT conditions can be retrieved for the non-separable case following the reasoning presented for the separable one.

IV. FUTURE SCOPE

The Developed System will be easy and user friendly to use by the customers as well as kitchen staff. The system can be used and implemented by any restaurant or café since the system needs minimum resources to run. The statistics calculated and reviews analyzed using Machine Learning (SVM Algorithm) will help in growing the restaurant business.

The system created is to be used and implemented by a restaurant or a café. The System is economically feasible and will be made accordingly with all functions that are required. In case of one or more modules of the system fail or a network failure happens, the system will fail to operate smoothly, still the orders can be placed manually and the issue will be fixed at the earliest. All the orders will be given through the customer interface. Only the food will be delivered by the waiter.

The SVM algorithm learns by taking data as input. The more the data and better tuning of the algorithm, more accurate results will be obtained. There might be a very few times where the customers will not get their proper recommendation or reviews might not be analyzed correctly. The chance of this happening is very less since the training of the machine learning model will be done precisely.

V. LITERATURE SURVEY

F.M Takbir Hossain, Md. Ismail Hossain, Ms. Samia Nawshin did an experimental study, where 827 reviews were collected of different restaurants to measure the efficiency of the approach. The techniques of finding positive/negative keywords as identifier from the text documents are described. By integrating all the reviews, the market position of a restaurant can be determined. For this experiment authors have used four proper algorithms for building proper model by using training dataset. Those are Multinomial Naïve Bayes (MNB), Support Vector Machine (SVM), K nearest neighbor (KNN) and Linguistic Regression (LR). For measuring the performance of their built model they have used k-fold cross validation which is very effective way for finding the accuracy of a predictive model. It divides the data set into training and testing set and computes the performance. They have used 10 fold in our experiment that means k=10. Linguistic Regression (LR) has given them the best result with more than 77% accuracy. [1]

Toshihiko Yamasaki, Masafumi Yamamoto, and Kiyoharu Aizawa presented a review-based service recommendation system using both a bag-of-words model and a skip-gram-based model. Namely, instead of checking the scores or review comments, a review is given to the system as a query, and the services to which similar reviews are given are recommended. They evaluated our algorithms by using Yelp's Academic Dataset. The experimental results show that the percentage of finding the corresponding restaurants as the top-1 candidate was 25% and that in the top-10 ranking was 51%. [2]

Yan Zhu, Melody Moh, Teng-Sheng Moh, presented a multi-layer architecture for sentimental analysis has been proposed, utilizing Naïve Bayes, SVM, Random Forest and Logistic Regression and classifiers, plus a voting scheme commonly used in ensemble learning. The architecture has been applied to the restaurant reviews of Yelp platform, grouping user reviews into five classes: 1-star to 5star. Performance evaluation is carried out by experiments based on a widely-used machine learning package sklearn. The results show that the multi-layer architecture has improved the results of single-layer architecture

by over 50%; with the use of voting scheme, a combination of feature set and word score, and a large training set would also improve the accuracy. Here it shows that when the training size is increased improvements of 5.6% and 6.3% respectively for hard voting and soft voting are achieved. The results show that the multi-layered architecture has improved the results of single layer architecture by over 50% with the use of scheme, a combination of feature set and word score and a large training set. [3]

Ling Li*, YaZhou, Han Xiong, Cailin Hu, Xiafei Wei have put forward a new method of user similarity calculation combined with traditional collaborative filtering algorithm. In the view of rating similarity, its intensified impact of the average score factors on similarity and introduced confidence level correction factor to improve the credibility. In addition, it exploited users' attributes to calculate similarity of users' attributes, realizing collaborative filtering based on users' ratings and attributes. It conducted a set of comparative experiments on the dataset crawled from dianping.com site. Experimental results indicate that our approach performing more effective and precise in restaurants recommendation. [4]

Sanjukta Saha, Dr.A. K. Santra took feedbacks of online customer as an integral part for deciding before purchasing any product. In this paper the customer's textual feedback has been considered for rating the restaurants available in Kolkata. The users' sentimental approach is measured first and then we have to calculate each user's sentiment on the food item. With that we can also consider an interpersonal sentimental influence. The performance evaluation of our experimental results is based on the concept of collaborative filtering and it is made on the real-life dataset collected from the "Zomato" rating website. This analysis can well denote the user's preferences which can upgrade our recommendation system. [5]

S. Prakhash, A. Nazick, R. Panchendrarajan, M. Brunthavan, S. Ranathunga, A. Pemasiri analysed different aspects such as food, service, and ambience that a customer would look for, when deciding on a restaurant to dine in. Among these aspects, the type of food it sells and the food quality are the most important. Therefore, when automatically rating restaurants based on customer reviews, the food aspect plays a major role. There exists some research on rating individual food items in a restaurant. However, a potential customer requires not the ranking of an individual food item, but the ranking of a particular food category in general. In order to do that, a categorization of food names is required. This paper presents two techniques for food name categorization using document similarity measurements. [6]

ZhenhaiMu, Zhongxuan Tan, Ge Zhu designed system using Android a system for restaurant is developed which will reduce the waiting time in the restaurants and they can not only order meals in advance so as to minimize the waiting time, but also avoid the embarrassment of being cheated by restaurants. This system features online ordering function, which can adequately guarantee consumers' rights and interests, and offer O2O catering services between restaurants and consumers. [7]

Fahim Rarh, Dastgir Pojee, Sajjad Zulphekari, Dr. Varsha Shah presented a system which gives a proper solution to the problem of long waiting times for the customers in a restaurant by allowing them to pre-book the table online. [8]

Kuan-Yu Lina, Chih-Hun Chen a, Zhe-Ming Zhang b, Sheng-Chuan Ou conducted a study which implements a restaurant ordering App system, by which a customer-end App allows preview of restaurant information, including menus and benefit information to resolve the problem of the huge space in food court. [9]

Hassain Saeed, Ali Shouman, Mais Elfar, Mostafa Shabka, Shikharesh Majumdar, Chung Horng-Lung presented a paper an approach to order food in the restaurant using android application and through the use of proximity sensors and NFC sensors, customers can also use the Android application on their personal smart devices to find available parking spaces, find an available table, and pay for their own bill. [10]

VI. CONCLUSION

Thus, in this project an efficient and user-friendly method is proposed which will provide automated systems in the restaurant and solve problems faced by the restaurants using technologies like Android, Web Development and Machine Learning. Interactive User Interfaces for the customers and restaurant staff will be provided and customers can order food directly through the module without interacting with waiter.

Using Machine Learning Models prediction of the food preferred by the customers and also information necessary for the restaurant to grow in business through customer reviews and other data. The system saves a lot of time of the customer as well as the restaurant staff and helps the restaurant in many ways.

VII. REFERENCES

[1] F.M Takbir Hossain, Md. Ismail Hossain, Ms. Samia Nawshin, "Machine Learning Based Class Level Prediction of Restaurant Reviews", 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) 21 - 23 Dec 2017, Dhaka, Bangladesh, Pg 420 – 423

[2] Toshihiko Yamasaki, Masafumi Yamamoto, and Kiyoharu Aizawa, "Review-Based Service Profiling and Recommendation", 2016 Joint 8th International Conference on Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems (ISIS), Pg 377-380

[3] Yan Zhu, Melody Moh, Teng-Sheng Moh, "Multi-Layer Text Classification with Voting for Consumer Reviews", 2016 IEEE International Conference on Big Data (Big Data), Pg 1991-1999

[4] Ling Li*, YaZhou, Han Xiong, Cailin Hu, Xiafei Wei, "Collaborative Filtering based on User Attributes and User Ratings for Restaurant Recommendation",2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Pg 2592-2597

[5] Sanjukta Saha, Dr.A. K. Santra , "Restaurant Rating Based on Textual Feedback", 2017 International conference on Microelectronic Devices, Circuits and Systems (ICMDCS)

[6] S. Prakhash, A. Nazick, R. Panchendrarajan, M. Brunthavan, S. Ranathunga, A. Pemasiri, "Categorizing Food Names in Restaurant Reviews", 2016 Moratuwa Engineering Research Conference (MERCon)

[7] ZhenhaiMu, Zhongxuan Tan, Ge Zhu, "Android based order recommender system", 2018 International Conference on Robots & Intelligent System, Pg 258-261

[8] Fahim Rarh, Dastgir Pojee, Sajjad Zulphekari, Dr. Varsha Shah, "Restaurant Table reservation using time-series prediction",2017 Proceedings of the 2nd International Conference on Communication and Electronics Systems (ICCES 2017) IEEE Xplore Compliant – Part Number: CFP17AWO-ART, ISBN:978-1-5090-5013-0, Pg 153 – 155

[9] Kuan-Yu Lina, Chih-Hun Chen a, Zhe-Ming Zhang b, Sheng-Chuan Ou, "NFC-based mobile application design restaurant ordering system APP", 2018 Proceedings of IEEE International Conference on Applied System Innovation 2018 IEEE ICASI 2018- Meen, Prior & Lam (Eds), 737- 740

[10] Hassain Saeed, Ali Shouman, Mais Elfar, Mostafa Shabka, Shikharesh Majumdar, Chung Horng-Lung, "Near-Field

Communication Sensors and Cloud-Based Smart Restaurant Management System", 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT), Pg686 – 691

[11] https://play.google.com/store/apps/details?id=com.ultimate.restaurantmanag

[12] https://play.google.com/store/apps/details?id=com.mcookin

Crack Detection Software for Building Surface

Craig Menezes, Pruthav Singhadia, Ashukalp Srivastava, Prachi Janrao
Computer Engineering, TCET University of Mumbai
craigpmenezes@gmail.com, pruthav8@gmail.com, aashukalp@gmail.com,
prachi.janrao@thakureducation.org

Abstract
Cracks are an important indicator reflecting the safety status of infrastructures. This paper presents an automatic crack detection and classification methodology for building safety monitoring. With the application of high-quality cameras, the building surface can be captured and stored in digital images. In a next step, the local dark regions with potential crack defects are segmented from the original gray-scale images by utilizing morphological image processing techniques and thresholding operations. In the feature extraction process, we present a though transformation-based shape descriptor that effectively describes the spatial shape difference between cracks and other irrelevant objects.

Keywords-Crack detection; crack classification; line scan cameras component.

I. INTRODUCTION

Cracks are an important indicator when it comes to the safety of an infrastructure. Effects such as cyclic loading, fatigue stress develop internal stresses in the structure and leads to the development of cracks. In addition, effects like thermal expansion and contraction, human damage and surface changes transcend material discontinuities which lead to failure. To prevent this damage a Crack detection process is employed.
The crack detection can be done using two ways:
(1) Manual Inspection
(2) Automated Inspection.
In conventional methods, manual inspection was done by set of skilled inspectors with the help of surveying instruments and visual examination to detect the irregularities and defects in the structure. However, this method has certain drawbacks, as it is impossible for a crew to detect the cracks in inaccessible areas such as large dams, monuments, buildings, etc. and also the estimation of size, length and width of the crack.

In order to get images of the structure in a required manner there is a need of a high resolution camera to get desired results through image processing. Because of its accuracy and simplicity in its results many of the image processing techniques were proposed. Image processing techniques such as Morphological approach, Hough transform, Edge detection Dijkstra's algorithm, Neural Network, Statistical approach, Segmentation are the process that can be used to detect the cracks in building.

Softwares like image processing lab, PyCharm and MATLAB Graphical User interface can be used for easy monitoring and real time inspection through which data can be extracted.
Initially the structure of the image processing based crack detection is proposed.
• Crack detection using image processing:

The steps involved in image processing techniques are as follows:
(1) Acquisition of an image with the help of a camera as a crack image as an input.
(2) After acquisition the collected images go through pre-processing where it undergoes processes such as scaling, RGB to HSV conversion and segmentation.
(3) In colour image processing, some of the techniques are employed to process the deducted image.
(4) Finally, the feature extraction of image is done to obtain the shape of the crack in an infrastructure.

II. LITERATURE SURVEY

Based on the reviewed paper, the analysis that was provided shows the various image processing techniques, objectives, accuracy level, error level, and the image data sets. The software presents the various research issues which can be useful to for further research work on the crack detection.

The paper which was referenced for studying image processing gives us a detailed survey of the image processing techniques used to detect cracks in engineering structures. The objectives such as width, depth, length and direction of propagation of crack can be analyzed.

E. Balasubramanian et al. [1] have proposed a technique that uses combination of Hat-transform and HSV thresholding for crack detection. They had proposed an algorithm which would enhance the output image by combining the output of the two filters. It was based on mathematical morphological approach and their study showed that bottom hat transform technique was better to identify cracks than Top-hat transform. The former method was then combined with the HSV thresholding which gave precise output of the cracks. Hough transform algorithm is used extensively applications like barcode scanning, verification and recognition.

III. PROPOSED SYSTEM

In this section, the crack detection procedure and method is explained thoroughly. The crack detection procedure starts by taking images captured from a camera, running these images through program that stitches the images together and generates a detailed crack report using Hough transform algorithm. The robustness of the crack detection code will be checked using the concept of precision and accuracy, which will be explained later.

Thresholding
Thresholding is a method of image segmentation. From a grayscale image, thresholding can be used to create binary image. Thresholding method extract useful information encoded into pixels while minimizing background noise. The idea is to separate the image into two parts; the background and

foreground. Thresholding method includes selecting the initial threshold value, which typically the mean 8-bit value of the original image. Dividing the original image into two portions; First - Pixel values that are less than or equal to the threshold; background and second - Pixel values greater than the threshold; foreground. The average mean of the two images will be calculated. If difference between the previous threshold value and the new threshold value are below a specified limit, task is finished. If not, then repeat the process.

Hough Transformation

In an image processing context, the Hough transformation of a mathematical method through which we can detect shapes even if it is broken or distorted. The Hough transform is a technique which can be used to isolate features of a particular shape within an image. It requires desired features to be specified in parametric form, the *classical* Hough transform is most commonly used for the detection of regular curves such as lines, circles, ellipses, *etc.*

The main advantage of the Hough transform technique is that it is tolerant of gaps in feature boundary descriptions and is relatively unaffected by image noise which is very important in detecting cracks, cracks in the walls are generally not regular. Lines [10] can be represented uniquely by two parameters. Often the form in Equation 1 is used

with parameters a and b.

$$y = a \cdot x + b. \qquad (1)$$

This form is, however, not able to represent vertical lines. Therefore, the Hough transform uses the form in Equation 2, which can be rewritten to Equation 3 to be similar to Equation 1. The parameters θ and r is the angle of the line and the distance from the line to the origin
respectively.

$$r = x \cdot \cos\theta + y \cdot \sin\theta \qquad (2)$$

$$y = -\frac{\cos\theta}{\sin\theta} \cdot x + \frac{r}{\sin\theta} \qquad (3)$$

All lines can be represented in this form when $\theta \in [0, 180[$ and $r \in R$ (or $\theta \in [0, 360[$ and $r \geqslant 0$).

The Hough space for lines has therefore these two dimensions; θ and r, and a line is represented by a single point, corresponding to a unique set of parameters (θ0, r0). The operation exceptionally straightforward. The picture is examined in a solitary pass and a running check of the quantity of pixels found at every power esteem is kept. When we apply image over the algorithm, it will then convert it from RGB to HSV and makes a black & white image which only have the desired crack in the wall. Figure 1 neatly describes the output. It is then further processed, and the length of the crack will be calculated which is then compared with a pre-stores database. If the length of the crack is more than the safety standards then it will be taken into consideration. The first parameter is an Input image which is used to apply threshold or we can also use canny edge detection before finding and applying hough transform. Second and third parameters are ρ and θ accuracies respectively. Fourth argument is the threshold, which means minimum vote it should get for it to be considered as a line. The number of votes depend upon number of points on the line. So it represents the minimum length of line that should be detected. Two types of lines will be considered as an output if the fall in the range of :

minLineLength and maxLineGap

In minLineLength, if the length of the line segment is shorter than the desired minimum length than it will be rejected.
In maxLineGap, a maximum gap between two line segments will be considered to treat them as a single line.



Fig 1: Sample picture of the crack before processing.

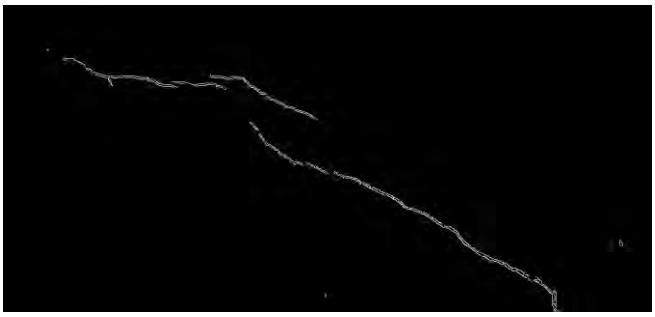


Fig 2 : Image of crack after processing using Thresholding

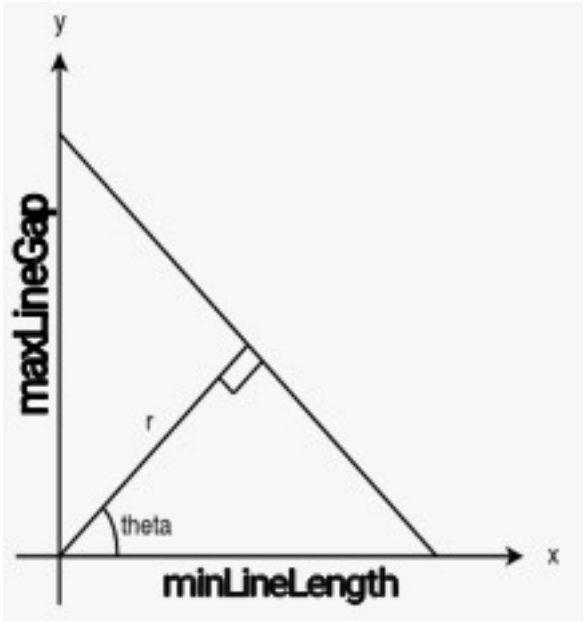


Fig 3

Algorithm

Notation: (x,y) will be the coordinates of a point and r is the distance between the straight line going perpendicularly to the

hypotenuse of the points formed like in figure 2. Θ will be the angle between the r and x-axis.

1. First it initialize a 2D array with zero value
2. Rows denote the r and columns denote the Θ theta.
3. Size of array is proportional to the accuracy you need. 180 degree means each angle denotes a pixel.
4. For r, the maximum distance possible is the diagonal length of the image.
5. It finds out the parameters with
   $r = x\cos\theta + y\sin\theta$

IV. FEASIBILITY

4.1 Economical Feasibility:

When we analyse the EFS there are two aspects:

- Business Case

The Business Case gives us a detailed study of the business environment which include-
Expected users: Public/Authority
The nature of the business: Public/Gov.
The Business Case also presents the benefits of the proposed project.

- Cost Benefit Analysis

The Cost Benefit Analysis outlines the incomes and costs required with the proposed project. As the proposed framework will be utilized for the advantages of the public, no extra expense will be paid by them. The only equipment used is a good quality camera to click images, so the equipment cost gets limited. Just insignificant sum will be required by certain product. Subsequently, our framework is Cost Efficient.

4.2 Technical Feasibility:

Technology used at front end: Python, CSS, PHP, OpenCV & Bootstrap.
Technology used at Back end: PHP & MySQL
Resources Required: Programmers, testers, debuggers
Software required: Testing Tools (perform black box and white box testing), Github
Editor: Pycharm
Hardware required: Camera for detection on laptop for development, server for deployment
Infrastructure: For scaling up the project

4.3 Managerial Feasibility:

Management support, employee involvement, and commitment are key elements required to gauge managerial feasibility in the proposed project. Management support, employee involvement, and commitment are key elements required to gauge managerial feasibility in the proposed project. The success and the profitability of the project partly depend on managerial competence of the major ingredients of the proposed project which is the System i.e. The system must detect with high accuracy. The capability of the infrastructure of a process is to achieve and sustain the properties of isolation, atomicity, durability, consistency etc in the matter of the data stored of users. Also, constant attention is required in the case of amount of log in files being created.

4.4 Operational Feasibility:

The proposed system is concerned with the Public sector companies involved in structural maintenance. At initial phase the scope of the system will be to detect cracks. The portal will be handled by a central manager who is responsible for all the collected data. The end users of the system will be the citizens or the government officials. The important stakeholder of the system will be the Common people, however they won't be directly using the system. Detection of cracks is the key process of the system which will be easily handle by an automated part of the system. If the usage of the system is very beneficial and helpful to Public sector then the system can be expanded at the global level also. At the end the system will contribute to improve the conditions of the buildings.

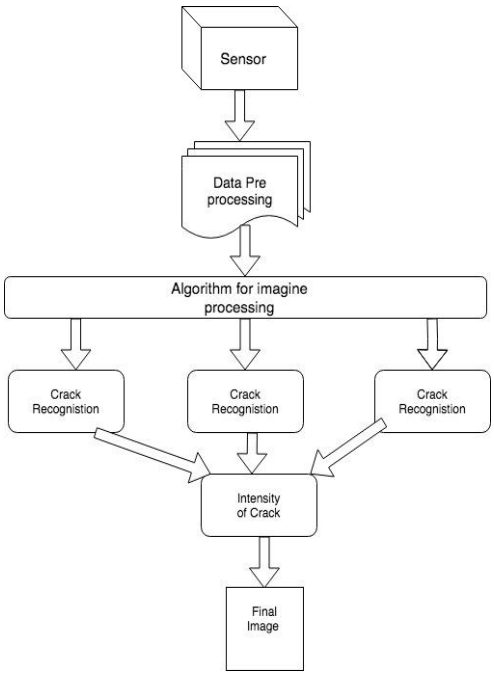


Fig 4: System Architecture

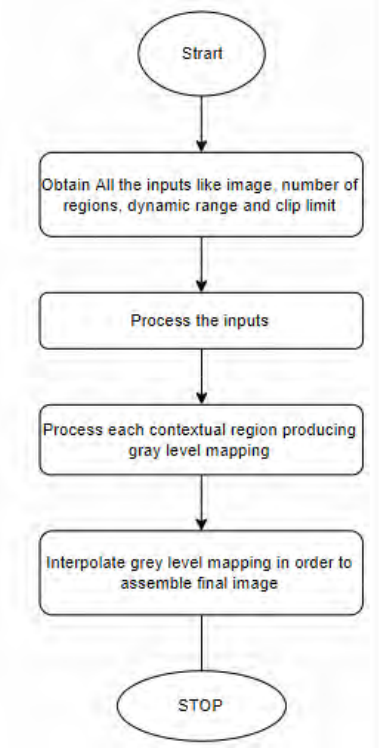


Fig 5: Flowchart

V. Results and Discussions

To do performance analysis, different metrics are considered to evaluate the performance of proposed technique. The overall objective of this study is to predict the presence of a crack accurately in minimum time. The image processing approach involves scanning of various images of cracks to check the vulnerability of the building. In the Figure 6 we have taken a sample image for the processing. Further Hough Transformation is applied to the crack.jpeg as shown in Figure 7. In the final model, only the more significant and suitable parameters are used to predict accurate result. From the result shown in Figure 8 it is observed that the accuracy of the system improved with the images that are processed. But in this technique also time complexity is high.



Fig 6. Image of crack (jpeg)

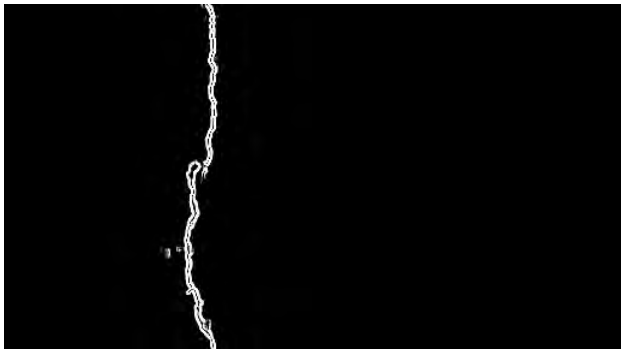


Fig7.Image after Hough transformation



Fig 8. Final image after processing

VI. Conclusion

This paper provides detailed survey of the image processing techniques used to detect cracks in engineering structures. The main intention of the study was to review papers and analyse the best technique that could be used for crack detection. Based on the paper reviewed we conclude that the image processing techniques such as Morphological and Thresholding are the most suitable for crack detection. The objectives such as width, depth, length and direction of propagation of crack can be analysed. The segmentation of images grouped into set of data give better and accurate results.

Refrences

[1][S. Sankarasrinivasan, E. Balasubramanian, K. Karthika, U. Chandrasekar and Rishi Gupta, Health Monitoring of Civil Structures with Integrated UAV and Image Processing System, Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015).]
[2] [The International Journal of Engineering and Science (IJES) || Volume || 6 || Issue || 6 || Pages || PP 59-63 || 2017 || ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805 DOI: 10.9790/1813-0606015963 www.theijes.com Page 59 A Study for Applications of Histogram in Image Enhancement]
[3] http://www.opencv.org
[4]Tobias OJ, Seara R. "Image Segmentation by Histogram Thresholding Using Fuzzy Sets," IEEE Transactions on Image Processing, 2002; 11(12):1457-1465.
[5] http://www.ijettcs.org/Volume4Issue5(2)/IJETTCS-2015-10-01-7.pdf
[6] http://en.wikipedia.org/wiki/Greyscale
[7]http://en.wikipedia.org/wiki/Digital_image_processing
[8] http://iosrjournal.org/iosr-jce/papers/Vol17-issue2/Version-5/B017250715.pdf
[9]https://www.sciencedirect.com/science/article/pii/0734189X86900733
[10]http://web.ipac.caltech.edu/staff/fmasci/home/astro_refs/HoughTrans_lines_09.pdf

A Comparative Analysis of Object Detection Methods

Niranjan Abhyankar, Rushikesh Shinde, Shubham Dukare, Dhananjay Gadakh, Rohini A. Patil

Department of Computer Engineering Terna Engineering College, Nerul, Navi Mumbai, Maharashtra

***Abstract*—Deep Learning is a sub field of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks. In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analyzing visual imagery. In this paper we try to compare and analyze the various aspects of object detection algorithms namely Single Shot Detector(SSD)[1], Fast R-CNN[2], Faster R-CNN[3], R-FCN[4] and YOLO[5]. This paper aims to find out the best combination of speed and accuracy by comparing different algorithms used for object detection using convolutional neural networks.**

I. INTRODUCTION

There exist many different state-of-the-art object detection technologies which use Convolutional Neural Networks. The technology to be used, namely SSD [1], Fast R-CNN [2], Faster R-CNN [3], R-FCN [4] and YOLO [5] depends upon the kind of application desired. Another improvement to object detection is their precise localization by using Deep Neural Networks (DNN)[9] in which a multi-scale inference procedure can produce high-resolution object detections at a low cost. In the case of applications like Google's SSD mobile net, memory efficiency is the most important factor while in the case of self-driving cars, accuracy is the most important factor. We have considered only such models which are similar in their overall architecture. These models are responsible for performing object detection whereas the feature extraction is done by image classification models. Object detection can be defined as a regression problem to the coordinates of several bounding boxes [7]. These object detection models, also known as meta architectures are combined with the feature extraction models to obtain the various evaluations. We would first describe the various meta architectures and then compare their attributes. In this paper we have tried to present an overall comparison of the object detection techniques in terms of the previously mentioned parameters.

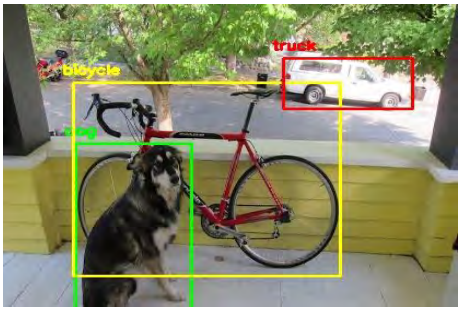


Fig. 1: An example of object detector in use

II. META ARCHITECTURES

Many modern meta architectures use CNN for object detection. Let us take a look at various meta architectures and discuss their working.

A. Single Shot Detector(SSD)

SSD [1] converts the discrete output space of bounding boxes into a set of default output boxes for every feature map location and for different aspect ratios. At the time of performing predictions, the model checks for presence of each object category in each default box and generates scores accordingly. It further adjusts the boxes to better match the shape of the contained object. The SSD network naturally handles objects of various sizes by combining predictions from multiple feature maps with different resolutions. SSD combines all computation in a single network. It also completely eliminates stages like subsequent pixel and feature resampling thus making it simpler than methods that require object proposal. All these utilities make SSD easy to train and integrate into systems which require an object detection component. The Single Shot Detector approach is based on the feedforward convolutional network. This network works by producing fixed size bounding boxes and assigns scores for each box by checking for an instance of the object class in those boxes.

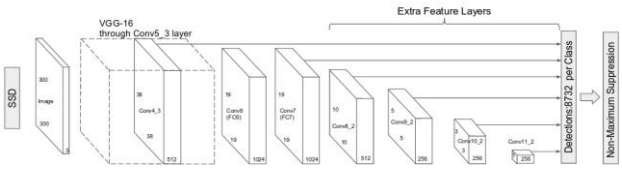


Fig 2: Single shot detector architecture

The base network consists of standard architecture used for high quality image classification. This base network is then appended by an auxiliary network to produce object detection. The auxiliary network consists of the following key features:

1) Multi-scale feature maps for detection: Convolutional feature layers which decrease in size progressively are added to the end of the base layer. They allow predictions of detection at multiple scales. Each feature layer consists of a different convolutional model for predicting detection.

2) Convolutional predictors for detection: A fixed set of detection predictions is performed by each feature layer using a set of convolutional filters. To predict parameters in an m x n sized feature layer consisting of p channels, the basic requirement is a small kernel of size 3 x 3 with p channels. This kernel produces either a score for the category or a shape offset of the object relative to default box containers. The output values are calculated by applying the kernel to each of the m x n locations.

3) Default boxes: Each feature map cell is associated with a set of default bounding boxes. The position of each default box with respect to its corresponding cell is fixed. They help to predict the

offsets relative to each default box shape in the cell as well as the probability of each box containing the presence of the class instance.

An improved version of SSD is the DSSD in which deconvolutional layers are placed after the original SSD setup to improve the resolution of feature map layers[6].

B. R-CNN

The R-CNN model was among the first models to use convolutional neural networks for object detection. The working of R-CNN is simple, it divides the given input image into bounding boxes and tries to find if any instance of the classes are present in those bounding boxes.

In R-CNN the CNN is forced to focus on a single region at a time because that way interference is minimized as it is expected that only a single object of interest will dominate in a given region. The regions in the R-CNN are detected by selective search algorithm followed by resizing so that the regions are of equal size before they are fed to a CNN for classification and bounding box regression. The final layer consists of a Support Vector Machine(SVM) that is used to identify the object, if any,in the bounding boxes.

The Region based Convolutional Neural Network uses a deep Convolutional Network to classify object proposals and thus achieves excellent accuracy. However, it has the following drawbacks:

1) Training is carried out in a multi-stage pipeline: Firstly, R-CNN fine-tunes a ConvNet on object proposal. Then it fits the Support Vector Machines (SVMs), which act as object classifiers, to the ConvNet features. Lastly it learns bounding box regression.

2) Training is expensive in space and time: For bounding-box regression training features need to be extracted from each object proposal in each image and written to disk. This process occupies a large amount of space on the disk and requires significant time to complete.

3) Object Detection is slow: As features from each object proposal in each image need to be extracted at test time, object detection can take a large amount of time.

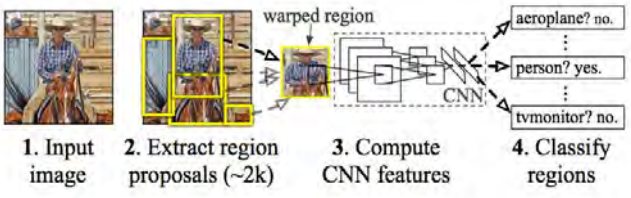


Fig. 3: R-CNN framework for object detection.

Below is a brief summary of the steps followed in RCNN to detect objects:

1) Take a pre-trained convolutional neural network.

2) Then, this model is retrained. Train the last layer of the network based on the number of classes that need to be detected.

3) The third step is to get the Region of Interest for each image. Reshape all these regions so that they can match the CNN input size.

4) After getting the regions, train SVM to classify objects and background. For each class, train one binary SVM.

5) Finally, train a linear regression model to generate tighter bounding boxes for each identified object in the image.

C. Fast R-CNN

R-CNN is accurate but quite slow. R-CNN forms bounding boxes across the image and tries to detect object instances in them. The creator of R-CNN realized that a significant amount of bounding boxes were overlapping leading to the same CNN computations multiple times. In order to solve this issue, the creator of R-CNN himself proposed another method known as Fast R-CNN[2]. He stated that such overlapping calculations can be reduced if they could be possibly shared across the proposals. Fast R-CNN achieves this by implementing a technique known as Region of Interest pooling (RoI pooling).

A Fast R-CNN network takes as input an entire image and a set of object proposals. The network first processes the whole image with several max pooling and convolutional layers to produce a convolutional feature map. Then, for each object proposal a region of interest (RoI) pooling layer extracts a fixed-length feature vector from the feature map.

The RoI pooling layer, which can be considered as a rectangular window in a convolutional feature map, converts the features inside a valid region of interest into smaller feature maps using max pooling. The RoI pooling layer works by dividing the RoI window into a grid of smaller sub windows and then max-pooling the values in each sub-window into the corresponding output grid cell.

Another insight of Fast R-CNN is to combine the CNN, classifier and bounding box regressions into a single model. Due to this, Fast R-CNN uses a single network to perform image feature extraction (CNN), classification (SVM) and tightening of bounding boxes (regression). Fast R-CNN used the softmax layer on top of the CNN instead of SVM to perform classification. A linear regression layer parallel to the softmax layer was also appended to the model to output bounding box coordinates. Thus, Fast R-CNN combined all the three models of the native CNN into a single model.

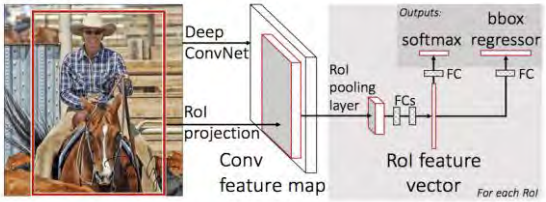


Fig. 4: Fast R-CNN architecture

Following are the steps followed in Fast R-CNN:

1) Take an image as an input.

2) This image is passed to a ConvNet which in turns generates the Regions of Interest.

3) A RoI pooling layer is applied on all of these regions to reshape them as per the input of the ConvNet. Then, each region is passed on to a fully connected network.

4) A softmax layer is used on top of the fully connected network to output classes. Along with the softmax layer, a linear regression layer is also used parallely to output bounding box coordinates for predicted classes.

To further improve Fast R-CNN, Zagoruyko and his colleagues presented a MultiPath Network for Object Detection [8], in which they modified Fast R-CNN by allowing information to flow through multiple paths in the network, including multiple object views and multiple network layers.

Another aspect in training an object detection model is to take care of deformations in the image.To address this issue, it is suggested by Xiaolong Wang and others to train the Fast R-CNN model along with an adversarial network that generates examples with occlusions and deformations[10]. It modifies the input image such that some

features of the image are hidden. This helps to train the model such that it can handle deformations in the input image.

D. Faster R-CNN

Faster R-CNN[3] is an improved version of Fast R-CNN object detection method. This technique consists of two modules. The first module is a deep fully convolutional network that proposes regions, and the second model is the Fast R-CNN detector. All the components of the system are integrated into a single, unified network used for object detection. In this technique, an RPN module directs the Fast R-CNN module to search the regions in the input image.

The major reason that the Fast R-CNN was still fairly slow was the region proposer. One of the first steps to detect the object locations is to generate a set of bounding boxes which can potentially contain an instance of the object classes. Fast R-CNN technique used the Selective Search method to generate these proposals. This technique was found to be the bottleneck of the overall process. Faster R-CNN uses the proposals generated by the forward pass of the CNN to generate region proposals instead of performing a separate algorithm for Selective Search. Faster R-CNN combines the region proposal module with the object classification module to generate a single CNN network. Thus, only one combined CNN model needs to be trained.

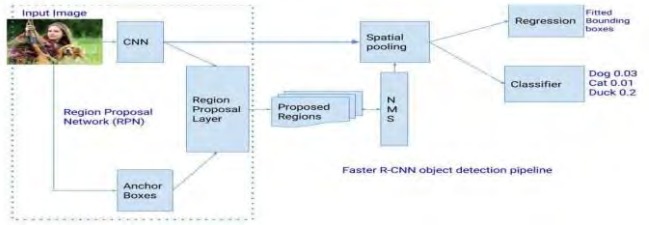


Fig 5: Faster R-CNN pipeline

Region Proposal Networks (RPN)

Faster R-CNN adds a Fully Convolutional Network on top of the features identified by the CNN module to create the Region Proposal Network. In object detection using Faster RCNN, RPN is the one true backbone and has proven to be very efficient till now. Its purpose is to propose multiple objects that are identifiable within a particular image. An RPN takes an input of any size and produces an output containing a set of rectangular object proposals or bounding boxes each with a score denoting the probability of finding an object in the box. Briefly, RPN ranks region boxes (called anchors) and proposes the ones most likely containing objects.

In order to generate these proposals for the region where the object lies, a small network is slid over a convolutional feature map that is the output by the last convolutional layer. This network takes an n × n spatial window of the input convolutional feature map as input. Each sliding window is mapped to a lower-dimensional feature. The output of this feature layer is simultaneously fed to a box regression layer (reg), that computes the box offsets, and the box classification layer (cls) that computes the confidence scores that are related to probability of objectness where objectness measures membership to a set of object classes.

The following steps are typically followed in a Faster RCNN approach:

1) Take an image as input and pass it to the ConvNet which returns the feature map for that image.

2) Region proposal network is applied on these feature maps. This returns the object proposals along with their objectness score.

3) A RoI pooling layer is applied on these proposals to bring down all the proposals to the same size.

4) Finally, the proposals are passed to a fully connected layer which has a softmax layer and a linear regression layer at its top, to classify and output the bounding boxes for objects.

Anchors

Anchors play an important role in Faster R-CNN. An anchor is a box. An anchor is used as a reference which are selected to have different aspect ratios and scales for differently shaped objects. For example, an object like a bus cannot be represented properly using square bounding boxes. In the default configuration of Faster R-CNN, there are 9 anchors at a position of an image representing 3 scales and 3 aspect ratios. So at each location of the convolutional layer, the bounding box regression heads outputs the bounding box offsets for each anchor box while the classification layer outputs the confidence scores that represents whether an object is present or not within each anchor box. Only those boxes with a corresponding high probability of object being present are further processed. Thus, the final proposals at each location are the anchor boxes along with the box offsets with a high probability of containing an object.

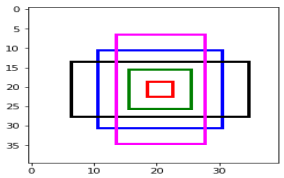


Fig 6: An anchor at position (20,20)

E. R-FCN

R-FCN[4] is an improved variant of Faster R-CNN. In the Faster R-CNN algorithm the region specific component had to be applied several times in an image. This increased computation caused degradation in performance. This issue was solved in R-FCN by making use of the crops of the image from the previous layer of features prior to predictions instead of taking them from the same layer where crops are predicted. This algorithm is faster than Faster R-CNN while also achieving comparable accuracy when Resnet101 feature extractor is used.

The basic architecture of R-FCN is same as of Faster R-CNN. The input to the algorithm is taken as an image. A base network known as ResNet-101 transforms an image to a feature map. After this step it is required to divide the image into bounding boxes. These bounding boxes act as containers to determine the presence of an instance of the objects in question in them. A region proposal network (RPN) uses the feature map to locate bounding box candidates in the image. A classifier uses the feature map and the bounding box candidates and classifies each one of them into C+1 classes, where C is the number of object classes to spot and 1 is added for the background. During that process, small sub regions of the feature maps which match the bounding box candidates must be extracted and converted to fixed-sizes matrices. The method to do that is called 'Region of Interest Pooling' (RoI-Pooling) and is based on max pooling.

A feature map M is created from the input image and divided into grids. With the help of M another feature map is created which is designed to recognize only a part of the image. For each individual grid block, a distinct feature map is created. These feature maps called position-sensitive score maps detect a sub-region of the object. The probability of finding an object in each block of the grid is calculated and the average of all these values is taken to calculate the class score. A softmax function is then applied on these scores to compute the probability for each class.

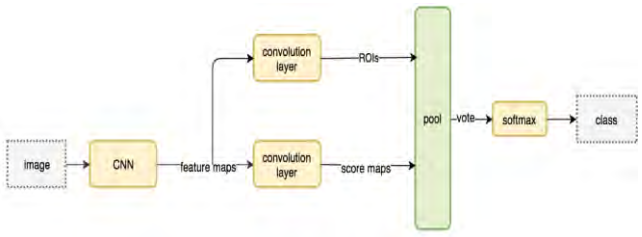


Fig 7: Network Flow of R-FCN architecture for object detection

F. YOLO

You only look once (YOLO)[5] is an object detection system targeted for real-time processing. YOLO is designed to be fast and to give results using just a single convolutional network. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Due to the whole detection pipeline being a single network, it can be optimized end-to-end directly on detection performance. Prior work on object detection models classifiers to perform detection. Instead, YOLO frames object detection as a regression problem to spatially separated bounding boxes and associated class probabilities.

Current object detection systems make the use of image classifiers to perform object detection. In such systems input is taken as an image and a classifying algorithm is evaluated in the image at various locations and scales to perform detection. Most of these approaches use a sliding window technique where the classifier is run at evenly spaced locations over the entire image.

In YOLO, object detection is reframed as a single problem which processes the image from image pixels to bounding boxes along with class probabilities. A single convolutional network predicts multiple bounding boxes in the image along with giving the class probabilities for those boxes.

YOLO is extremely fast. There is no need of any complex pipeline architecture here because the detection is framed as a regression problem. The neural network can simply be run on any new image to predict detection. It is possible to process streaming videos in real time to detect the objects present with less than 25 milliseconds latency. Also, YOLO takes the entire image into considerations unlike the earlier algorithms which use sliding window based techniques. Due to this YOLO encodes the contextual information about the classes along with their background. Unlike Faster R-CNN which mistakes the patches in backgrounds for objects, YOLO avoids such erroneous detection by seeing the larger context that is the objects with their backgrounds. Another advantage of YOLO is that it can be highly generalized. Hence, it is less likely to break down when applied to new domains or when unexpected input is given to it.

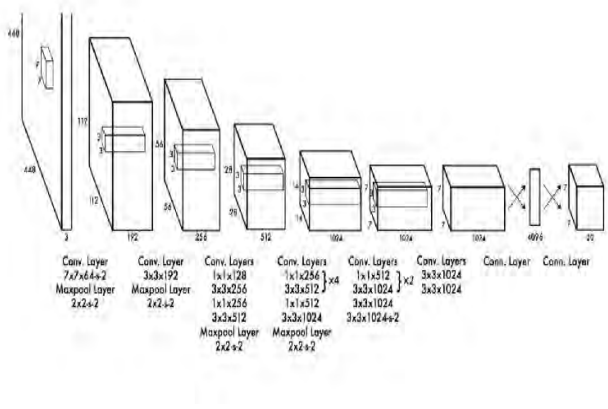


Fig 8: YOLO framework for object detection.

III. COMPARISON OF OBJECT DETECTION ALGORITHMS

We analyze the aforementioned research papers and found out some key points of comparison. Here we try to present the results obtained from the papers in a tabular format.

This will help us to compare each attribute of the various object detection techniques side by side. It will also aid us in knowing which methods are better for which application.

Table I. Comparison between different object detection algorithms

	Training Time	**Testing Time**	**Usable in real time detection**	**No. of feature maps**	**Accuracy**	**Prediction time**	**Section of image accessed**
SSD	Fastest	Fastest	Yes	Multiple	Higher than Faster R-CNN	Very small	Whole image
R-CNN	Large	Large	No	1	Low	Too large	Part of image
Fast R-CNN	Less than R-CNN	Less than R-CNN	No	1	Low	Large	Part of image
Faster R-CNN	Lesser than Fast R-CNN	Lesser than Fast R-CNN	Yes	1	Highest	Small	Part of image
R-FCN	Lesser than Faster R-CNN	Lesser than Faster R-CNN	Yes	1	High	Large	Part of image
YOLO	Lesser than Faster R-CNN	Lesser than Faster R-CNN	Yes	1	High	Very small	Part of image

IV. CONCLUSION

This paper aims to provide an insight into various object detection algorithms such that it is possible to view all their aspects side by side. However, it is very difficult to decide what the best approach for a particular application is as factors such as running time and memory usage of the servers are also critical along with standard accuracy metrics such as map. It is found that fewer proposals for Faster R-CNN can speed it up substantially without much compromise of the accuracy. It is also seen that the performance of SSD is less sensitive to the quality of

the feature extractor as compared to Faster R-CNN and R-FCN. It is observed that R-FCN and SSD models are faster on average while Faster R-CNN tends to lead to slower but more accurate models.

However it is also true that there are many factors such as object size, image size, number of proposals, feature extractor used etc. which affect the performance of these models to a great extent[11]

REFERENCES

[1] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C., 2016,October. Ssd: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham.

[2] Girshick, R., 2015. Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).

[3] Ren, S., He, K., Girshick, R. and Sun, J., 2015. Faster RCNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99)

[4] Dai, J., Li, Y., He, K. and Sun, J., 2016. R-fcn:Object detection via region-based fully convolutional networks. In Advances in neural information processing systems (pp. 379-387).

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. arXiv preprint arXiv:1506.02640, 2015.

[6] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg. Dssd: Deconvolutional single shot detector. arXiv preprint arXiv:1701.06659, 2017

[7] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov. Scalable object detection using deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2147–2154, 2014.

[8] S. Zagoruyko, A. Lerer, T.-Y. Lin, P. O. Pinheiro, S. Gross, S. Chintala, and P. Dollar. A multipath network for object detection. arXiv preprint arXiv:1604.02135, 2016.

[9] C. Szegedy, A. Toshev, and D. Erhan. Deep neural networks for object detection. In Advances in Neural Information Processing Systems, pages 2553–2561, 2013.

[10] Wang X, Shrivastava A, Gupta A: A-Fast-RCNN: Hard Positive Generation via Adversary for Object Detection. arXiv preprint arXiv:1704.03414

[11] J Huang, V Rathod, C Sun, M Zhu, A Korattikara, A Fathi, I Fischer, Z Wojna, Y Song, S Guadarrama, K Murphy: Speed/accuracy trade-offs for modern convolutional object detectors. arXiv preprint arXiv:1611.10012

[12] N Yadav , U Binay: Comparative Study of Object Detection Algorithms. Volume: 04 Issue: 11, Nov 2017, International Research Journal of Engineering and Technology (IRJET), (pp. 586-591).

Automatic BrainTumor detection on T1- weighted MR image using K-mean clustering and HCSD

Akshay A. Jain1 Rajashree Gadhave2
1BE, dept. of Computer Engineering, PHCET, Rasayani, master1spidy@gmail.com

2Prof., dept. of Computer Engineering, PHCET, Rasayani, rajashree86@gmail.com

Abstract— With the advancement in medical field, use of technology has increased tremendously. Brain tumour is one of the most tedious thing to detect accurately without the help of technology. Magnetic Resonance (MR) allows to localize a brain tumour tissue and a mass of abnormal cells in a slice. This process automation is useful for experts for post processing and extraction of region in interest like tumour identification and segmentation. Techniques in MR images contain various processing modalities that scans and store the insidious structure of brain. This paper introduces a novel technique that uses two-step procedure; the HCSD (Hierarchical centroid shape descriptor) and K-mean clustering. The Hierarchical Centroid Shape Descriptor selects only those with specific shape. The stage of clustering is applied to recognize a distinction between structures with specified pixel intensity. Then the process automatically places a bounding box to describe the region precisely where the tumour was found. An expert performed the distinction to exactly locate the tumour; he reached accuracy of 91% us9ing Dice coefficient. He tested on 254 T1-weighted Magnetic Resonance Image of 18 patient having brain tumour.

Keywords— ***Brain tumour detection, HCSD,K-mean algorithm, magnetic resonance.***

I. INTRODUCTION

MRI is a standard Computed Radiography particularly used in medical for diagnosis and curing it. It's an advantage to be a non-introductory of instrument into human body and spreading undesirable. Early stage of tumor detection in brain of patient lead's to saving life with proper treatment. With the incremental flow of medical data, it is been fastest method in detection of tumors using MRI slices. MRI not only helps medical experts in identifying brain tumor but also for creation of segmentation and 3D construction. This paper introduces the method for identifying the tumor with all detailed information and reduces work by doing it automatically in preoperative images. The visual inspection and portraying it precisely will be limited which will consume less time for medical experts.

Critical decision are been lead due to automatic detection and segmentation of tumour play a vital role in it. From few years, many works concentrated on this problem, but still it's not solved completely.

This work of our contributes automatic detection of brain tumour using T1-weighted MRI by using vigorous technique for preparation of shape variation, size, texture, pixel intensity and tumor position. Algorithm like k-mean is associated with shape based feature in hierarchical descriptor. The preprocessing is done to remove skull and take only the brain.

Anatomy of brain is classified in three groups. If physical or mental disease tissues such as tumors appear, the number of groups increase to four and the gray matter (GM), White Matter (WM), Cerebrospinal Fluid (CSF) and tumor.

II. RELATED WORK

Some authors have showed that detection automatically can be done by referring features like shape, texture, symmetry and intensity. Tumours are of different types which leads to in homogeneity to their texture, shape, position and intensities, the challenging task is to detect abnormal tissues automatically.

The detection of brain tumour automatically was presented by [5] that deals with the methods of modified texture based region growing and cellular automatic edge detection. A pattern based model was used to extract tumour texture was proposed in [6]. But only the use of texture was not sufficient, some genuine data does not have correct texture feature which leads to unsatisfying results. Numerous types of tumour texture can be confusing with other tissues, considered alone.

Algorithm such as Fuzzy C means is used in [7] for classifying brain tumour images and it works well for tumour detection. Like the mean-shift [8], this method has a high computation complexity. In [9], the brain symmetry was used for tumour segmentation and detection by using the texture and intensity. The application of this kind of feature is limited on axial and coronal planes because there are no symmetric structures in the sagittal plane. The authors in [11] propose to classify brain image as normal or abnormal by using neural network.

III. MATERIALS AND METHODS

The involvement of physical or mental disease in the brain leads to inappropriate results. As done in various works, thresholding technique and the morphological operations such as erosion, dilation, closing and opening gets involve in pre-processing for removing of skull.

To achieve the ineffectiveness in algorithm to detect and localize tumor in human brain automatically, we have use the k-means clustering technique in order by the selection of shape that perfectly detail the tumor. The selection of k-mean algorithm was practice with a prior knowledge of clustering of brain tissues and its less

computational complexity. The algorithm Otsu was tested on first stage and provided results are close to those got with k-means algorithm. By calculating the intensities based on the pixels, the high intensity clusters are highlighted structures like tumor. The threshold value used for image which is binarized in order to allow only pixel in Kth cluster. Some studies shows that this method is sufficient for detection of brain tumor with many time in real data, always abnormal tissues are not selected. By this another stage is required in brain tumor detection.

In the cluster of interest, misclassified healthy tissues are discriminated. To select only tumor structure, a flexible descriptor based on kd-tree decomposition is used. The technique used for extraction centroid-based tree structure feature.
The obtained vector from k-d tree decomposition process is defined as

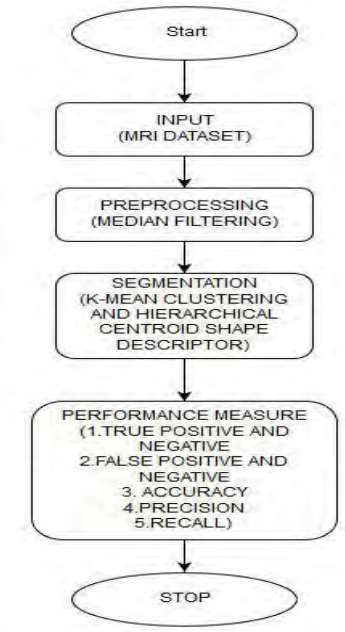


Fig:- Proposed Brain tumor detection algorithm

$$\tilde{X} = (a_0^0\, b_0^1\, b_1^1\, a_0^2\, a_1^2\, a_2^2\, a_3^2\, b_0^3\, b_1^3 \ \dots\dots\ b_7^3) \qquad (1)$$

Where a_0^0 shows the root level coordinate, the m*th* y coordinate at level *n* and the m*th x* coordinate at level *n*

A. k-mean Algorithm

This algorithm was introduced and received many contribution and it's one of most popular clustering algorithms. Given y_I data of vector(i = 1…n), the categorization of its value in *k* clusters began by randomly describing k points as centroid of every group in the space of data. By the repetitive process the elements are associated to the closest center in k classes. The group means are increased by considering the newest element to each of them. This technique minimize the function detailed as the equation sum of squared errors.

B. Hierarchical Cnetroid Shape Descriptor

The Hierarchical centroid shape descriptor is binary descriptor build with centroid coordinates removed from a binary image and based on the kd-tree technique rotting. The HCSD is a shape remover that continuously by rotting the image in sub-images. Spatial distribution of pixels are used to describe the image, by using kd-tree algorithm in the pixel domain by decomposing image. The neighboring information such as centroid coordinates of local regions is removed. The descriptor length is 2 * (2_d – 2) where d is the depth of the feature extraction process. HCSD is built as follows

1.Proceed with input as *I* and calculate the transpose *IT,*
2. Compute for each input, centroid C(Xa, Xa) to the root level by using formula.
3. Divide continuously the image in two different images based on the centers of gravity($x = x_c$ or $y = y_c$)
4. The obtained vector is in the range of [-0.5,0.5], the point 0 tends to the centroid of the root level. The values which are negative are features from tree decomposition of the left part of the image, positive shows to the right side.
5. Combine the features removed from the image I and I_T.

IV. IMAGES ILLUSTRATING THE FLOW

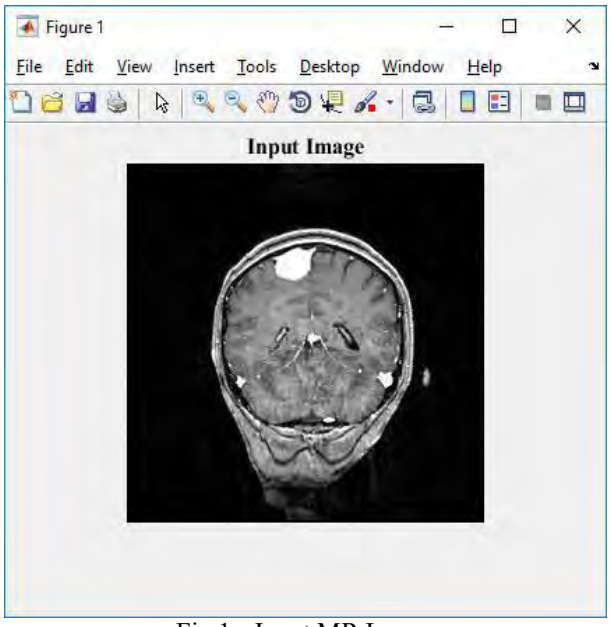


Fig 1:- Input MR Image

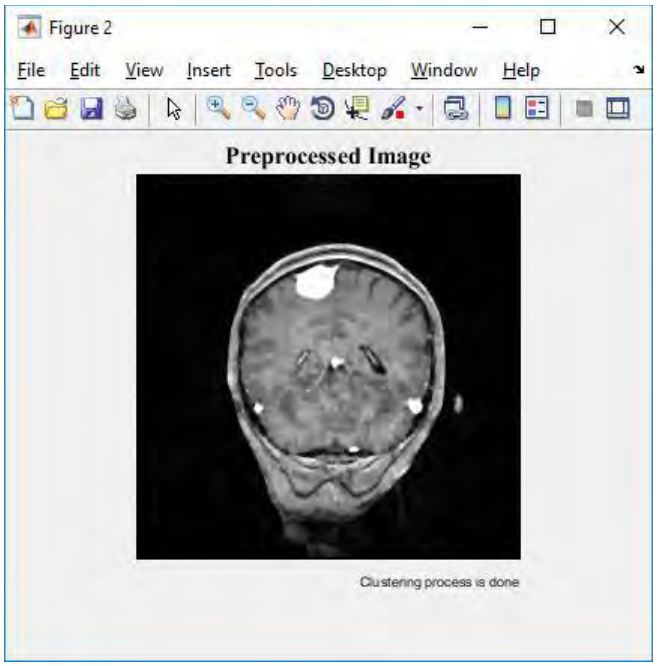


Fig 2:- Preprocessed Image(Median filtering)

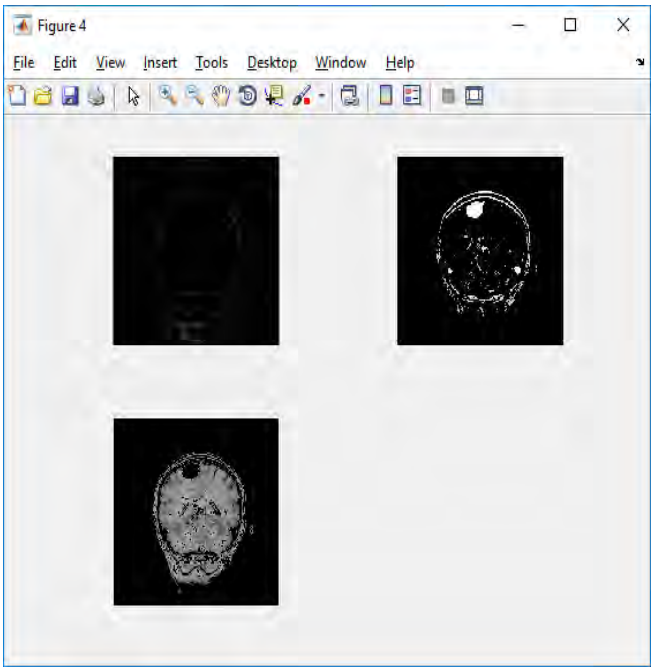


Fig 4:- Hierarchical Clustering

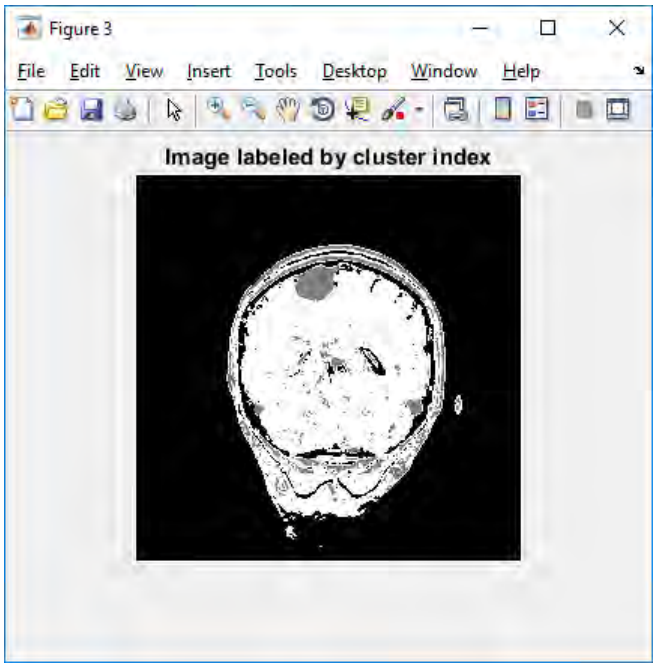
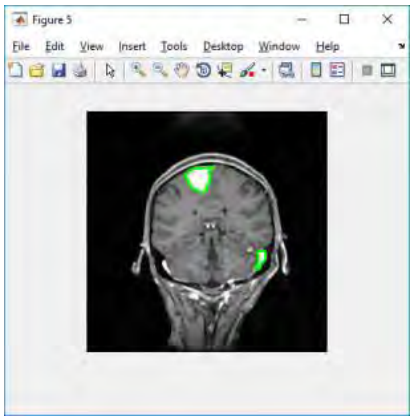


Fig 3:- Cluster Index



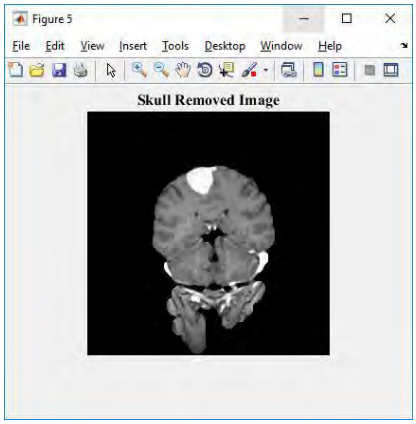


Fig 5:- Skull removed image

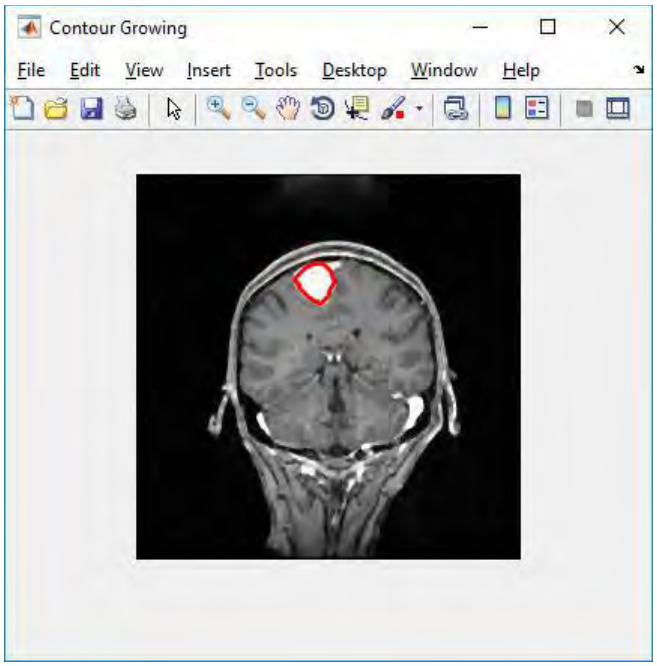


Fig 6:- GLCM

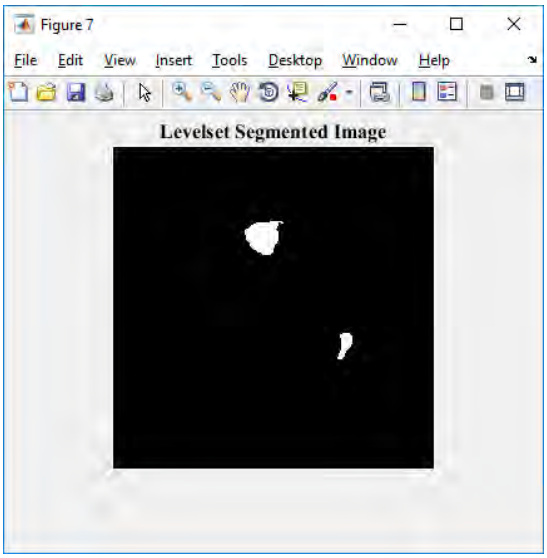
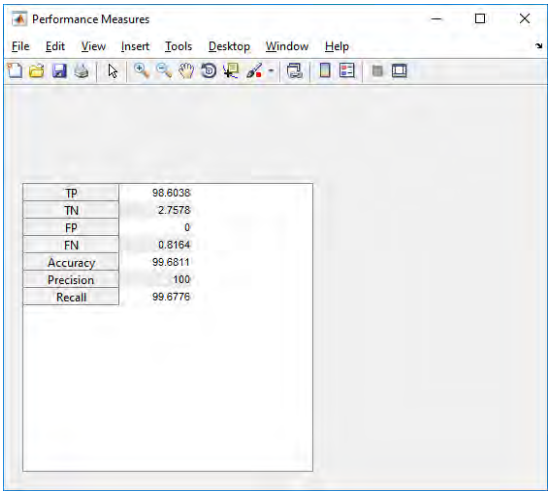


Fig 7:- HCSD



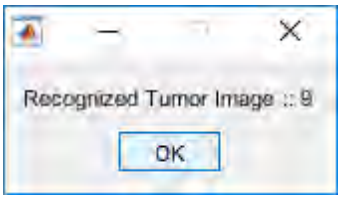


Fig 8:-Performance Measure

V. RESULT

The obtained results of proposed technique are viewed in this section. The execution was done with Intel i5 processor,2.4 Ghz and 4 GB of memory using Matlab v.2016 tool. Test performed were done on set of 254 T1-weighted MRI having brain tumors. These images have been taken from open neuro science and Department of Neurosurgery, University of Leipzig, Germany.

The Hierarchical Centroid Shape Descriptor technique performed the obtained results with one-step method like as k-means or Otsu Multilevel thresholding. The mark reached by this technique is of 0.9245 with Jaccard index and 0.95 based on the Dice index, and it illustrate that the presented technique is effective approach for brain tumor detection.

VI. CONCLUSION

In this paper, we introduced a two-step method for detection of brain tumor tissues. K-mean and HCSD method were combined for perfectly identifying the shape and size of the tumor. In the first step, k-means groups the pixels of the image in k clusters, after that image is binarized by threshold value equal to k. In the remained binary structure tumor was found but with the healthy tissues surrounded to it. In the second stage it was discarded in order to detect the tissues belonging to the tumor only. The technique which was experimented have given the desired results which means the method is robust in detecting and adjoining the abnormal tissues in MRI rather than the in homogeneity intensity or the complex shape of the brain tumor. This approach is vigorous to the shape change

References

[1] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, “A survey of MRI-based medical image analysis for brain tumor studies,” Physics in Medicine and Biology, vol. 58, no. 13, pp. 97 – 129, 2013.

[2] S. Aswathy, G. Glan Deva Dhas, and S. Kumar, “A survey on detection of brain tumor from MRI brain images,” in Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014 International Conference on, July 2014, pp. 871–877.

[3] K. Somasundaram and T. Kalaiselvi, “Automatic brain extraction methods for t1 magnetic resonance images using region labeling and morphological operations,” Computers in Biology and Medicine, vol. 41, no. 8, pp. 716 – 725, 2011.

[4] S. Ghanavati, J. Li, T. Liu, P. Babyn, W. Doda, and G. Lampropoulos,

"Automatic brain tumor detection in magnetic resonance images," in Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on, May 2012, pp. 574–577.

[5] S. Charutha and M. Jayashree, "An efficient brain tumor detection by integrating modified texture based region growing and cellular automata edge detection," in Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014 International Conference on, July 2014, pp. 1193–1199.

[6] A. Islam, S. Reza, and K. Iftekharuddin, "Multifractal texture estimation for detection and segmentation of brain tumors," Biomedical Engineering, IEEE Transactions on, vol. 60, no. 11, pp. 3204–3215, Nov 2013.

[7] R. Preetha and G. Suresh, "Performance analysis of fuzzy c means algorithm in automated detection of brain tumor," in Computing and Communication Technologies (WCCCT), 2014 World Congress on, Feb 2014, pp. 30–33.

[8] C. Farmaki, K. Mavrigiannakis, K. Marias, M. Zervakis, and V. Sakkalis, "Assessment of automated brain structures segmentation based on the mean-shift algorithm: Application in brain tumor," in Information Technology and Applications in Biomedicine (ITAB), 2010 10th IEEE International Conference on, Nov 2010, pp. 1–5.

[9] T. E. M. A. Bianchi A, Miller JV, "Brain tumor segmentation with symmetric texture and symmetric intensity-based decision forests," in Proceedings / IEEE International Symposium on Biomedical Imaging: from nano to macro IEEE International Symposium on Biomedical Imaging., 2013, pp. 748 – 751.

[10] K. W. . B. K. Dvok, P., "Automatic brain tumor detection in t2weighted magnetic resonance images." Measurement Science Review, vol. 13, no. 5, pp. 223 – 230, 2013

[11] Y. Zhang, Z. Dong, L. Wu, and S. Wang, "A hybrid method for MRI brain image classification," Expert Systems with Applications, vol. 38, no. 8, pp. 10 049 – 10 053, 2011.

[12] E. Ulku and A. Camurcu, "Computer aided brain tumor detection with histogram equalization and morphological image processing techniques," in Electronics, Computer and Computation (ICECCO), 2013 International Conference on, Nov 2013, pp. 48–51.

[13] P. Dhage, M. Phegade, and S. Shah, "Watershed segmentation brain tumor detection," in Pervasive Computing (ICPC), 2015 International Conference on, Jan 2015, pp. 1–5.

[14] M. Ozic, Y. Ozbay, and O. Baykan, "Detection of tumor with otsu-pso method on brain mr image," in Signal Processing and Communications Applications Conference (SIU), 2014 22nd, April 2014, pp. 1999–2002.

[15] S. Satheesh, R. Kumar, K. Prasad, and K. Reddy, "Skull removal ofnoisy magnetic resonance brain images using contourlet transform and morphological operations," in Computer Science and NetworkTechnology (ICCSNT), 2011 International Conference on, vol. 4, Dec 2011, pp. 2627–2631.

[16] J. Chiverton, K. Wells, E. Lewis, C. Chen, B. Podda, and D. Johnson, "Statistical morphological skull stripping of adult and infant MRI data," Computers in Biology and Medicine, vol. 37, no. 3, pp. 342 – 357, 2007.

[17] B. Dogdas, D. W. Shattuck, and R. M. Leahy, "Segmentation of skull and scalp in 3-d human MRI using mathematical morphology data," Human Brain Mapping, vol. 26, no. 4, p. 273285, 2007.

[18] K. Somasundaram and T. Kalaiselvi, "Automatic brain extraction methods for t1 magnetic resonance images using region labeling and morphological operations," *Comput. Biol. Med.*, vol. 41, no. 8, pp. 716–725, 2011.

Proposed Software Process Model and Template for Integrating Disparate Systems

Dr. Nazneen Ansari, Shraddha Pawar
Department of Information Technology
St. Francis Institute of Technology
Mumbai, India
naz2ensfit@gmail.com
shraddha.pawar400060@gmail.com

***Abstract*—To rationalize business processes and enhance productivity in business, organizations globally invest in software systems like ERP, supply chain management (SCM), and other BI software that are owned by different vendors and run on different platforms. The usefulness of such software systems can only be comprehended if they are suitably integrated with one another. There exists a lot of scope for integrating such systems. Instead of developing a contemporary system, there should be a template for integrating disparate systems into a single system. Additionally, the integration of disparate software systems not only saves time but also money. The prime focus of integrating disparate systems is to create communication among them, which is a challenging task because different systems possess different configurations and function on different platforms that use different technologies. This paper proposes software process models for integrating disparate systems. Moreover, it proposed a template with 35 steps to merge disparate software solution.**

Keywords—disparate software systems, software process model, template

I. Introduction

Organizations own legacy software systems that are considered to be esteemed resources and substantial to existing business setups. In today's world, a major challenge that is faced by organizations is the need to competently and successfully integrate legacy software systems with other internal and external business processes. Rahimi et al. [1] state that information and communication technology expansion has considerably strengthened global market competition. For example, the use of enterprise resource planning (ERP) systems has ascertained to reinforce process integration and mechanization, performance improvement, and cost reduction. ERP consists of organizing and handling an organization's assets efficiently and productively. Business intelligence (BI) is a quintessential way to encourage an organization's core competency. It is an array of know-hows that expedite data collection, integration, and analysis in an organization thereby enabling the process of decision making. If ERP and legacy systems have previously been implemented in an organization and the organization attempts to integrate a BI system, then ERP–BI integration would be extremely challenging due to a variety of factors. Nonetheless, challenges can be surmounted through appropriate planning, and organizations can acquire enormous benefits by investing in both ERP and BI systems.

The rest of this paper is organized as follows. Section II provides an overview on software engineering (SE) process model. Section III discusses proposed software process model for integrating disparate systems. Section IV describes 35 steps for merging two disparate software solutions through a template. Section V concludes the study.

II. Literature Survey

Software systems act as a backbone in today's digital world. Right from number crunching applications that ran on main frame computers (colloquially referred to as big iron) to today's web applications that handle massive warehouses of data, software systems have gained considerable popularity and importance. Software engineering is the product of a systematic and quantifiable approach for the development and maintenance of software [2].

A software engineering procedure defines "how" to perform a task and is normally applied to a particular role. In other words, SE procedures outline "how" to perform tasks and associations among those tasks. There are numerous advantages of SE procedures, such as providing visibility in terms of quality, schedule, and cost; improving communication and understanding; abetting in scheduling and plan execution; providing the need to capture lessons learned; facilitating the execution of organization-wide processes; and providing a foundation for training and assessment.

Software process, in other words, denotes a collection of activities that are needed for producing a software system, which is implemented by a team of experienced people that is organized as per a prearranged organizational structure based on techno-conceptual tools.

An intellectual representation of a software process is known as a software process model, and the practice of representing a software process is called software process modeling. IEEE 1974–1991 [3] and International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 12207 [4] are the two main international standards that recommend processes for evolving and conserving software. Both these standards regulate a group of vital, albeit disorganized activities, which need to be finalized for acquiring a software product. In other words, a software process model is nothing but a conceptual illustration of style, blueprint, or distinctness of a software process [5].

III. Proposed Model

A. Proposed software process model for integrating disparate systems

A five-step software process model is proposed that consists of phases such as planning, define, design, build, and deploy, as

shown in Figure 1. Detailed explanation of the proposed model is as follows [6]:

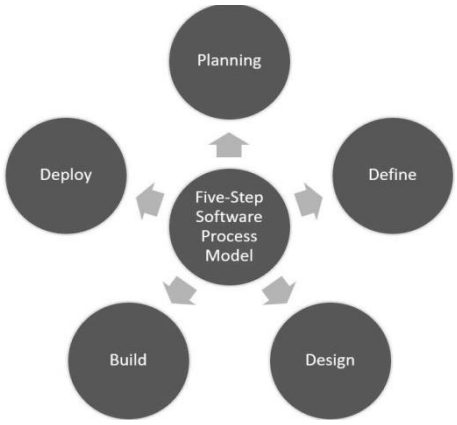


Fig. 1. Proposed software process model for integrating disparate systems

1) Planning: In this phase, a foundation for a recommended project is laid. During the laying of foundation, several factors such as price, advantages, and assessment structure for evaluating integration technologies and solutions must be considered.

This phase includes the following activities:

a) Functional requirements analysis: It aims at providing an analysis of proposed data exchange points for determining the needs that are to be met for each integration point.

b) Non-functional requirements analysis: It aims at identifying necessities for integration functionality. It considers parameters like execution, error handling, and supervising necessities for business dealings and interfaces.

c) Application analysis: It aims at capturing the important details of applications like operating location, technologies, and application interfaces.

2) Define: It is mainly used to identify a project's scope, goals, and objectives. It also performs other functions like performing comprehensive business process analysis, identifying procedural needs, and arranging for software quality assurance activities.

3) Design: This phase is used for developing analytical design, selecting data format, implementing data plotting, defining schema, and developing reference architecture. This phase aims to create models where construction, solutions for integration, and required amendments to current systems are established.

The activities that are considered under this phase are as follows:

a) Message design: It defines data that is delivered between systems, data format, and the way of transforming data from one process to another.

b) Choosing of appropriate integration solution: There exist several technologies for performing integration under differing circumstances as distinguished by perceptible features of detailed data exchange.

While selecting technology for a particular case, different factors have to be considered such as:

a) Data format: It is essential that applications that are integrated should adopt suitable data format. However, in certain cases, if prevailing applications disagree on a particular data format and demand another data format, then an intermediate decoder should exist to maintain standardization.

b) Data timeliness: Note that data timeliness is a must while performing integration. It is necessary that when an application shares some data to another application and that application already possesses that data, then, in such cases, integration must significantly reduce time interval between the two applications so that integration occurs faster.

4) Build: This phase mainly includes the development of software codes for objectifying any alterations to affected systems and for integration purposes.

The activities that are performed under this phase are as follows:

a) Selection of a specific technology: Several techniques of integration need countless dedicated hardware as well as software, which could be costly, produce seller lock-in, and upsurge learning curve. Conversely, the creation of an integration solution devoid of tools may signify reinventing the wheel.

b) Integration of applications: Dependency between applications should be minimized so that each application progresses without instigating problems to other applications.

c) Intrusiveness: Software developers should not only reduce modifications to applications that need to be integrated in an organization but also reduce integration code. In addition, they must also develop integration facilities, fulfill technical documentation, and execute test cases.

5) Deploy: With the help of quality assurance/user acceptance testing, this phase mainly checks and ensures whether integrations that were stated in the define phase meet desired requirements, and only after meeting quality standards, it deploys integrations in a production environment.

IV. TEMPLATE

A. Steps for merging two disparate software solutions through a template

Vicens et al. [7] state that for developing an integrated system, methodology, architecture, and tools are wanted. Without considering them, it would really be difficult to develop an integrated system. Project development takes place through methodology, architecture is responsible for providing necessary elements that are desirable for an enterprise structured description in an enterprise integration context, and tools support the development of methodological concepts.

First, we begin merging of two disparate software solutions by examining their requirements and understanding their scope. Moreover, we use a cost-effective integration method that is uncomplicated and has a direct approach in satisfying integration objectives. Next, we obtain a solution that uses the systems' built-in capabilities and/or off-the-shelf alternatives for meeting the requirements [8]. We then identify different approaches to integration and evaluate or compare them based on parameters such as technical expertise, time taken to design, implementation, cost, and flexibility.

B. Thirty-five steps for merging two disparate software solutions are as follows [9]:

- ***Step 1:*** Start an integration project.
- ***Step 2:*** Develop an integration plan by highlighting on sufficient amount of execution details, i.e., concentrate on integration points and ETL approaches. Note that E and T approaches should take place at source level rather than target level, whereas L approach can take place at target level.

- ***Step 3:*** Data points and sources that need to be integrated must be identified by an integration team in order to avoid partial integration failures.
- ***Step 4:*** An integration team should understand and identify key business processes in a business area and give less priority to least important processes [10].
- ***Step 5:*** An integration team should give higher priority to logical flow implementation that is defined in terms of functional, operational, and integrated business flows.
- ***Step 6:*** An integration team must understand and identify suitable data structures. For example, many integration projects are integrated at data level, and so, it necessary to understand and identify appropriate data structures and descriptive data schemas for determining which data streams coincide in an integrated system and where and why to use these data for the purpose of data mappings and transformation.
- ***Step 7:*** An integration team must thoroughly understand all processes for acquiring an accurate functional model in an integrated system. Moreover, project definition is reliant not only on how business processes are noticed but also on the accurateness of data model definition.
- ***Step 8:*** An integration team should give more importance to developing methods and technologies in place of work processes.
- ***Step 9:*** UI plays a vital role in understanding integration points between disparate systems.
- ***Step 10:*** It is necessary to understand events in an integrated system because they are responsible for the accuracy of functional model. Events can be synchronous and asynchronous. Asynchronous events are responsible for automated enterprise integration solution.
- ***Step 11:*** A specification of integration points and a list of data sources need to be preserved. Before the commencement of a project and resource reservation, a thorough plan is a must.
- ***Step 12:*** Appropriate use of information technologies for developing an integrated solution is essential. Even if a mistake occurs in selecting a suitable technology, a lot of time would be wasted, which is precious. Therefore, it is advisable that a pilot project should be built first before the actual one.
- ***Step 13:*** Understand the software applications that need to be integrated.
- ***Step 14:*** Understand the technical features of the software applications that are to be integrated.
- ***Step 15:*** Understand the architecture of two disparate software systems (for example, ERP and BI).
- ***Step 16:*** Identify the deployment type of both ERP and BI systems [11].
- ***Step 17:*** Identify where to fit or connect the two disparate systems in the architecture (for example, identify where to fit or connect BI in ERP architecture).
- ***Step 18:*** An integration team should give priority to the prototype installation of intermediate versions of integrated system before running the final installation and launching it in product environments.
- ***Step 19:*** An integration team must take into account the following parameters [12]:

  1) Extent of integration: Even though rigid integration is a perfect solution, cost and financial plans might affect the extent of integration. For instance, it is profitable to select a loosely coupled integration approach for an integration system that does not require tightly coupled integration approach. Tightly coupled integration approach is apt for transactions that have a serious effect on other systems.

  2) Affluence of maintenance: While selecting integration approaches, it is necessary to keep in mind the affluence of maintenance. Even if the implementation of some integration approaches may be easier, cost and labour required will be more so that a system may last for a longer time.

  3) Access control and data security: Security must be present at all levels of system integration. Only those users must be given access to data that own appropriate credentials.

  4) Budget: Costs and benefits differ for different integration approaches. Nonetheless, an appropriate integration approach should be considered on the basis of efficiency rather than cost.
- ***Step 20:*** Identify integration system requirements.
- ***Step 21:*** Identify the modules of two systems that are to be merged. Tabulate the important modules of application as well as integration direction, as shown in Table I.

TABLE I. APPLICATION INTEGRATION DIRECTION

Application A	***Integration Direction***	***Application B***
Modules X	One way	Modules P
Modules Y	Two way	Modules Q
Modules Z	--	Modules R

- ***Step 22:*** Reduce both integration codes and application modifications as accentuated in the tactlessness of integrated applications [13].
- ***Step 23:*** Summarize integration requirements and define integration point.

  An integration point is data cross-section between two integrated systems where data is shared.

  Important factors for integration points:

  • *Source:* Where is the data coming from?

  • *Target:* Where is the data going?

  • *Direction:* Which direction does the data need to flow? Either one- or two-way.

  • *Frequency:* How often is the need for an integration point? Hourly, nocturnally, on-demand, real time, etc.
- ***Step 24:*** Understand the purpose of each integration point (i.e., how integration is executed: real time, manual, or scheduled?).

  Table II shows the details of integration points by considering the above factors.

TABLE II. DETAILS OF INTEGRATION POINTS

Integration point	***Source***	***Target***	***Direction***	***Frequency***	***Key fields***	***Notes***

- ***Step 25:*** Identify integration architecture.
- ***Step 26:*** Identify integration methods.

  Identify different integration approaches and compare them based on parameters such as technical expertise, time taken to design, implementation, cost, and flexibility. Table III lists different integration approaches.

TABLE III. COMPARING INTEGRATION APPROACHES [14]

Integration solutions	*Technical expertise*	*Time taken to design*	*Implementation*	*Cost*	*Flexibility*
Approach 1	++	+++	++	+++	+++
Approach 2	++	+++	++	+++	+
Approach 3	+++	+	++	++++	++++
Approach 4	++	+++	+	+++	+++

++++ Optimum, +++ Good, ++Average, and + Poor

- ***Step 27:*** Select appropriate integration tools.

  Select appropriate integration tools by considering features such as cost, speed, UI, scheduling, scalability, and customization options. Furthermore, select the most appropriate integration tool by taking into account the abovementioned features.

  Table IV shows an example for selecting appropriate integration tools from the aforesaid features.

TABLE IV. SELECTION OF APPROPRIATE INTERGRATION TOOLS

Tools	*Features*	*Notes*
SQL	High speed	Requires SDK
Smart Connect	Strong UI and low to medium cost	

Integration tool task

• *Connectivity:* Technical links between applications.

• *Data transformation:* Data conversion format and significance between applications.

• *Data integrity:* To ensure data is consistent between applications with required performance metrics.

- ***Step 28:*** Identify and select appropriate integration platform and technology.
- ***Step 29:*** Selection of technology leads to various integration styles because diverse integration methods need diverse software and hardware [15]. Selected tools can be costly and cause vendor lock-in, which can ultimately increase a developer's learning curve. The other point that needs to be considered is data format in which integration style is done.
- ***Step 30:*** Integrated solution must handle the following parameters [16]:
  - *Interoperability*: In an integrated solution, various components are integrated that use different platform, language, and operating system (OS), and so, it is essential that all these components are interoperable.
  - *Data integration*: When data flows from one application to another in an integrated solution, data format that is used may be different, and so, it is necessary to integrate data by changing it to a suitable format as per the system.
  - *Robustness, stability, and scalability*: Since all components are present together as one solution, it is necessary that the integrated solution must be extremely robust, stable, and scalable.
- ***Step 31:*** Test the integrated solution by creating sufficient test plans.
- ***Step 32:*** After an integration solution runs, an integration team should identify who can update the data? Does there need to be something in place to block updates to certain data?
- ***Step 33:*** Consider the system performance and test the system before it is used in real time.
- ***Step 34:*** Determine the business value of the integrated system.
- ***Step 35:*** Determine various ways of maintaining the integrated system, i.e., maintenance effort and support are vital in software integration solution, which needs to be treated separately.

V. CONCLUSIONS

This research work proposed a five-phased software process model which includes planning, define, design, build, and deploy for coupling disparate enterprise systems. we created a template with thirty-five steps to merge two disparate software solutions. This template covers all information required from the beginning to the end in integrating disparate software systems. Based on the template, organizations can make a suitable plan for integration.

REFERENCES

[1] Rahimi, E., & Abbasi Rostami, N. (2015). Enterprise Resource Planning and Business Intelligence: The Importance of Integration. *International Journal of Management Academy*, 3(4), 7-14.

[2] Bakshi, A., & Singh, R. (2013). Component based development in software engineering. *Int. J. Recent Tech. Eng.*, 2(1), 48-52.

[3] IEEE Standard for Developing Software Life Cycle Processes, *IEEE Standard 1074-1991*.

[4] Singh, R. (1996). International Standard ISO/IEC 12207 software life cycle processes. *Software Process Improvement and Practice*, 2(1), 35-50.

[5] Asmita, Kamlesh, Usha. Review On Comparative Study Of Software Process Model. *International Journal of Science, Technology & Management*, vol. no 04, Special Issue No. 01, March 2015.

[6] Moturi, Christopher A., Gabriel Githehu Kinu, and Andrew Mwaura Kahonge. Process Model for Enterprise Application Integration: Case for a Customs Department. *International Journal of Applied Information Systems* 6.2 (2013): 1-16.

[7] Vicens, E., Alemany, M. E., Andres, C., & Guarch, J. J. (2001). A Design and Application Methodology for Hierarchical Production Planning Decision Support Systems in an Enterprise Integration Context. *International Journal of Production Economics*, 74(1), pp. 5-20.

[8] AFDteck Solutions, Systems Integration Overview, [Online] Retrieved from http://www.afdtek.com/downloads/whitepapers/White_Paper-Integration_Overview.pdf

[9] Ansari, N., & Thampi, G. T. (2016, April). Investigation of different approaches, feasible solutions, and estimates for disparate enterprise solution synchronization. *In Computing, Communication and Automation (ICCCA), 2016 International Conference on* (pp. 214-218). IEEE.

[10] Surugiu, I. (2012). Integration of Information Technologies in Enterprise Application Development. *Database Systems Journal*, 3(1), 21-31.

[11] 147. Microsoft Dynamics GP, Connector for Microsoft Dynamics for Microsoft Dynamics GP, [Online] Retrieved from https://mbs.microsoft.com/customersource/northamerica/GP/downloads/service-packs/mdgp_dynamicsconnector

[12] Esichaikul, V., & Nuankhieo, P. (2004). Connectivity of ERP with Legacy Systems. *Journal of Information and Organizational Sciences*, 28(1-2), 165-174.

[13] Ramadhani, A., Sam, A., & Kalegele, K. (2017). Framework for integrating fragmented information systems: Case of livestock information systems. *International Journal of Computer Engineering Research*, 7(1), 1-10.

[14] http://www.panduit.com/ccurl/135/838/PanduitCW_WPinfra_0922.pdf

[15] Klemm, D. M., & Chang, R. A. (2006). U.S. Patent No. 7,124,413. Washington, DC: U.S. Patent and Trademark Office.

[16] Understanding Enterprise Application Integration: The Benefits of ESB for EAI, [Online] Retrieved from https://www.mulesoft.com/resources/esb/enterprise-application-integration-eai-and-esb

Solar based laptop charger

Bhagyashri N Bachhav, Darshana A Deore, Pranjali B Deore, Sarika M. Jagtap

Electronics & Telecommunication NDMVP's KBTCOE, Nashik
pranjalideore1997@gmail.com, pranjalideore1997@gmail.com, pranjalideore1997@gmail.com, jagtapsarika@kbtcoe.org

***Abstract*—The objective of this research is to design a Solar Powered based Portable Power Bank for laptop using sunlight as its ultimate power, which can be used effectively during disaster events. It has in-built solar panel, which converts the solar energy to electrical energy. The charge is then transferred to a battery for storage of charge for further use, with the battery. When battery voltage is sufficient, then this voltage is used to charge a laptop, when battery voltage is not sufficient laptop charges through AC mains. Tilt and IR sensors are used for security purpose. If system is ok, authentication is check by using RFID tag, and respective person can charge a laptop for respective time and if authentication fails then person is not allow charging a laptop. When system has any problem then message is send to owner by using GSM module.**

Keywords—Solar panel, GSM module, PIC16F877a, TILT and IR sensor, Power supply.

I. INTRODUCTION

The consumption of non-renewable sources like oil, gas and coal is increasing at an alarming rate. The time has finally come to look after some other renewable sources of energy i.e. solar, wind and geothermal energy. Solar power is a renewable source of energy, which has become increasingly popular in modern days. Today 80% energy we use comes from fossil fuels and about estimated that the world's oil reserves will last for 30 to 40 years, whereas solar energy is forever. Burning of fossil fuels introduces many harmful pollutants into the atmosphere and contributes to global warming and acid rain. Solar cell directly converts solar energy into electricity. The solar cells that are connected together make up the solar panel. This can last up-to several decades without replacement. However, there is a drawback of solar power: energy can be produced only in the presence of sunlight. To overcome this, the solar panels are coupled with the rechargeable batteries, which can store excess power generated and provide energy in the absence of sunlight. Solar energy has advantages over other renewable energy sources including wind and water power: solar power is generated using solar panels, which do not require any major mechanical parts, such as wind turbines. These mechanical parts can breakdown and cause maintenance issues and can also be quite noisy. Both of these issues are virtually non-existent with solar panels. This project aims at harvesting solar energy and storing it in a rechargeable battery. Using this battery various low-voltage devices can be charged.

Since many years all over the world people face electricity problem. Various traditional sources used for generating electricity are water, coal, natural gas, hydroelectric wind generator etc. These conventional resources has some disadvantage like; environmental consequences, cost, droughts, and limited reservoirs. Laptop is an essential daily gadget of our tech-savvy life. We need to perform various tasks on laptop every day. Many times when we are out of office or home, we face the problem of laptop battery charging. Other hurdles are frequent power failure problem, non-availability of electrical connection etc. In these situations, we need to find the solution for this laptop charging problem. Proposed system suggests the solution to the problem of power source for the laptop charging. [1]-[7]

A. Motivation

India along with the world is facing Energy Crisis. There is a significant gap in the demand and supply for electricity. Day by day as our country progresses towards development, this gap is increasing and tackling this situation is very important to continue our country ascending path. In order to meet this situation a number of options are being considered with a large focus on renewable energy research amp; development. The options considered are solar energy, because it is Good for the Environment. The most commonly known fact about solar energy is that it represents a clean, green source and free source of energy of energy.

B. OBJECTIVES

- To charge laptop anywhere by using solar energy.
- Only authorized person should get the benefit of this system by using RFID reader.
- More accessible and affordable to the public.

II. LITERATURE REVIEW

Ref no.	Year	Name of Researcher	Name of Paper	Technique Used
1	2012	Soren BaekhojKjaer	Evaluation of the Hill Climbing and the Incremental Conductance Maximum Power Point Trackers for Photovoltaic power System.	Mppt is an important part Of Every PVPS, Employed in order to increase the Yield. A model for the loss was established on the basis of the single Diode model of the PV Cell. The model Predicts 0.1per loss for a perturbation step size of 2per of the MPP voltage, at steady state irradiance and temperature.
2	2014	Girish Makarabbi	Compatibility and Performance Steady of Home Appliances in a DC Home Distribution system	This Paper Presents a study of the Home appliances which can be a operated on the 220v DC supply with no or minor modification, in the devices and also not Deviating from its ac

				performance. This has decreased power loss in the standby mode of the appliances.
3	2015	Dr.D.P.Kadam	Solar Panel Based Multi-mobile charger with LED Illuminatio n	Charge Controller is used to prevent over charging deep discharging. LDR Sensor is used, which senses the light intensity. We get the mobile charging station.
4	2016	S.M. Dehghan	Solar charger system with LED Driver using capacitor-Less Multi-Port Converter.	In This paper a Three Port Converter Was used For Solar Charger application with lighting load. The Used Converter is based on the connection of three conventional two-port converter with non-constant common bus.
5	2016	G. Escoar	Multi-Sampling Maximum Power Point Tracker (MS-MPPT) to compensate irradiance and temperature changes	Experiments showed that the MS-MPPT method was able to provide an excellent tracking of the MPP during irradiance changes following profiles from standard EN50530. Method guaranteed a fast response under rapidly and gradually changing irradiance changes

III. ANALYSIS AND APPROACH

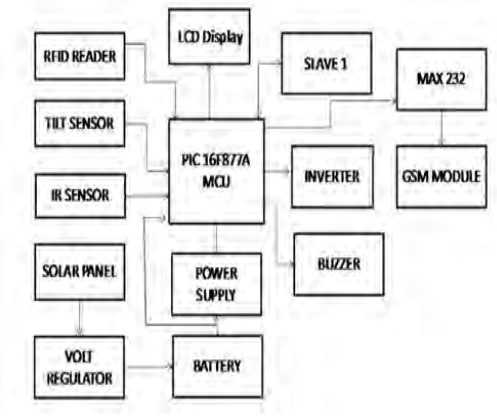


Fig 3.1: Block Diagram of System

A. GSM (Global system mobile):

GSM technology is used for intimating the owner. An alert message is sent to the owner if anyone tries to disturb the system. And information about location of laptop is sent to the owner through message. The module sends SMS to the owner of the laptop using a predefined time interval which is neither too periodic nor too overdue.

SIM900A GSM/GPRS MODULE:

This is an ultra-compact and reliable wireless module. The SIM900A is a complete Dual-band GSM/GPRS solution in a SMT module which can be embedded in the customer applications allowing you to benefit from small dimensions and cost-effective solutions. Featuring an industry standard interface, the SIM900A delivers GSM/GPRS 900/1800MHz performance for voice, SMS, Data, and Fax in a small form factor and with low power consumption. With a tiny configuration of 24mm x 24mm x 3 mm, SIM900A can fit in almost all the space requirements in your applications, especially for slim and compact demand of design. [6-7]

B. RFID Reader

By using RFID Reader only authorized person can take benefit of laptop charging. This is a low frequency RFID reader with serial output with a range of 0.5-1cm. It is a compact units with built in antenna and can be directly connected to the PC using RS232 protocol. RFID is an acronym for radio-frequency identification and refers to a technology whereby digital data encoded in RFID tags or smart labels are captured by a reader via radio waves. RFID is similar to barcoding in that data from a tag or label are captured by a device that stores the data in a database. RFID, however, has several advantages over systems that use barcode as set tracking software. The most notable is that RFID tag data can be read outside the line-of-sight, whereas barcodes must be aligned with an optical scanner.

RFID tag consists of an integrated circuit and an antenna. The tag is also composed of a protective material that holds the pieces together and shields them from various environmental conditions. The protective material depends on the application. For example, employee ID badges containing RFID tags are typically made from durable plastic, and the tag is embedded between the layers of plastic. Two-way radio transmitter-receivers called interrogators or readers send a signal to the tag and read its response. RFID tags can be either passive, active or battery-assisted passive. [13]

Features of RFID Reader

- Serial and TTL output. Along with two RFID cards.
- Excellent read performance without an external circuit.
- Compact size and cost-effective.

C. IR Sensor

IR sensor is use for security purpose which is use to continuous detection of device. The IR transmitter consists of the LED that emits the Infra-Red (IR) radiation. This is received by the photo diode, which acts as IR receiver at the receiving end. Due to the falling of light on the receiver a potential difference is created across the ends. This potential difference is recognized by a microcontroller as HIGH or LOW. Since the IR radiation is invisible to human eye it is perfect for using in wireless communication. An electronic remote device mainly consists of this IR transmitter and receiver. A remote control patterns a flash of invisible light which is turned into an instruction and is received by the receiver module. IR Sensor Pair for transmitter and receiver. An infrared sensor is an electronic device that emits in order to sense some aspects of the surroundings. An IR sensor can measure the heat of an object as well as detects the motion.

This sensor is analogous to human's visionary senses, which can be used to detect obstacles .IR detectors are digital out. IR Sensors are basically electronic devices which are used to sense the changes that occur in their surroundings. The change may be in color, temperature, moisture, sound, heat etc. They sense the change and work accordingly. In IR sensor there is emitter and detector. Emitter emits the IR rays and detector detects it. [12]

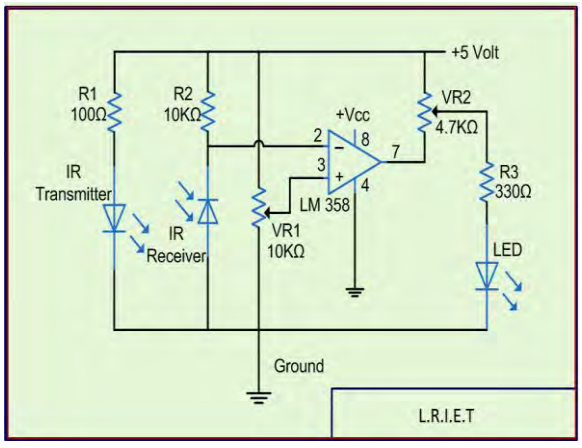


Fig 3.2: Internal circuitry of IR Sensor Block Diagram

Features of IR Sensor

- Resolution: 1%.
- Response Time: 60Sec
- Temperature range: -20C To +50C
- Single-supply operation: 1.8 V to 3.6 V
- Repeatability: 2%
- Output Linearity: Linear
- Relative Humidity: 15 To 90%

D. Solar Panel



Fig 3.3: Solar panel

The advantage is that when solar panel is installed it can produce electricity virtually free. The principle used to generate electricity from the solar panels, sis same as that used to generate electricity from the chemical reaction using a standard battery. The basic working of solar panels depends on the semiconductor property of silicon. The silicon is a unique substance that has revolutionized the way electronic appliances work. This property is used to generate electricity from the solar panels. In order to understand how Solar panels work we need to understand how silicon works at an atomic level. Silicon in its pure form i.e. when all the impurities have been removed a silicon atom is bonded with another silicon atom. Since the valence of silicon atom is 8 which means that there are 8 electrons in its outermost orbit.

However, in its natural state there are only 4 electrons bounded in the outermost orbit. Hence these 4 electrons are able to bond with another 4 electrons with 4 silicon atoms around it. The 4 free electrons that can move around throughout the substance. During the absence of electric potential these free electrons tend to remain close to their parent atoms so that they are at minimum energy level. However when the electric potential is applied across the substance these free electrons receive energy and move direction of the potential difference applied thus producing electric current. However the current produced in a pure semiconductor is quite less because these free electrons wants to remain at minimum energy level possible. [9-10]

E. Microcontroller

In our system we want to convert analog signal into digital form. PIC16f877a has inbuilt ADC, so we choose this microcontroller. The PIC microcontroller PIC16f877a is one of the most renowned microcontrollers in the industry. This controller is very convenient to use, the coding or programming of this controller is also easier. One of the main advantages is that it can be write-erase as many times as possible because it use FLASH memory technology. It has a total number of 40 pins and there are 33 pins for input and output. [11]

PIC 16F877A:

PIC microcontrollers are a family of specialized microcontroller chips produced by Microchip Technology in Chandler, Arizona.

The acronym PIC stands for" peripheral interface controller," although that term is rarely used nowadays.

A microcontroller is a compact microcomputer designed to govern the operation of embedded systems in motor vehicles, robots, office machines, medical devices, mobile radios, vending machines, home appliances, and various other devices. A typical microcontroller includes a processor, memory, and peripherals. The PIC microcontrollers appeal to hobbyists and experimenters, especially in the fields of electronics and robotics. Every PIC microcontroller has a set of registers that also function as RAM (random access memory). Special purpose control registers for on-chip hardware resources are also mapped into the data space. Every PIC has a stack that saves return addresses. [11]

F. Tilt Sensor



Fig. 3.4. TILT sensor

Tilt sensor used for security purpose. It is an instrument which can measure the tilt in multiple axes of a reference plane. Tilt sensors measure the tilting position with reference to gravity, and are used innumerous applications. Similar to mercury switches, it is also be known as tilt switch or rolling ball sensor. Tilt sensors are devices that produce an electrical signal that varies with an angular movement. These sensors are used to measure slope and tilt within a limited range of motion. Sometimes, the tilt sensors are referred to as inclinometers because the sensors just generate a signal but inclinometers generate both readout and a signal. A tilt sensor is an instrument that is used for measuring the tilt in multiple axes of a reference plane. Tilt sensors measure the tilting position with reference Solar Based Laptop Charger. [12]

Features of Tilt Sensor:

Number of Axes: The number of axes are varies from application to application. Dual axes tilt sensor is used in Robotics. In video game controllers and joysticks, a 3 axes tilt sensor is required, and in some smart phones, 4 axes sensors are used.

1) **Resolution:** The minimum inclination detected by the sensor.
2) **Measuring Range:** The range of inclination that can be handled by the sensor is important. Some sensors are capable of measurements up to 10 while others can cover a range of up to 60.
3) **Sensitivity:** The sensor's ability to react to small changes.

G. Power Supply

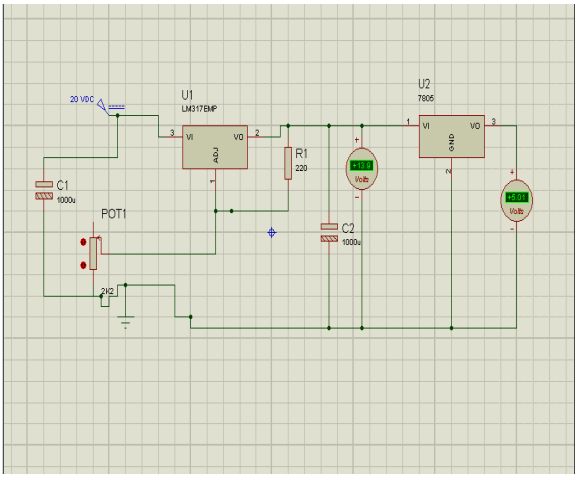


Fig. 3.5. Simulation of power supply

In proposed system 12V supply is required for LCD, relay, RFID reader. It is design using IC LM317. Similarly 5V supply is design using IC 7805, which is required for microcontroller, driver IC, and buzzer. It clearly indicates regulated output voltages of 12.3V and 5.01V, which is sufficient for the proposed system. Solar panel output of 20V is obtained for simulation purpose using LM317 and variable potentiometer of 2.2K.

IV. ADVANTAGES

1) **Emergency Purposes:** Another benefit is that it hardly requires any electrical outlet. It can therefore be used during emergencies and outdoor purposes.
2) **Uninterrupted power supply:** One of the greatest advantages of solar laptop phone charger is that it can be used to charge laptop even during power outages.
3) **Maintenance:** Low maintenance, emission free and environment friendly power source.
4) **Backup power Sources:** Its benefit that it provides power by battery in case of mains failure.
5) **Renewable Energy Source:** Its provides renewable source of energy
6) **Reduces Electricity Bills:** It provides free source of energy without using electricity.

Application

- In rural areas where ac source is not available.
- Mobile phone charger.
- Solar based car system.
- Mini power station.

V. FLOW CHART

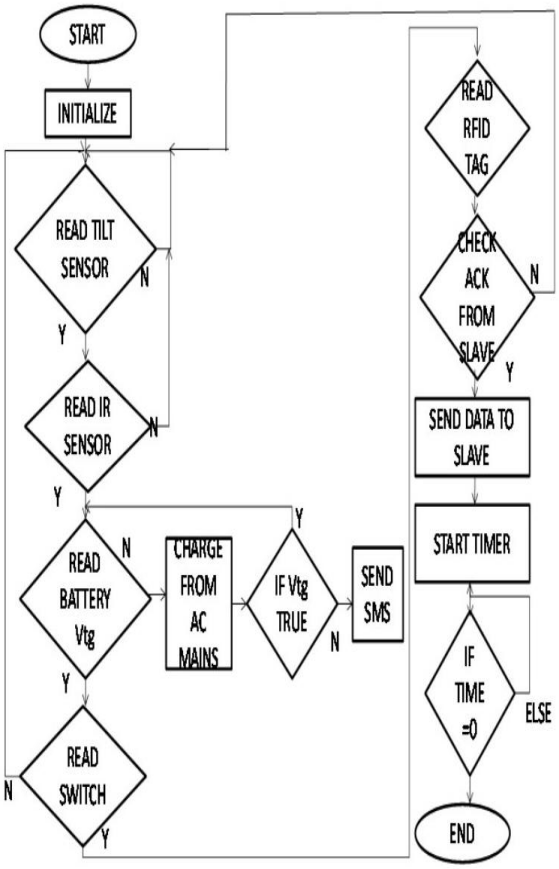


Fig 6.1: Flow chart

VI. CONCLUSION

Renewable energy is not a new concept, nevertheless at an exponential growing population, the development and improvement of them are essential to sustain world power hunger. In 2050 the population expectation on earth is about 9 billion people, where approximately 3 billion will use laptops. The application of renewable energy at portable devices starts to plays a significant role at global energy saving. Solar chargers are simple, portable and ready to use devices which can be used by anyone especially in remote areas.

REFERENCES

[1] Jan Krikke, ”Sunrise For Energy Harvesting Products” Published by the IEEE CS and IEEE , 2005.

[2] RaJeev BansalThe Future of Wireless Charging IEEE Antennas and Propagation Magazine, 49, 1, February 2007

[3] Soren Baekhoj Kjaer,Evaluationof the Hill Climbing and the Incremental Conductance Maximum Power Point Trackers for Photovoltaic Power Systems ,2012 IEEE.

[4] G. Escobar,Multi-Sampling maximum power point tracker MS-MPPT to compensate irradiance and temperature changes, 2016 IEEE

[5] S.M. Dehghan,Solar Charger System With LED Driver Using CapacitorLess Multi- Port Converter, 2016 IEEE.]

[6] Dr. D.P. Kadam,Solar Panel Based Multi-Mobile Charger with LED Illumination 2015 IEEE.

[7] Girish Makarabbi, Compatibility and Performance Study of Home Appliances in a DC Home Distribution System,2014 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES).

[8] K.B.Khanchandani and M.D.Singh , ”Power Electronics”,2006

[9] Chetan Singh Solanki ,”Solar Photovolatics Fundamentals and Technologies”, 2009.

[10] K.C Singal ”Renewable Energy Sources And Emerging”, 25 Nov 2011.

[11] Muhammad Ali Mzidi “PIC Microconrollers and Embedded Systems”,2009.

[12] www.elprocus.com

[13] http://www1.cse.wustl.edu/~jain/cse574-06/index.html

Image Processing For Hazy Weather Conditions

Atanu Roy, Swapnil Kothari, Kaustubh Chakkarwar

Department of Electronics and Telecommunication, MAEER's MIT College of Engineering, Pune, India.

atanu226@gmail.com, swapnilkothari106@gmail.com, kaustubhchakkarwar1234@gmail.com

Abstract: **Accidents due to the hazy vision in adverse atmospheric conditions is not an issue limited to a region. It is observed worldwide that even after developing the road safety knowledge the accidents in hazy atmosphere are not reduced. These are due to some constraints like blurred vision in hazy conditions in which human eyes cannot detect vehicles or other objects coming in front of vehicle. To overcome this issue there should be some interface between the driver and the objects coming in front of vehicle. The issue occurs while driving in hazy environment and can be avoided by providing an interface which will store the video in form of frames at a speed specified by the camera module used on the front panel of the vehicle. The video stored in format of frames will be processed through the Raspberry Pi which is a series of small single board computer. There will be a provision for the driver to turn on the system whenever it is necessary by pressing a button and then there will be a TFT display provided inside the vehicle to visualize the objects coming in front of the vehicle in advance to the driver.**

I. INTRODUCTION

Nowadays low visibility conditions are a frequent occurrence. Due to fog and haze, visibility is impaired. Poor visibility causes flight delay, diversion, cancellation as well as automobile accidents. Driving at night in case of fog forms a challenging condition which may lead to an accident. We have seen a change in the environmental condition. Also pollution has increased the problem in driving and proving to be fatal. This issue is on a larger scale in the industrial region but with the increasing development and industrialization, it won't take long to spread across the countries. Thus, the problem is to deal with such situations in which human performance is limited by low visibility. The resolution of the human eye is 324 to 576 megapixels depending upon the angle of vision, but in the case of fog or other weather haze, the visibility is disturbed. With the advancement in every field of image processing and computer vision from last few decades it is possible to visualize even in bad weather. The performance of many computer vision applications, such as object detection and tracking, video surveillance, FVES and target identification may fail due to degraded images. Therefore improving visibility through fog removal is an inevitable task. Research is still going on to implement such a system so that it will assist the driver in any weather condition.

With the advancement of technology, many single image fog removal methods [7] [8] [9] [10] have been proposed. R.Fattal [7] has proposed a method which solves a non-linear inverse problem and its performance is greatly dependant on quality of input data. If insufficient signal to noise ratio or the absence of multiplicative variation are seen on insignificant portion of the image then there may be a chance of proposed method failure. R. Tan [10] proposed a method in which they first estimate the atmospheric light, from which they obtain the light chromaticity. Based on the estimated air-light, they compute the direct attenuation that represents the scene with enhanced visibility. Since the actual value of air-light is unknown the output tends to have larger saturation values than those in actual clear day images. To solve these problems, we have introduced a method that only requires a single input image that neither requires geometrical information of the input image nor it is dependent on quality of input image.

In this paper a different approach for fog removal technique is proposed and partially implemented. It produces a more enhanced edge preserving image using preprocessing technique like histogram. Proposed algorithm uses color correction at initial stage followed by gamma correction for contrast adjustment and to visualize the fine details within in an image. Contrast enhancement is applied to the gamma corrected image which is preprocessed using histogram equalization. A Wiener filter is used to diminish ringing effect while de-blurring images in presence of noise.

This paper is organized into five sections. In the section II of this paper we explain in detail the proposed fog removal algorithm which consists of color correction, gamma correction, contrast enhancement and histogram equalization. In section III, contrast manipulation with preserving fine details using Wiener filter is discussed. Further in section IV, simulation and results has been discussed. The remaining of this paper gives conclusion and future scope.

II. ALGORITHM

Fig. 1 shows the detailed algorithm of the proposed method. An input image is acquired in the step I of the algorithm. This image is then converted from color space to gray-scale using color conversion in step 2 of algorithm. Uneven luminance is compensated or corrected in step 3 by using gamma correction. Step 4 deals with the contrast enhancement of the gamma corrected image. Histogram equalization is then applied to equalize the brightness and contrast of the image respectively.

A. Input Image:

Input image is the hazy image. In this proposed algorithm we have taken a random hazy image from the internet.

B. Color image to grayscale image conversion:

RGB image contains lots of data which are not required for the processing. When an image is converted into grayscale, lots of information is discarded which are irrelevant to processing. In case of a RGB scale image, for each of the components i.e. R, G, B, image holds different intensity labels. RGB color model, a color image can be represented by the intensity function.

$$I\text{-}_{RGB} = (F_R, F_G, F_B)$$

Where, $F_R(x,y)$ is the intensity of the pixel (x,y) in the red channel, $F_G(x,y)$ is the intensity of the pixel (x,y) in the green channel and $F_B(x,y)$ is the intensity of the pixel (x,y) in the blue channel.

Each of the R, G and B component consist of 8 bits. If only the brightness information is needed, color images can be transformed to gray-scale image. The transformation can be made using proposed equation [11].

$$I_y=0.333F_R+0.5F_G+0.1666F_B$$

Where, F_R, F_G and F_B are the intensity of R, G and B component respectively and I_y is the intensity of equivalent gray level image of RGB image.

C. Gamma correction:

Gamma correction, the second step in this flow is a function which maps the uneven luminance in the image and compensates or syncs that according to the human eye perspective. This is because the video recorders or cameras do not correctly capture the images with proper luminance.

The gamma correction actually plots a graph of voltage on x-axis versus the levels brightness of an image on y-axis. It plots an inverse curve to that of original image which is gamma correction function and a resultant gamma function corrected curve which describes the voltage and brightness level of gamma correction of an image. Fig.2 describes the related theory of gamma correction in an easy way.

$$C(x,y) = A(x,y)^{\alpha}$$

Equation 1: Gamma correction function
Where A(x,y) =Input foggy image with (x,y) pixels
C(x,y) =Output or required image
α =gamma correction factor.

- Solid red curve is of voltage and brightness ratio.
- The dash red curve is its inverse function i.e. gamma correction function.
- The gray dotted line is the corrected result.

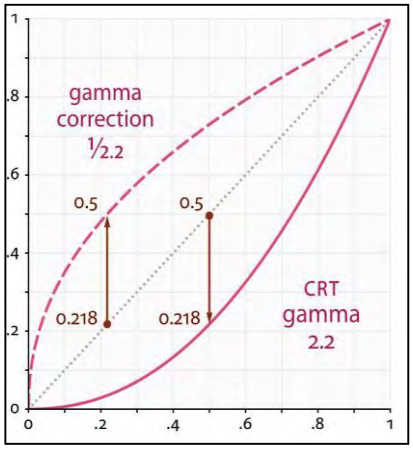


Fig.2 Gamma correction function graph

D. Contrast enhancement:

Contrast enhancement is a process that makes the image features stand out more clearly by making optimal use of the colors between the ranges of gray color levels.

E. Histogram Equalization:

Histogram equalization is one of the techniques for contrast enhancement. In histogram equalization a graph between the values of gray level on x-axis versus the frequency of occurrence of gray colors on y-axis is plotted. After that the graph is analyzed and the frequency of specific values of gray levels is equalized throughout the range of gray-scale.

Implementation:
Consider the discrete grayscale input Image $X = x(i,j)$, with the L discrete levels, where $x(i,j)$ represents the intensity levels of the image at the spatial domain (i,j) Let histogram of Image X is $H(X)$. Now the probability density function $pdf(X)$ can be defined as-

$$Pdf(X_k) \text{ or } p(X_k) = n_k/N$$

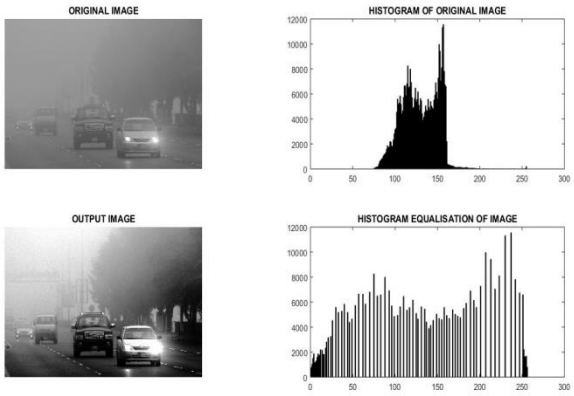
Equation 2: Probability density function (PDF)

Where, $0 \le k \le (L - 1)$
L is the total number of gray level in the image
N is the Total number of pixels in the image,
n_k is the total number of pixels with same intensity level k.

From the *pdf(X)* (1) the cumulative distribution function *cdf(X_i)* is defined as
Cdf(X_i)or c(X_i)= $\sum^{k}_{i=0}p(X_i)$
Equation 3: Cumulative distribution function (CDF)

Note that cdf(X_{L-1})=1 from equation (2) and (3)
Fig.3 (a) input image having low contrast (b) histogram of input image (c) shows histogram equalized image and (d) shows histogram of processed image

Histogram equalization is a scheme that maps the input image into the entire dynamic range [X_o, X_{L-1}] by using



the cumulative distribution function as a transform function. Let's define the transform function f(X) using cdf(X_i) as:-
f(X)=X_o+(X_{L-1}-X_o) x cdf(X_i)
Then the output image of histogram equalization, Y=y(i,j) can be expressed as
Y ={f(x(i,j))| x(i,j) X}

Above shown results in Fig. 3 consist four images. Fig. 3(a) is the original image with haze and as it can be observed objects in that image are not easily detectable. Fig. 3(b) shows the plot of histogram for original image. After histogram equalization the results are shown in Fig. 3(c) with clearer vision of objects and Fig. 3(d) is the equalized plot of resultant histogram equalized image.

F. Haze free image:

This improved algorithm will produce an image which will be haze free. However in the obtained image effects like blurring and ringing are seen, which are not expected and need to be removed with the post-processing techniques.

III. POST PROCESSING

This section deals with further refinement including local contrast enhancing and increasing the fine details over the restored image to clearly observe the object in an image. For this purpose, an alternative edge preserving Wiener filter is used. Wiener filter is specifically used where the ringing effect in an image is observed. The wiener filtering is the linear estimation of the original image. Wiener filter removes the additive noise and reduces the blurring effect simultaneously.

$$W(f_1,f_2) = \frac{H^*(f_1,f_2)Sxx(f_1,f_2)}{|H(f_1,f_2)|^2Sxx(f_1,f_2)+S\eta\eta(f_1,f_2)'}$$

Equation 4: Weiner filter equation.
Where
$Sxx(f_1,f_2)$, $S\eta\eta(f_1,f_2)$ are respectively power spectra of the original image and the additive noise,
And $H(f_1,f_2)$ is the blurring filter

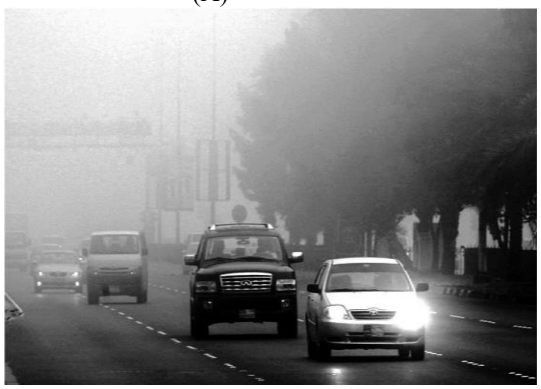
IV. SIMULATION RESULTS

To demonstrate the effectiveness of our method, we have used real images of outdoor scenes in out experiment. The computational time for the 600x400 image using MATLAB software with processor of i3 generation is approximately 15-20 seconds.

Fig. 4(a) and Fig.4(a) shows a road scene partially covered by haze. The distant objects in the images are quite not visible. Fig. 4(b) and Fig. 4(b) shows our result for proposed method. In the result, one can easily observe that the objects which are at far distance are easily visible as compared to original image and even the contrast and brightness is enhanced in our result.



(A)



(B)

Fig. 4 A) Original Image B) Haze free Image



(A)



(B)

Fig. 5 A) Original Image B) Haze free Image



(A)



(B)



(E)



(F)

Fig. 6 shows different outputs of gamma correction function. Image A), B), C) and D) are the outputs of gamma correction function at the values of α = 2, 3, 4



(C)



(D)



(G)



(H)

and 5 respectively. And the images E), F), G) and H) are the final output images.

Fig. 6(A) shows gamma corrected image for gamma value α=2. The corresponding output for the

gamma corrected image can be seen in Fig. 6(E). Fig. 6(B), (C) and (D) shows different gamma corrected images for different gamma values. It can be observed that with increase in value of gamma correction the luminance in the resultant images get properly mapped such that the objects in the image get easily detected.

V. CONCLUSION

In this paper, we presented a framework able to detect and clear the fog. We have successfully removed the fog from randomly selected foggy images. The proposed system can be implemented for various applications like computer vision system for object tracking, night vision surveillance. The proposed method is an efficient single image fog removal method which is used for gray-scale images. As shown in the simulation results the obtained fog free image is clearer than the original image. As the proposed system can be used as a versatile tool in driving assistance, object detection and it can be implemented further on color images or videos.

VI. REFERENCES

[1] N. Hautiere, R. Labayrade, D. Aubert, "Detection of visibility conditions through use of onboard cameras," Intelligent Vehicles Symposium, Proceedings, IEEE, pp.193-198, 6-8 June 2005.

[2] Nicolas Hautiere, Raphaël Labayrade, and Didier Aubert. "Estimation of the Visibility Distance by Stereovision: A Generic Approach", IEICE - Trans. Inf. Syst. E89-D, 7 (July 2006)

[3] Hautiere, N.; Labayrade, R.; Aubert, D., "Real-time disparity contrast combination for onboard estimation of the visibility distance", IEEE Transactions on Intelligent Transportation Systems, vol.7, no.2, pp. 201- 212, June 2006

[4] Pomerleau, D., "Visibility estimation from a moving vehicle using the RALPH vision system", IEEE Conference on Intelligent Transportation System, 1997. ITSC '97, pp. 906-911, 9-12 Nov 1997

[5] N. Hautiere, D. Aubert, "Contrast restoration of foggy images through use of an onboard camera," Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE, pp.601-606, 13-15 Sept. 2005.

[6] Mori, K.; Kato, T.; Takahashi, T.; Ide, I.; Murase, H.; Miyahara, T.; Tamatsu, Y., "Visibility Estimation in Foggy Conditions by In-Vehicle Camera and Radar," Innovative Computing, Information and Control, 2006. ICICIC '06. First International Conference on, vol.2, no., pp.548- 551, Aug. 30 2006-Sept. 1 2006.

[7] S. G. Narasimhan, S. K. Nayar, "Shedding light on the weather. In Proceedings of the 2003 IEEE computer society conference on Computer vision and pattern recognition (CVPR'03). IEEE Computer Society, Washington, DC, USA, pp. 665-672.

[8] R. Fattal, Single image dehazing. In SIGGRAPH, pages 1-9, 2008.

[9] S. G. Narsimhan and S. K. Nayar. Contrast restoration of weather degraded images. PAMI, 25:713-724, 2003.

[10] R. Tan. Visibility in bad weather from a single image. CVPR, 2008.

[11] Tarun Kumar and Karun Verma, A Theory based on Conversion of RGB image to gray image, International journal of computer application(0975-8887), Vol. 7- No. 2, Sept 2010.

[12] Omprakash Patel, Yogendra P. S. Maravi and Sanjeev Sharma, A Comparative Study Of Histogram Equalization based image enhancement techniques for brightness preservation and contrast enhancement, An International Journal (SIPIJ), Vol-4, N0. 5, Oct. 2013

[13] Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing, third edition, Pearson publication.

Ferromagnetic Detection System For MRI Safety

Yugantari Kailas Mahajan, Pooja Pramod Saner, Mansi Sanjay Bihani, Dr. V. M. Birari

Electronics & Telecommunication
NDMVPS'S KBTCOE
NASHIK

ABSTRACT — According to world health organization (WHO) , MRI accidents were increasing at a rapid rate .to overcome this problem Ferromagnetic detection system was introduced. It is specially designed to only detect ferromagnetic objects. Other materials, such as aluminum and copper are nonferromagnetic and, therefore, are not detected by an FMDS. For example, an FMDS will detect a steel gas cylinder and indicate a positive alarm but it will not detect or alarm on an aluminum one. Thus, the FMDS will only alarm on potentially dangerous objects relative to issues related to magnetic field interactions. The ambient field is a combination of the fringe field of the magnet and the Earth's magnetic. An FMDS ignores static magnetic fields (i.e., magnetic fields that do not change with time).it is basically used in MRI where Magnetic resonance imaging (MRI) is a safe and painless test that uses a magnetic field and radio waves to produce detailed pictures of the body's organs and structures. When ferrous objects reach the MRI magnet's fringe field, they can be strongly attracted, at high speed, towards the bore of the magnet. This frightening and dangerous phenomenon is known as the projectile effect. Although MRI safety procedures are in place in all MRI facilities, intended to prevent such incidents, the number of MRI projectile events continues to rise year-on-year. Ferroguard for MRI safety clearly identify ferromagnetic hazards and provide valuable objective data to help MRI staff make accurate safety-decisions to significantly enhance the overall effectiveness of MRI safety processes.it Gives advance visual warning of the presence of dangerous ferromagnetic objects that is view technology.

Key words — FMDS, MRI, Projectile Effect, Ferroguard.

I. INTRODUCTION

To verify the successful screening of persons and identifying ferromagnetic object which were not discovered by conventional screnning method, Ferromagnetic detection system was introduced. Magnetic resonance imaging (MRI) is a safe aend painless test that uses a magnetic field and radio waves to produce detailed pictures of the body's organs and structures. An MRI differs from a CAT scan (also called a CT scan or a computed axialomography scan) because it doesn't use radiation.

An MRI scanner consists of a large doughnut-shaped magnet that often has a tunnel in the center. Some centers have open MRI scanners that have larger openings and are helpful for patients with claustrophobia. MRI scanners are located in hospitals and radiology centers.

MRI is used to detect a variety of conditions, including problems of the brain, spinal cord, skeleton, chest, lungs, abdomen, pelvis, wrists, hands, ankles, and feet. In some cases, it can provide clear images of body parts that can't be seen as well with an X-ray, CAT scan, or ultrasound. MRI is particularly valuable for diagnosing problems with the eyes, ears, heart, and circulatory system. Now to reduce the MRI accident ferromagnetic detection system is specially designed to only detect ferromagnetic objects. Other materials, such as aluminum and copper are nonferromagnetic and, therefore, are not detected by an FMDS.For example, an FMDS will detect a steel gas cylinder and indicate a positive alarm but it will not detect or alarm on an aluminum one.

Thus, the FMDS will only alarm on potentially dangerous objects relative to issues related to magnetic field interactions.The ambient field is a combination of the fringe field of the magnet and the Earth's magnetic. If it is brought close to an FMDS, the distortion is detected as a changing magnetic field and an alert is triggered. Importantly, an FMDS ignores static magnetic fields (i.e., magnetic fields that do not change with time).

II. MOTIVATION

In 2001, a 6-year-old boy named Michael Colombini died in Westchester,New York,after an oxygen canister flew at his skull during an MRI for abenign brain tumor. Similarly, In 2014, a technician at another hospital in Mumbai spent 4 hours wedged inside an MRI machine after he was pinned between ward assistant carrying an oxygen cylinder and the scanner. As the accidents were rising day by day so there was a need to stop it.

The man, Rajesh Maru, was visiting a relative at a hospital in Mumbai and had been handed a metal oxygen cylinder to carry,according to the Agence France-Presse. He etered the MRI room after being told the machine was off, but the powerful magnet that runs the machine was functioning and pulled the oxygen cylinder toward it. Maru may have died from inhaling liquid oxygen from the damaged cylinder, according to Mumbai police. The police also said two hospital staff members had been arrested for causing death by negligence.



In 2014, a technician at another hospital in Mumbai spent

4 hours wedged inside an MRI machine after he was pinned between a ward assistant carrying an oxygen cylinder and the scanner. The technician lost blood circulation below the waist and was temporarily paralyzed; he also suffered organ damage and internal bleeding, according to the Mumbai Mirror.Last year, the maker of the machine, General Electric, paid the technician asettlement of 10 million rupees.The most common MRI injuries, though, are burns, according to a 2008 report by The Joint Commission, a nonprofit healthcare accreditation agency. When metal is left inside a patient's body or a tattoo containing metallic pigments is overlooked the magnetic fields induced by the MRI can create electrical currents in that metal, potentially heating up the soft tissue around it.

III. OBJECTIVE

1. To describe the projectile effect and how magnetic fields interact with metal.

2. To describe ferromagnetic detection systems and methods of detection including advantages.

IV. ANALYSIS & DESIGN APPROACHES

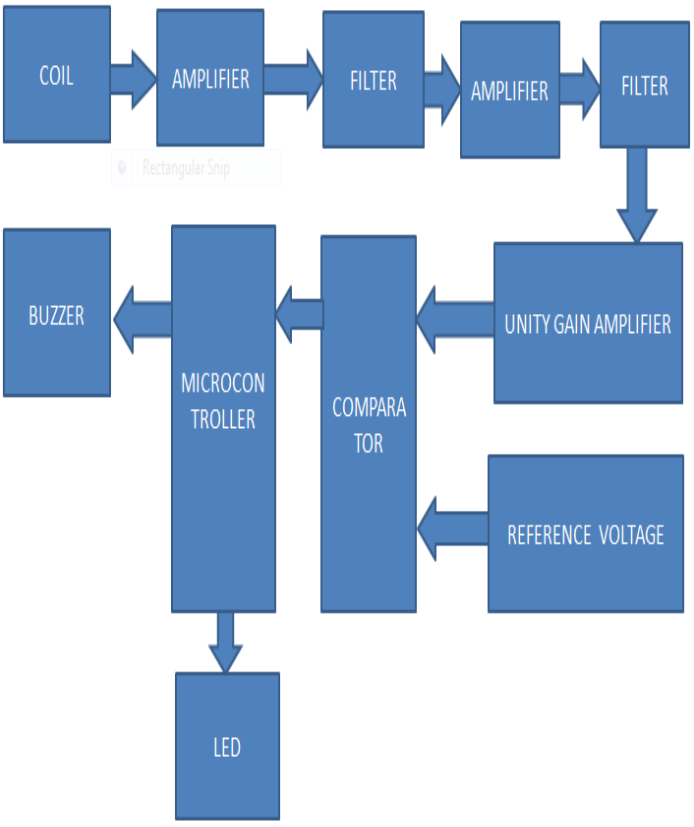


Figure 1: Block Diagram of System

4.1 ELECTROMAGNETIC COIL

When we place a electromagnetic coil in earth's magnetic field and then we bring metal at the same time towards the coil then it will disturb the magnetic field. So due to change in magnetic field current is induced in the coil and in the case of ferrous object field arround the coil wil get disturb, but what Lenz's law will do is it will try to keep field stable and therefore ferrous object get detected and current is induced. Ferrous object got detected because it has high permiability and there is change in the magnitude of signal.

4.2 AMPLIFIER

A transimpedance amplifier, (TIA) is a current to voltage converter, almost exclusively implemented with one or more operational amplifiers. It is also possible to construct a transimpedance amplifier with discrete components using a Field effect transistor for the gain element. After the output of the electromagnetic coil which was in the current form so there was a need to convert it into voltage to work on it.

Now the second amplifier we are using for the purpose ofis an electronic device that can increase the power of a signal (a time-varying voltage or current). It is a two-port electronic circuit that uses electric power from a power supply to increase the amplitude of a signal applied to its input terminals, producing a proportionally greater amplitude signal at its output. The amount of amplification provided by an amplifier is measured by its gain: the ratio of output voltage, current, or power to input. An amplifier is a circuit that has a power gain greater than one.

4.2 FILTER

Active High Pass Filter as its name implies, attenuates low frequencies and passes high frequency signals. It consists simply of a passive filter section followed by a non-inverting operational amplifier. The frequency response of the circuit is the same as that of the passive filter, except that the amplitude of the signal is increased by the gain of the amplifier and for a non-inverting amplifier the value of the pass band voltage gain is given as 1 + R2/R1, the same as for the low pass filter circuit.the filter where it performs signal processing function that is it removes the unwanted signal from the output of amplifier.

4.3 BUFFER

Buffer is also called as unity gain amplifier.It is used for impedance matching which is going to compare or going to match the impedance of amplifier and comparator.

A unity gain buffer (also called a unity-gain amplifier) is a op-amp circuit which has a voltage gain of 1. ... The reason it is called a unity gain buffer (or amplifier) is because it provides a gain of 1, meaning there is no gain; the output voltage signal is the same as the input voltage.

4.4 COMPARATOR(LM339)

Comparator will detect the smaller and bigger object in doing so, it is going to compare the voltage level which has minimum voltage reference of +/- 20 mV and maximum voltage reference is 2.3V. Therefore, the objects will get detected according to voltage level.

This comparator is designed for use in level detection, low-level sensing and memory applications in consumer automotive and industrial electronic applications.

Features

1. Single or Split Supply Operation
2. Low Input Bias Current: 25 nA (Typ)
3. Low Input Ofset Current: +/- 5.0 nA (Typ)
4. Input Common Mode Voltage Range to Gnd
5. Low Output Saturation Voltage: 130 mV
6. TTL and CMOS Compatible
7. ESD Clamps on the Inputs Increase Reliability without Affecting Device Operation

4.5 MICROCONTROLLER

Microcontroller will take input from comparator and according it will perform necessary operation.If the object is ferromagnetic in nature.It will allow the LED and buzzer to get ON, if object is non-ferromagnetic LED will remain green and buzzer will be OFF.

A Microcontroller is a VLSI IC that contains a CPU (Processor) along with some other peripherals like Memory (RAM and ROM), I/O Ports, Timers/Counters, Communication Interface, ADC, etc.8051 is an 8 { bit Microcontroller i.e. the data bus of the 8051 Microcontroller (both internal and external) is 8 { bit wide. It is a CISC based Microcontroller with Harvard Architecture (separate program and data memory).

8051 Microcontroller Features

- 8 Bit ALU:ALU or Arithmetic Logic Unit is the heart of a microcontroller. It performs arithmetic and bitwise operation on binary numbers. The ALU in 8051 is an 8 Bit ALU i.e. it can perform operations on 8 bit data.
- 8 Bit Accumulator:The Accumulator is an important register associated with the ALU. The accumulator in 8051 is an 8 bit register.
- RAM:8051 Microcontroller has 128 Bytes of RAM which includes SFRs and Input / Output Port Registers.
- ROM:8051 has 4 KB of on-chip ROM (Program Memory).
- I/O Ports:8051 has four 8 { bit Input / Output Ports which are bit addressable and bidirectional.
- Timers / Counters:8051 has two 16 { bit Timers / Counters.
- Serial Port:8051 supports full duplex UART Communication.
- External Memory:8051Microcontroller can access two 16 bit address line at once: one each for RAM and ROM. The total external memory that an 8051 Microcontroller can acess for RAM and ROM is 64KB (216 for each type).
- Additional Features:Interrupts, on-chip oscillator, Boolean Processor, Power Down Mode, etc.

4.6 BUZZER

Buzzer is activated to alert the nearby people that the object is ferromagnetic in nature and stop the person from entering inside the MRI room.

4.7 LED

This system has two lights, Red and Green . When object is ferromagnetic in nature, red LED will be glow and when object is non ferromagnetic LED will remain green.

POWER SUPPLY

The system is powered by battery present ,which is generally of 12 Volts. IC7805 is used to provide supply regulated voltage of 5 Volts to MICROCONTROLLER, and other units in the system.

ADDITIONAL COMPONENTS

Along with the major modules and components, a number of other components are also used in our project.
Some of them are listed below:

- 12V Battery
- LEDs, Buzzers, Indicators
- Voltage Regulators & Other Electronic Component.

V . FLOWCHART

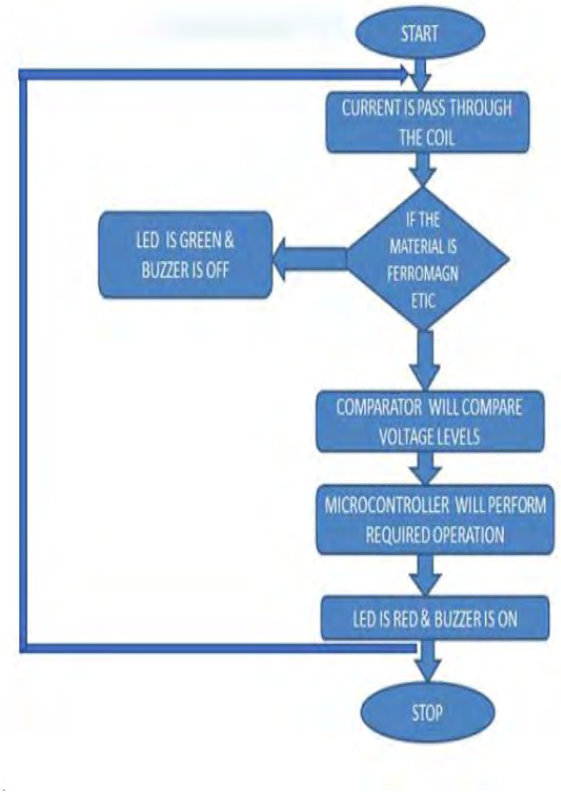


Figure 2: Flow Chart

5.1 ADVANTAGES

1. Provides continuous monitoring of the local ferromagnetic environment.
2. Reduced chance of the \projectile effect".
3. Increased staff awareness of projectile.
4. Visual feedback can be used to test MRI Safe equipment.
5. Reduced chance of equipment damage.
6. Reduce chance of the need to repeat a scan.
7. Reduced costs to hospital facility.
8. Lifetime cost effective with payback after.

5.2 LIMITATIONS

1. FMDS will not detect smaller objects like hair pin, metallic paper pins.
2. It will detect 25mm object from a distance of only 300mm, if it is more than 300mm it will not detect the object.To detect it we will have to increase the size of object.

5. 3 APPLICATIONS

1. The Patient Screening Process is used in an FMDS which is an additional step in the screening process and is the last step before the MRI examination.
2. It is used for detection performance where due to the range when using a patient screening FMDS, magnetically weaker objects can be more reliably detected. In general, bobby pins, hair barrettes, andsome jewelry can be reliable detected with the best performing patient screening FMDS.

VII. CONCLUSION

In this project, we have described the specification and architecture of a Ferromagnetic Detection system for MRI safety. Hence we have learnt how Projectile effect can be eliminated because of which magnetic field interact with metal and responsible for MRI related accidents.It ignores static magnetic field therefore ignoring static object.
Ferroguard for MRI safety clearly identify ferromagnetic hazards and provide valuable objective data to help MRI staff make accurate safety-decisions to significantly
enhance the overall efectiveness of your MRI safety processes.

. ACKNOWLEDGMENT

I would like to express my very great appreciation to Dr.V.M.Birari Sir for his valuable and constructive suggestions during the planning and development of this research work. His willingness to give his time so generously has been very much appreciated.

I would also like to thanks the Sivananda Electronics for enabling me to visit their offices to observe their daily operation.

REFERENCES

[1] Hundley WG, Bluemke DA, Finn JP et al. ACCF/ACR/AHA/NASCI/SCMR 2010 expert consensus document on cardiovascular magnetic resonance: a report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents. Circulation 2010;121:2462-508.

[2] Epstein AE, DiMarco JP, Ellenbogen KA, et al. ACC/AHA/HRS 2008 Guidelines for Device-Based

Therapy of Cardiac Rhythm Abnormalities: A report of the American College of Cardiology/American Heart

Association Task Force on Practice Guidelines (Writing Committee to Revise the ACC/AHA/NASPE 2002 Guideline Update for Implantation of Cardiac Pacemakers and Antiarrhythmia Devices). Developed in collaboration with the American Association for Thoracic Surgery and Society of Thoracic Surgeons. Circulation 2008;117:e350-408.

[3] Kalin R, Stanton MS. Current clinical issues for MRI scanning of pacemaker and de_brillator patients. Pacing Clin Electrophysiol 2005;28:326-8.

[4] Kanal E, Borgstede JP, Barkovich AJ et al. American College of Radiology white paper on MR safety: 2004

update and revisions. AJR Am J Roentgeno

2004;182:1111-4.

[5] Avery JK. Loss prevention case of the month. Not my responsibility! J Tenn Med Assoc 1988;81:523.

[6] Coman JA, Martin ET, Ramza BM, Margolis PD, Fair RD. Pacemaker safety during magnetic resonance imaging at 1.5 Tesla. J Am Coll Cardiol 2001;37(Suppl. A):436.

[7] Irnich W, Irnich B, Bartsch C, et al. Do we need pacmakers resistant to magnetic resonance imaging? Europace 2005;7:353-65.

[8] Pavlicek W, Geisinger M, Castle L, et al. The e_ects of nuclear magnetic resonance on patients with cardiac pacemakers. Radiology 1983;147:149-53.

[9] Fetter J, Aram G, Holmes DR, Gray JE, Hayes DL. The efects of nuclear magnetic resonance imagers on external and implantable pulse generators. Pacing Clin Electrophysiol 1984;7:720-7.

[10] Erlebacher JA, Cahill PT, Pannizzo F, Knowles RJ. Effect of magnetic resonance imaging on DDD pacemakers. Am J Cardiol 1986;57:437-40.

[11] Holmes DR, Jr., Hayes DL, Gray JE, Merideth J. The effects of magnetic resonance imaging on implantable pulse generators. Pacing Clin Electrophysiol 1986;9:360-70.

[12] Iberer F, Justich E, Stenzl W, Tscheliessnig KH, Kapeller J. Nuclear magnetic resonance imaging of a patient with implanted transvenous pacemaker. Herz 1987;7:196-9.

[13] Hayes DL, Holmes DR, Gray JE. E_ect of 1.5 Tesla nuclear magnetic resonance imaging scanner on implanted permanent pacemakers. J Am Coll Cardiol 1987;10:782-6.

Improved Machine Learning Algorithm for Predicting Movie Success

Jaineel Parmar , Pranit Patil , Ritwik Singh, Ms.Meghrani Patil
Computer Department, Thakur College of Engineering & Technology, Kandivali (East), Mumbai-400101, India
parmarjaineel@gmail.com, patilpranit@rediffmail.com, ritwiksingh2108@gmail.com megharani.patil@thakureducation.org

ABSTRACT

Highlight films are a multi-billion industry. Here forecast of a film's prosperity is anticipated dependent on its highlights like cast, class of motion picture, month of discharge, run time, chiefs, makers and so forth. In light of numerous such highlights and with the database of past films measurements, machine learning calculation like Straight Relapse can anticipate the rough appraisals the motion picture can get once it is really discharged and consequently characterize a motion picture as a hit or a flounder. A lot of information speaking to highlight films is kept up by the Web Motion picture Database (IMDb). A considerable lot of the motion pictures recorded on IMDb contain a normal client rating on a size of 0 to 10 which relates to popular feeling of that film. Because of the expansive number of movies created and additionally the level of investigation to which they are uncovered, it might be conceivable to anticipate the achievement of an unreleased film dependent on freely accessible information This information can be separated and arranged for use in preparing machine learning calculations .The objective of this paper is to talk about a framework that can nearly foresee normal client rating by gaining from verifiable motion picture information and henceforth decide whether it is probably going to be a flounder or a hit.

Keywords

Web Motion picture Database (IMDb), Bayesian Belief Network, Straight Relapse.

1. INTRODUCTION

Given the low achievement rate of motion pictures, models and systems can be utilized to foresee the accomplishment of a motion picture. It will help the business essentially. Different partners, for example, performers, makers, chiefs and so on can utilize these expectations to settle on more educated choices. They can settle on the choice before the motion picture discharge. Verifiable information of every segment, for example, on-screen character, performing artist, and executive, author that impacts the achievement or disappointment of a motion picture is offered because of its weightage.. This proposed work expects to build up a model dependent on the information mining procedures that may help in foreseeing the accomplishment of a film ahead of time accordingly lessening certain level of vulnerability. The framework is utilized to anticipate the past and also the fate of motion picture with the end goal of business assurance or basically a hypothetical condition in which basic leadership (the accomplishment of the film) is without hazard, in light of the fact that the chief (motion picture creators and partners) has all the data about the correct result of the choice, before he or she settles on the choice (arrival of the motion picture). With more than two million onlookers per day and movies sent out to more than 100 nations, the effect of Bollywood film industry is impressive. Specifically, we focus on credits significant to the achievement expectation of films, for example, regardless of whether a specific performing artists or on-screen characters are probably going to assist a motion picture with succeeding. The proposed framework provides details regarding the methods utilized, giving their usage and handiness. The imperative issue associated with the expectation framework is, IMDb is hard to perform information mining upon, because of the arrangement of the source information. We likewise found that, the financial plan of a film is no sign of how all around appraised it will be, there is a descending pattern in the nature of movies after some time. Another critical variable are the chief and performers/on-screen characters associated with a film.

2. LITERATURE SURVEY

SR. NO	TITLE	YEAR AND AUTHOR	SUMMARY	GAP IDENTIFIED
1	Raja Harishchandra	Dhundhiraj Govind (Dadasaheb) Phalke (1913 to 1981)	Anatomy of Movie achievement indicator with more than two million Audience daily and movies sent out to more than 100 nations ,the effect of Bollywood film industry is considerable.	Only the movies from Hollywood are included in the database thus leaving other film industries without any way to predict movie success.
2	Utilization of Calculations like RF and SVM for foreseeing acievement.	Meyer (2003)	Assessing RF and SVM among different calculations over various 33 datasets individually.	RF seldom gives better results than SVM. SVM mostly gives better results than RF, but not always.
3		Statnikov (2008)	Assessing RF and SVM among different calculations over various 22 datasets individually.	RF seldom gives better results than SVM.

3.PROPOSED WORK

We will likely foresee the appraisals of new motion picture dependent on the evaluations of the preparation set models. We have thought about 50000 preparing precedents, every one of which has highlights and the relating human appraisals of some specific quality. We at that point shape an expectation demonstrate that can be utilized to foresee the human rating when the computational highlights are known. The objective is to characterize a connection between the forecast esteem and the highlights by settling for the straight coefficients, θ that best guide the highlights to the expectation esteem. Where, the evaluations have been gathered in a vector Y. Y is a (m x 1) vector. For our situation m=50000.

The whole framework predicts the future evaluations dependent on the preparation set of 50000 motion picture subtle elements. To bring the highlights and related records of the highlights we will utilize affiliation control mining. This motion picture set will be pruned to choose an arrangement of highlights that have been found to have a noteworthy effect on the achievement or disappointment of a film. After the distinguishing proof we discover includes every one of the makers, chiefs, on-screen characters and on-screen characters were appraised dependent on their past execution at the Box Office. Correspondingly months of the year and runtimes office motion picture will be allotted a score dependent on similar criteria.

4. SYSTEM ARCHITECTURE

Algorithm

1. Start
2. Initialize associated array F
3. U, ~∑ ,Vt ← svd(M,K)
4. ~M ←U × ~∑ × Vt
5. For f Є S do
6. Wf ←getTermVectorFromTermDocMatrix(f,~M)
7. Initialize similarity list sim
8. i←1
9. foreachcolumn vector w of ~M do
10. sim[i]←wf . w
11. i←i+1
12. end
13. sort(sim)
14. relatedFeaturelist←getTopRelatedFeatures(sim,n,~M)
15. F[f]←relatedfeaturesList
16. end
17. return F
18. end

System Architecture:

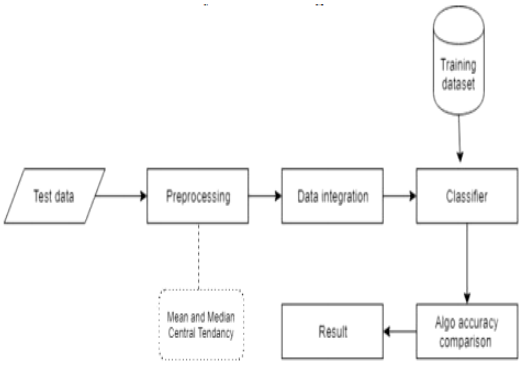


Fig 1. System Architecture

Flowchart:

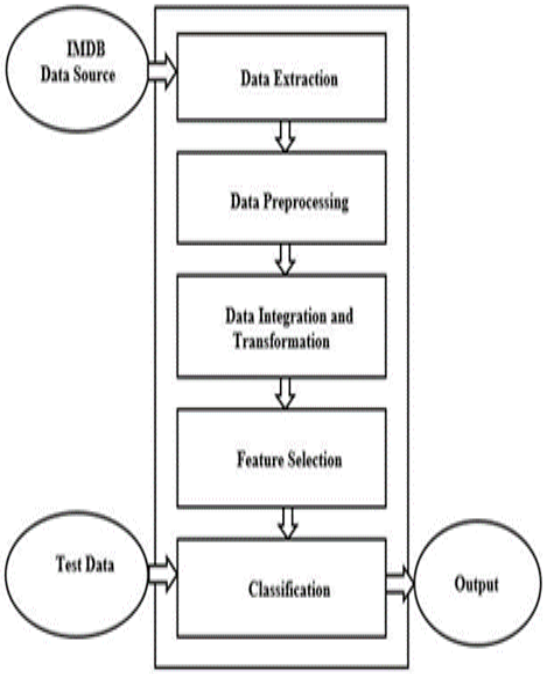


Fig 2. Flowchart

Database Design:

Value	Type	From	To
Title	Title		
Run time	Int	68	271
Year	Int	1918	2017
Month	Int	1	12
Day	Int	1	31

Weekday	Int	1	7
Genre	Genres	1	324
Rating Value	Float	1,6	9,6
Rating Count	Int	9999	1793960
Meta Score	Int	1	100
Oscar	Int	0	11
Budget	Int	0	384
Gross Revenue	Int	1	1301
Director	Name	1	1445
Director Oscars	Int	0	6
Director Rating	Float	5,0	8,0
Writer	Name		
Writer Oscars	Int	0	6
Writer Ratings	Float	4,0	8,0
First Actor	Name		
First Actor Oscars	Int	0	4
First Actor Rating	Float	5,0	8,0
First Actress	Name		
First Actress Oscars	Int	0	4
First Actress Ratings	Float	5,0	8,0

Fig 3: Database Design

5. FEASIBILITY STUDY

The specific first stage in any framework creating life cycle is fundamental examination. The attainability contemplate is a noteworthy piece of this stage. A proportion of how gainful or useful the advancement of any data framework would be to the association is the achievability think about. The achievability of the improvement programming can be examined as far as the accompanying viewpoints:

Operational feasibility: The site will diminish the time devoured to keep up manual records and isn't tedious and lumbering to keep up the records. Henceforth operational possibility is guaranteed.

Specialized feasibility:
At slightest 166 MHz Pentium Processor or Intel good processor.
At slightest 16 MB RAM.
14.4 kbps or higher modem.
A video illustrations card.
A mouse or other pointing gadget.
At slightest 3 MB free hard plate space.
Microsoft Internet Explorer 4.0 or higher.

Prudent feasibility:
Once the equipment and programming necessities get satisfied, there is no requirement for the client of our framework to spend for any extra overhead. For the client, the site will be financially attainable in the accompanying viewpoints:
The site will diminish a great deal of printed material. Consequently, the expense will be lessened.
Our site will decrease the time that is squandered in manual procedures. The stockpiling and dealing with issues of the registers will be understood.

Legitimate feasibility:
The authorized duplicate of the required programming is very shabby and simple to get. So from lawful perspective the proposed framework is legitimately attainable.

6.RESULTS AND DISCUSSION

The customer ought to have the capacity to associate with the server appropriately with no issues. The association foundation between the cell phone and the server should take insignificant time. The cell phone ought to be capable get information from the server uninterruptedly. Information given by the application ought to be right and according to the user's require.

7.CONCLUSION

This undertaking had as a fundamental objective to build up a model capable of anticipating the movies money related achievement of a specific arrangement of films through particular factors and recorded information. It was conceivable to presume that the level of achievement of the cinematographic income expectation is very unique dependent on the typology of the needy variable utilized in the investigation. The exact model showed great measurable outcomes when the reliant variable was twofold and interim. Nonetheless, in what respects the multiclass expectation, the outcomes were exceptionally a long way from the real world, contrarily affecting the model. In Table 9, the last order blunders for every single one of the models and their separate prescient instruments are spoken to. It ought to likewise be noticed that the high misclassification mistake of the multiclass variable displays a level of accomplishment of around 40%, under 12.6% contrasting with the rate gotten by Sharda e Delen

in 2009. The twofold factor shows a very ideal misclassification mistake since one of the investigations with the greatest level of achievement turning to neural systems and a similar double typology (Rhee and Zulkernine, 2016), acquired a positive execution of 88.8%, around 4.5 rate focuses beneath the aftereffect of this examination. Then again, the normal squared mistake came to in the persistent model got exceptionally ideal outcomes and survived, fundamentally, the one acquired by a comparable report that utilized a similar typology (Hunter et al., 2016), and that was not ready to achieve blunder esteems second rate compared to 0.427. As to utilized techniques, neural systems was the one that exhibited a superior prescient power, with an unmistakable favorable position in the majority of the three created models. This is an outcome in accordance with the writing audit made in this examination, as neural systems as often as possible exhibited the most elevated rates of accomplishment (Ghiassi et al., 2015; Kaur and Nidhi, 2013; Rhee and Zulkernine, 2016). In what concerns the impact of the factors present in the examination it was conceivable to infer that some of them are not applicable to a greater film industry income, being hence, evacuated. The variable "performing artist" in the specific instance of multiclass and interim benefit did not present an incredible clarification esteem relatively to the staying ones, a circumstance that happened at the same time with the parallel ward variable in regards to the variable "Oscars". Then again, there is a gathering of factors that clarify the model great and that contribute unmistakably to its prescient achievement: "spending plan", "executive" and "continuation".

8.FUTURE SCOPE

It can be additionally reached out for creating surveys identified with the items in Online shops.

• It can be additionally reached out for creating surveys identified with the items in Online shops.

• It can be utilized for creating surveys for the online recordings.

• It can be utilized for creating surveys identified with the universities amid affirmation process.

• It can be utilized for creating surveys of the competitors in the race.

ACKNOWLEDGEMENT

We sincerely thank our guide Ms. Megharani Patil for her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

We thank the HOD, Dr. Sheetal Rathi, Dean Academic, Dr. R.R Sedamkar, the Principal, Dr. B. K. Mishra and the college management for their support.

REFERENCES

[1] W. Zhang and S. Skiena, “Improving movie gross prediction through news analysis”, IEEE/WIC/ACM International Conference on Web In-telligence and Intelligent Agent Technology, Milan, 2009

[2] Sitaram Asur and Bernardo A Huberman. “Predicting the future with social media”. In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on, volume 1, pages 492–499. IEEE, 2010.

[3] Karl Persson. “Predicting movie ratings : A comparative study on random forests and support vector machines”, 2015.

[4] Business Intelligence from Social Media: “A Study from the VAST Box Office Challenge, Computer Graphics and Applications”, IEEE Oct 2014 (Volume:34 ,Issue: 5),10.1109/MCG.2014.61

[5] “Predicting iPhone Sales from iPhone Tweets”,2014 IEEE 18th International Enterprise Distributed Object Computing Conference, Pages 8190, 10.1109/EDOC.2014.20

[6] “Content Based Prediction of Movie Style, Aesthetics, and Affect: Data Set and Baseline Experiments”, IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 16, NO. 8, DECEMBER 2014

[7] Improving motion vector prediction using linear regression, “Proceedings of the 5th International Symposium on Communications”, 10.1109/ISCCSP.2012.6217750, Control and Signal Processing IEEE, ISCCSP 2012, Rome, Italy, 24 May 2012

[8] The International Movie Database (IMDb). https://www.imdb.com/

[9] The Movie Database (TMDb). http://www.themoviedb.org/?language=en-US

[10] http://www.pwc.in/en_IN/in/assets/pdfs/PwC- Indian-Entertainment-and-Media-Outlook- 2009.

[11] Utilization of Calculations like RF and SVM for foreseeing accomplishment, Verikas (2015), Inside late investigation have studied various expansive and also little scale examinations on information mining and machine adapting, all of which incorporate the RF calculation, particularly issuing its forecast execution in contrast with different calculations and the utilization of the variable significance gauges accessible from RF.

[12] Utilization of Calculations like RF and SVM for foreseeing accomplishment, Meyer (2013), Assessing RF and SVM among different calculations over various 33 datasets individually.

[13] RF seldom gives better results than SVM. SVM mostly gives better results than RF, but not always, Statnikov (2014)

[14] Predicting IMDb Movie Ratings Using Google Trends,

Deniz Demir and Olga Kapralov And Hongze Lai(2012)

Clump Approach for Animal Concern System

[1]Aruna Pavate, [2]Pranav Nerurkar,[3] Mansi Shah, [4] Rajesh Bansode, [5] Anita Chaudhary
[1,4,5] Research Scholar,IT Thakur college of Engineering, Mumbai, India,
[2] Research Scholar,IT Veermata Jijabai Technological Institute, Mumbai, India,
[3] CMPN Dept, Rizvi college of Engineering Mumbai, India
[1] arunapavate@ atharvacoe.ac.in, [2]pranavn91@gmail.com, [3]mansi.shah@eng.rizvi.edu.in, [4]rajesh.bansode1977@gmail.com,
[5]anitac@sjcet.co.in

***Abstract*—This Machine learning approach have proved best results in many applications from healthcare, Videos Surveillance, Social Media Services, Email Spam and Malware Filtering, Search Engine Result Refining, product recommendations, online fraud detection etc. Machine learning can provide an important part of the solution for cutting the cost and improving the quality with fewer efforts. This work implements machine learning technique to create a network of users who can assist an organization and or an individual to spread awareness by helping them to reach proper animal care center. Here adopted unsupervised learning for clustering animal care centers. The developed model is claimed to make predictions with relatively high accuracy. Furthermore, the result of the clustering-guided approach of prediction is discussed in general.**

Keywords—Machine learning, unsupervised learning, k-means algorithm, animal care system, clustering

I.INTRODUCTION

This Clustering is a type of unsupervised technique which helps to group the data samples based on similarities of the data. The result of clustering gives the groups of samples clumped together based on the similarities with labels [1]. Clustering methods developed on the data without labels i.e. unknown data class. Therefore, clustering is often classified as an unsupervised learning method [2]. Now days, good and legal vet hospitals are difficult to find. The designed system helps the individuals to find the legal vet hospitals/animal centers. Animal care system allows individuals to create a network of users who can assist an organization /an individual, for spreading the awareness by helping the strays reach a proper care center, which is difficult in current scenario. Direct contact to the vets is not possible, a user or an organization can connect too many different animal lovers through this system. It's not only changes the environment but does also helps the various systems to enact upon it and have a complete healthcare system for animal in one roof. It is necessary to get information of the all useful vet hospitals so animals can be admitted. Animal care system helps care center to manage the information about types of animals wanted to adopt by individuals. Also, a system informs the availability of animals in the specific center and allows the persons to get through the center. The randomness of the Animal Center managed and controlled, to put a limit on the number of iterations to be carried out in the conventional algorithm.

A. Significance of the study

There are very less Animal care centers available in India. Many pet animals lost. It is necessary to ensure the return of missing animals and also identification of animals in the form of licenses. It is difficult to search missing pets, if the full time service centers available which helps in searching pet. Also an animal care center helps the officers to assist animals in the field. An individual helps in the placement of shelter of animals and also give fundraising support and valuable administrative. Animal care centers helps animals to rescue in need, create a threat to public safety, injured animals ,stray dogs that need extra help and also can assist to answer to phone calls from the public have queries regarding animals. It is requisite to deal with the issues such animal testing, Animals in zoos and related facilities, abandoned pets, hunting, poaching, Cruelty to animals etc.

Cluster Analysis is one of the most important data mining techniques which help the researchers to analyze the data and categorize the attribute into various groups. [3] K Means is one the frequent partitioning algorithm used in clustering. The improvement in the K- means algorithm clustering can be done by choosing appropriate initial cluster centers to converge quickly to the local optimum. [2] The paper suggests that initialization of the cluster centers cannot be separated from effectiveness and the concept of success and failure. The randomness of the cluster centers needs to be managed and controlled to put a limit on the number of iterations to be carried out in the conventional algorithm with decreased complexity and increased accuracy.

B. Applications

- Individuals can operate this application from any place to adopt a pet by contacting vet care centre.
- When an individual comes across an injured stray, he/she can use this application to call an ambulance or take them to a nearby vet hospital/animal centre.
- Animal care center as well as NGOs can spread the awareness about the strays with the help of this application.
- In case of misplaced pet an individual post the details of the pet on the portal for assistance as well as can contact to the doctor in illness or in case of injury.

The study initially started with the aim of grouping vet care centers from a data-driven standpoint [5]. In this concern, unsupervised clustering using the K-Means algorithm was used. The learned clusters were then utilized to train the forecast models as presented at the following sections in detail.

II.LITERATURE REVIEW

There are many applications where K-means algorithm applied for clustering data. A few of them are mentioned below.
Mahmoud Elbattah, Owen Molloy implemented unsupervised and supervised approach for clustering patients individually in Ireland. This study describes the machine learning problem in healthcare. K means algorithm suggested three coherent clusters of patients. The study adopted random forest approach for classification and regression. Author claimed that approach applied provide more accurate predictions compared to non-clustering-based predictions. [6]
Parasian D.P Silitonga, used K-means algorithm to cluster the patients disease at Haji Adam Malik Hospital in Medan. K means clustering algorithm applied using Weka tool. Results showed that the pattern of disease tendency of patients like Mild Septicemia Disease with the highest tendency pattern. The results showed that in future hospitals can predict the priority of service based on the example of trends received. [7]
Pranav Nerurkar, Archana Shirke, Madhav Chandane, Sunil Bhirud , have studied different clustering approaches in the context of massive data set with their strength and weaknesses. Results showed that current clustering algorithms needed parameter tuning as well as sensitive to noise. [8]

Shenghui Wang, Rob Koopman, used unsupervised learning such as K-Means and the Louvain community detection algorithm to cluster the document which helps in topic identification. This work concentrated on semantic representation of an article from the entities associated with it. [9]
Marco Capo, Aritz Pérez, Jose Lozano proposed recursive and parallel algorithm to cluster massive data. Author showed that
K means algorithm works well on both the problems that are dimensionality and number of instances without affecting the quality of approximation & presented a new Boundary weighted K-means algorithm.[11]

There are variety of clustering algorithms used among that, the k means algorithm remains one of the most popular algorithms[10][12][13][14][15].

Oyelade, O. J , Oladipupo, O. O , Obagbuwa, I. C designed system for analysing students results using clustering technique. In this work, k means algorithm combined with deterministic model to analyses the student result. Author claimed that this model is improved over the previous model such as Fuzzy logic and reduced the limitations. Clustering algorithms serves as best algorithms and enhances the decision-making power. [16]

Author's discussed the shortcomings of K-means algorithm such as efficiency of the clustering algorithm, calculation of distance between the object. In this work they proposed an improved k means algorithm to eliminate the shortcomings. Modified algorithm improved speed of clustering and reduced the complexity of the standard K means algorithm. [17]

Huda Hamdan Ali , Lubna Emad Kadhum, discussed K means algorithms used in various applications from economic science to Image processing. This work focused on the image processing and data mining applications such as efficient customer segmentation, Pattern Recognition, to detect an infected part of fruit, Leukemia Image Segmentation and discussed the advantages and disadvantages of the algorithm. [18]

Author proposed new algorithm based on standard K-means algorithm for face extraction. Improved algorithm helps reduction in intra cluster distance, number of iterations, processing times, and an increase in the accuracy rate. [19]

Unnati R. Raval, Chaita Jani, suggested new algorithm based on the traditional K means algorithm. Author claimed that the time complexity of the proposed algorithm is much faster than traditional K-means. There are many advantages of the K means algorithm ad compared to disadvantages of, but it still needs some improvements. Also showed that the value of K affects the clustering decision.[20]

Author addressed the problem of the traditional k means algorithm. In this work the classical k means algorithm modified by keeping the advantages of the original and concentrated on improving centroids. The updated algorithm computes new centroids by finding the mean of all data points within the same cluster. The updated and classical K means algorithm applied on ten different datasets and comparative analysis done based on the measurement metrics recall, F1-score, accuracy, precision. Also it is proved that classical k –means failed to support outliers. [21]

Wang proposed updated k-means algorithm which is feasible and effective for grid simulation analysis. The working of the proposed algorithm showed on the simulator. The results showed that there is difference between the theoretical results and the actual results. The proposed generative model detects the anomalous data sequences. [22]

This work clusters the text with high similarity. The initial cluster selected randomly and the two samples chosen based on the largest distance from the initial cluster. The work showed that the selection of the initial cluster center affects the time and precision of the clustering. Author proposed the vector space model to cluster the similar text. The result showed that the total time utilized for clustering has been reduced as well as the accuracy of the system has been improved. [23]

Dhyaa Shaheed, worked on clustering of Arabic words by using k- means algorithm. Clustering of words done based on the characteristic like syntactic, morphological, and Semantic. The performance of the system evaluated based on parametric measures likes Precision, F-Measure, and recall. [24]

III.APPROACH OVERVIEW

In Machine learning, learning can be possible with three ways: 1. Supervised Learning, 2. Unsupervised Learning and 3. Reinforcement Learning, Algorithms related to Unsupervised Learning have no attribute to predict fixed to the data. In place of output the data has an input which would be multiple attribute that describe the data.
The study implemented an approach that composed of an unsupervised learning. Initially, the study aimed to explore the potential presence of coherent clusters of Vet care centers. The K-Means algorithm was used to discover the groups of vet care center nearby location. The prediction model developed to clump the vet care centers and helps the centers as well as the individuals to access the information about animals.

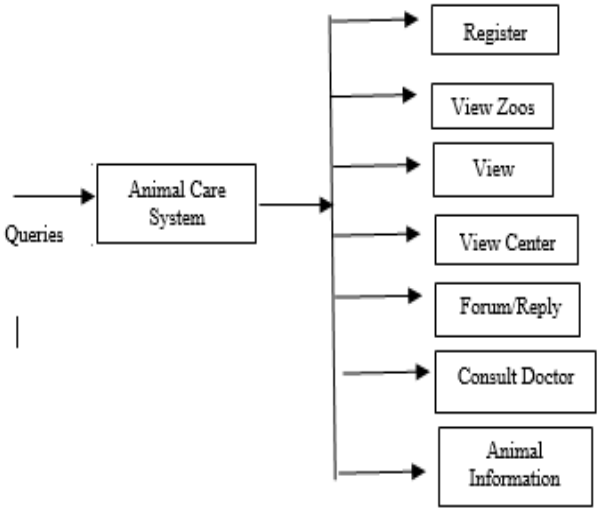


Fig. 1. Approach Overview

The figure 1 describes various modules such as Register, view zoos, etc. which can be accessed by the User through the proposed Complete Animal Care system. The User can Register and then Login to access the various modules.
K-Means is based on the concept of centroid-based clustering. A centroid is an any data point at the center of a cluster. In K-means clustering, clusters are represented by a centroid. It is not necessary that centroid contains in the dataset.

A. Data Description

Table 1 lists the attributes that were initially considered as probable features for the clustering and prediction models as well. Dataset consists of the database records of Veterinary hospitals, adoption centers, forum details record and doctor forum.

TABLE I. ATTRIBUTES EXPLORED AS POSSIBLE FEATURES.

Attributes Explored		
NGO Id	Address	Forum Id
Name	Operational Since	Creation Date
Registration id	Logo	Title
Phone number	Website	Description
Email id	Area of Interest	Gender
Google Map Link	Doctor Id	Solution
Concern Id	Query	User Id

B. *Methodology*

For the development of k-mean clustering algorithm the following steps were followed:

- **Step 1**. Loading the dataset

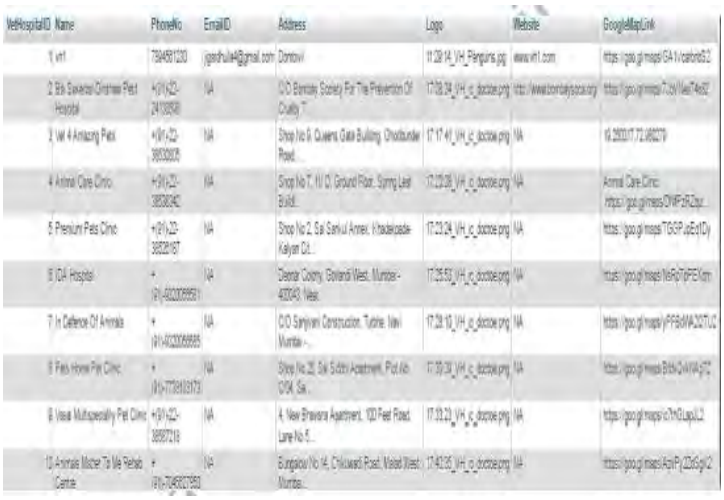


Fig. 2. Approach overview : Dataset Attributes 1

- **Step 2.** Cleaning the dataset to remove all the null values..
- **Step 3.** First 300 records taken from the data set for analysis.



Fig. 3. Approach overview : Dataset Attributes 2

- **Step 4 :** Attributes NGO name and location (googleMaplink) used to make clusters.
- **Step 5** Determined the number of clusters
- **Step 6.** Analysed cluster formation
- **Step 7.** Examined the coordinates of the cluster centroids
- **Step 8.** Examined how the cluster assignment relates to individual characteristics.

C. Result and Analysis

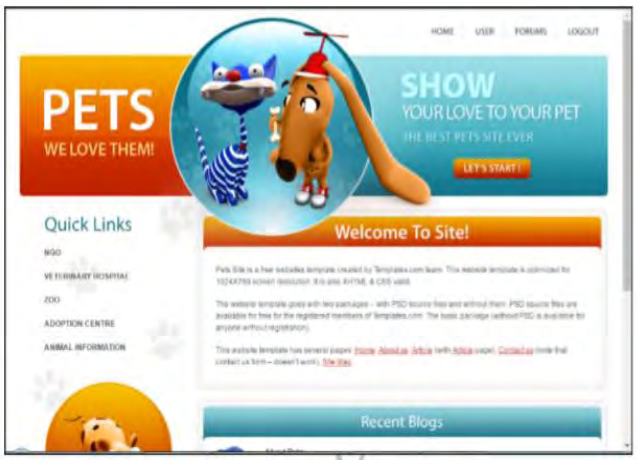


Fig. 4. Homescreen

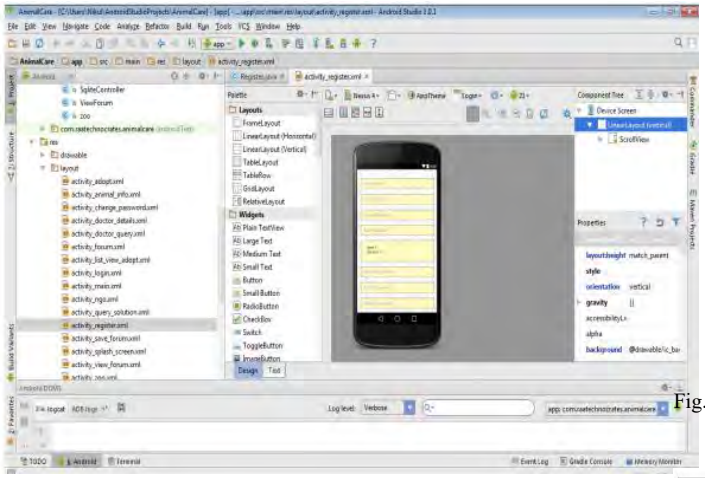


Fig. 5. Simulation Environment for system

The working and simulation environment of animal care system implemented using Android studio and the Website designed using PHP environment as shown in Figure 5.

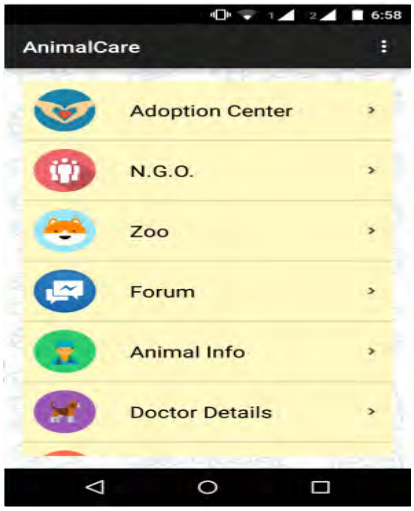


Fig. 6. Home Page Android Environment

Figure 6 shows the Homepage in Android Environment. After successful login we get this page where user can check all the centers, a user can ask some queries in the forum section and can check animal information.



Fig. 7. Adoption center details



Fig. 8. Adaption center details

Figure 8 is the page where an admin can add, update or delete any Adoption Centers.

IV.CONCLUSION AND FUTURE WORK

In this paper, we designed animal care system and analyzed clustering algorithm (K-means) in order to obtain optimal clusters for an animal care data set.

The future work mainly concentrates on improving the healthcare of all the animals. To give them a proper medical facility, to give them a house, to help other stray find animal centers, to make everything in one portal.

Currently, good and legal vet hospitals are difficult to find and with our proposed System we shall be able to find the legal vet Hospitals/animal centers. We can create a network of users who can assist an organization and or an individual, for spreading the awareness by helping the strays reach a proper care center, which is difficult in current scenario. A user or organization can connect to many different animal lovers. Direct contact to the vets which was not possible can be made available in the current system.

V.REFERENCES

[1] Jiawei Han and Micheline Kember, Data Mining: Concepts and Techniques Second Edition , Morgan Kaufmann Publishers. San Francisco: 2006.

[2] Ian. H and Eibe Frank, Data Mining Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers. San Francisco: 2005.

[3] Merlin Jacob, Anina John, “Improved Clustering of Documents using K-means Algorithm ”,International Journal of Engineering Research & Technology ,Volume. 5-Issue.07, July-2016.

[4] Sudhir Singh and Nasib Singh Gill, “Analysis and Study of K-Means Clustering Algorithm”, International Journal of Engineering Research & Technology, Vol.2 -Issue 7 (July - 2013).

[5] Gracy Joseph, Shravan Rao, Spurthi Shetty, Prof. Aruna Pavate,” Complete Animal Care System using Clustering Algorithm: CACS”, IJERT ICIATE – 2017 (Volume 5 – Issue 01)

[6] Mahmoud Elbattah, Owen Molloy,” Clustering-Aided Approach for Predicting Patient Outcomes with Application to Elderly Healthcare in Ireland”, The AAAI-17 Joint Workshop on Health Intelligence WS-17-09,pp 533-541

[7] Parasian D.P Silitonga,”Clustering of Patient Disease Data by Using K-Means Clustering”, International Journal of Computer Science and Information Security (IJCSIS), Vol. 15, No. 7, July 2017 pp 219-221

[8] Pranav Nerurkar, Archana Shirke, Madhav Chandane ,sunil Bhirud,” Empirical Analysis of Data Clustering Algorithms”, 6th International Conference on Smart Computing and Communications, ICSCC 2017, 7-8

[9] December 2017, Kurukshetra, India procedia Computer Science 00 (2018) 000–00

[10] Shenghui Wang, Rob Koopman ”Clustering articles based on semantic similarity” , Journal Scitometrics archive Volume 111 Issue 2, May 2017 Pages 1017-1031 Springer-Verlag New York, Inc. Secaucus, NJ USA

[11] Jain A. K.: Data clustering: 50 years beyond K-means. Pattern recognition letters 31(8) 651−666(2010).

[12] Marco Capó, Aritz Pérez, and Jose A. Lozano,” An efficient K - means clustering algorithm for massive data”, JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015

[13] Na, S., Xumin, L., & Yong, G. (2010, April). Research on k-means clustering algorithm: An improved k-means clustering algorithm. In Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on (pp. 63-67). IEEE.

[14] Godwin Ogbuabor1 and Ugwoke, F. N2,” Clustering Algorithm For A Healthcare Dataset Using Silhouette Score Value”, International Journal of Computer Science & Information Technology (IJCSIT) Vol 10, No 2, April 2018

[15] Zheng, B., Yoon, S. W., & Lam, S. S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Systems with Applications, 41(4), 1476-1482.

[16] Oyelade, O. J , Oladipupo, O. O , Obagbuwa, I. C ,”Application of k-Means Clustering algorithm for prediction of Students’ Academic Performance “ , (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, o. 1, 2010

[17] S. Na, L. Xumin and G. Yong, "Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm," *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*, Jinggangshan, 2010, pp. 63-67

[18] Huda Hamdan Ali1 , Lubna Emad Kadhum2,” K- Means Clustering Algorithm Applications in Data Mining and Pattern Recognition ”, International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2015): 78.96

[19] Yousef Farhang ,” Face Extraction from Image based on K-Means Clustering Algorithms”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 9, 2017

[20] Unnati R. Raval, Chaita Jani,” Implementing and Improvisation of K-means Clustering “, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.11, November- 2015, pg. 72-76

[21] Peter O. Olukanmi , Bhekisipho Twala,” K-means-sharp: modified centroid update for outlier robust k-means clustering ”, Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech) Bloemfontein, South Africa, November 29 - December 1, 2017

[22] Wang, T., Bucci, D. J., Liang, Y., Chen, B., & Varshney, P. K. “Exponentially Consistent K-Means Clustering Algorithm Based on Kolmogrov-Smirnov Test.”, 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).doi:10.1109/icassp.2018.8461730

[23] Huan, Z., Peng zhou, Z., & Zeyang, G. (2018),” K-means Text Dynamic Clustering Algorithm Based on KL Divergence”, 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS).doi:10.1109/icis.2018.8466385

[24] Al-Azzawy, D. S., & Al-Rufaye, F. M. L. (2017), “ Arabic words clustering by using K-means algorithm.”, 2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT).doi:10.1109/ntict.2017.7976098

Invisible Video Watermarking for Data Integrity and Security based on Discrete Wavelet Transform – A Review

Nitin A. Dhawas, Sambhaji A. Patil

Nutan Maharashtra Institute of Engineering & Technology, Talegaon, Dabhade, India
dhawasnitin2010@gmail.com, sapatil.mah@gmail.com

***Abstract*—To maintain the data integrity of transmitted video is very essential to secure any web-based system. Implementation of the invisible video watermarking is the best solution by using discrete wavelet transform gives us secured transformation in image processing applications. This method is also implemented as a 2-factor authentication (2FA) to secure the system. To protect the piracy of important data we implemented the above method. The major issue of illegal manipulation and sending/receiving of digital video becomes a big problem. To resolve this kind of issue, a latest and new technology has been suggested or we can say that proposed. Basically, it is a mechanism of implant copyrights information into a bit streams of any kind of video, the video may be any type i.e. it will be study purposed, presentation purposed, organization purposed and entertainment purposed. In this kind of scheme, the random segmentation and reconstruction of implants secrete data is done without having or knowing the original provider video. Throughout this process, confidential data or sensitive data is implanted in separate video frames using the DWT's frequency territory. DWT means Discrete Wavelet Transform. In the DWT's i.e. discrete wavelet transforms video watermark mechanism has been inaugurated for securing the confidential or sensitive data. We proposed a Haar wavelet ground (based) digital watermark mechanism to secure confidential or sensitive data.**

Keywords: Digital Video, DWT, Discrete Wavelet Transform; Principal Component Analysis.

I. INTRODUCTION

Watermarking is a technique for securing various copyright data such as text, image, audio, and video. Copyright protection is the major part of the watermarking. Since the 13th century, Digital Watermarking came into existence [1]. In case visible watermarks, visual patterns like logos were used to get inserted into digital data. To convey the hidden information, most watermarking systems involve marking alteration on the cover data. This technique is known to be invisible watermarks. Digital watermarks are found with the advancement of the internet. Thus, the thinking and way of the idea of watermarking into digital data can be naturally extended. Due to the escalation of digital data it may be an image, audio, and video, the popularity of digital watermarking has been increased. Recently the copyright of digital images and to verify the multimedia data integrity, numerous digital watermarking algorithms are developed [2][3]. Mainly due to the superior energy compaction property of wavelets, the use of the wavelets in the image and video coding has increased significantly over the years. The less need for a watermark is being invisible. Many authors require a watermark image to have the same quality as the original one; this has been found after studying several research papers. A number of properties of wavelet transform will be discussed in the next section. In one of the research, they proposed a novel watermarking method to embed QR codes in digital images. The mechanism is root or we can say that based on Discrete Wavelet Transform i.e. DWT.

The very fast expansion of the Internet in the last few years has fast increased the availability of digital data like audio, images, and videos to the public. The main goal of robust watermarking of the video is to implant information data within the video with a senseless form for the human understanding scenario but in a way that protects from attacks such as common video processing operations. The main approach is to produce an image/video that looks the same as the same to a human eye but still allows its positive identification in comparison with the owner's key if necessary. In fact any images or video watermarking technique can be extended to watermarking videos, but in reality video watermarking techniques need to meet other challenges like video coding technologies, large size of data, unseen watermarking detection, the unbalance between the motion and motionless region, some attacks like frame averaging, frame swapping, statistical analysis and other real-time features than that in image watermarking scheme. To be effective, the watermark should possess properties such as:

1) **Robustness:** The watermark should be impossible to remove even if the algorithmic principle of the watermarking method is public.
2) **Unambiguous:** The retrieved watermark should uniquely identify the copyright owner of the content, or in case of fingerprinting applications, the authorized recipient of the content.
3) **Loyalty:** A watermark has a high reliability if the degradation it causes is very difficult to perceive for the viewer.
4) **Computational Cost:** Embedding and extraction of the watermark from the video both should be fairly fast and should have low computational complexity
5) **Interoperability:** Watermark system must be interoperable for the compressed and decompressed operations.
6) **Universal:** The same digital watermarking algorithm needs to be applicable to all three media under consideration. This feature is favorable for the implementation of audio and image/video watermarking algorithms on common hardware as well.

7) **Unobtrusive:** The watermark needs to be perceptually invisible.

II. LITERATURE REVIEW

[1] Proposed a practical video watermarking technique on the compressed domain that is real-time and robust against video processing attacks. In particular, they focused on video processing that is commonly used in practice such as downscaling resolution, format changing, and transcoding. [2] Proposed QR Code (Quick Response Code) embedded technique for invisible watermarking by using Discrete-Fourier-Transform (DFT) compare with Discrete-Wavelet-Transform (DWT). [3] This proposed a novel watermarking method to embed QR codes in digital images. The method is based on a discrete wavelet transform (DWT). [4]In this it proposed an innovative practical video watermarking scheme based on MPEG-2. Watermark can be extracted blindly without original images and complete decoding is unnecessary. [5] Proposed a hybrid digital video watermarking scheme based on Discrete Wavelet

Transform (DWT) and Principal Component Analysis (PCA). [6] In this they proposed video data embedding scheme the embedded secret data is randomly segmented and reconstructed without knowing the original host video. Secret data is embedded in individual video frames using the frequency domains of DWT. [7] Image is represented as a two dimensional (2D) array of coefficients, every coefficient representing the brightness level at that point. Most natural images have smooth color variations, with the fine details being described as sharp edges in between the smooth variations. Technically, the smooth variations in color could also be termed as low-frequency components and therefore the sharp variations as high-frequency components. The low-frequency components constitute the base of an image, and also the high-frequency components add upon them giving a detailed image. Hence, the averages/smooth variations are demanding a lot of importance than the details [imp DWT]. DWT is used to implement a simple watermarking scheme. The 2-D discrete wavelet transforms (DWT) decomposes the image into sub-images.

The approximation looks like original only on the 1/4 scale. The 2-D DWT is an application of the 1-D DWT in both the horizontal and also the vertical directions. The DWT decomposes an image into a lower resolution approximation images as well as horizontal vertical and diagonal detail components. Due to its excellent spatial-frequency localization properties, DWT is very suitable to identify areas in the host video frame where a watermark can be embedded imperceptibly. Embedding the watermark in low frequencies obtained by wavelet decomposition increases the robustness with respect to attacks that have low pass characteristics like lossy compression, filtering, and geometric distortions. The video is nothing but a collection of still images. The original video is converted into frames discrete wavelet transform (DWT) and principal component analysis (PCA) is applied to each frame. The watermark image is converted into vectors and embedded in the low-frequency DWT sub-bands of each decomposed frame. Embedding the watermark in both LL and HH makes the scheme robust to a variety of low and high-frequency characteristic attacks [10]. Then inverse DWT and inverse PCA is applied to get watermarked video. One frame is chosen from the watermarked video for implementation.

III. APPLICATIONS OF DIGITAL VIDEO WATERMARKING

The major applications of digital video watermarking include copyright protection, video authentication, broadcast monitoring, copy control, fingerprinting, taper resistance, video tagging, ownership identification and enhance video coding. [20][21][23] Some of them are explained below:

1) **Copyright protection:** In digital multimedia, watermarking is used as copyright protection to identify the copyright owner.

2) **Video authentication:** Authentication means storing the signature into the header section, but the header field still is prone to tempering. So we can directly embed this type of authentication information directly as a watermark.

3) **Broadcast monitoring:** In television network, different products are distributed over the channel. A broadcast observation system must be built in order to check the entire broadcasted channel. Watermark is used for this type of broadcast monitoring system by putting a unique watermark for each video to broadcast.

4) **Copy control:** Watermarking system has the available technologies in which the information is secured into the header and it prevents from copying of that data.

5) **Fingerprinting:** Pay-per-view and Video-on-demand are two real-time applications of video streaming, in which digital watermarking is used to enforce a fingerprinting policy.

6) **Data Hiding (Covert Communications):** The transmission of private data is probably one of the earliest applications of watermarking. As one would probably have already understood, it consists of implanting a strategic message into an innocuous one in a way that would prevent any unauthorized person to detect it.

7) **Medical Safety:** Embedding the date and patient's name in medical images could increase the confidentiality of medical information as well as the security

IV. TECHNIQUES OF VIDEO WATERMARKING

Current video watermarking techniques can be grouped into two major classes; spatial-domain watermarking techniques and frequency-domain watermarking techniques. Spatial-domain techniques embed a watermark in the frames of a given video by modifying its pixels directly. These techniques are easy to implement and require few computational resources; however, they are not robust against common digital signal processing operations such as video compression. On the other hand, transform-domain watermarking techniques modify the coefficients of the transformed video frames according to a predetermined embedding scheme. The scheme disperses the watermark in the spatial domain of the video frame, hence making it very difficult to remove the embedded watermark.

1) **Spatial Domain Video Watermarking Techniques:** The watermark design and the watermark insertion procedures do not involve any transforms. Simple techniques like addition or replacement are used for the combination of the watermark with the host signal and embedding takes place directly in the pixel domain.

2) **Least Significant Bit modification (LSB):** Least Significant Bit (LSB) method is the simplest technique of this domain. In this scheme, the watermark is simply embedded into the least significant bits of the original video or flips the LSB. Due to its simplicity, it is the most

popular scheme, but some limitations are also there like, poor quality of the produced video, inefficient in dealing with the various attacks, least robustness and lack of imperceptibility.

3) **Correlation-based techniques:** A pseudo-random noise (PN) pattern W(x, y) is added to the cover image I(x, y), according to the equation shown below

$$Iw(x,y)=I(x,y) + k * W(x,y) \ldots\ldots\ldots (1)$$

In equation (1), k is a gain factor and Iw is the watermarked content. As we increase the value of k, it will expense the quality of watermarked contents.

4) **Frequency Domain Video Watermarking Techniques:** In frequency domain techniques, the watermark is embedded by modifying the transform coefficients of the frames of the video sequence. The most commonly used transforms are the Discrete Fourier Transform (DFT), the Discrete Cosine Transform (DCT), and the Discrete Wavelet Transform (DWT). Generally, the main drawback of transform domain methods is their higher computational requirement.

This approach first extracts the brightness of the watermarked frame, computing its full-frame DFT taking the magnitude of the coefficients. The watermark is composed of two alphanumerical strings. The DFT coefficient is altered, then IDFT. Only the first frame is watermarked, which was composed of twelve frames, leaving the other ones uncorrupted. It is good robustness to the usual image processing as linear/non-linear filtering, sharpening, JPEG compression and resist to geometric transformations like scaling, rotation, and cropping. The watermark design and the watermark insertion procedures do not involve any transforms. Simple techniques like addition or replacement are used for the combination of the watermark. DFT-based watermarking scheme with template matching can resist a number of attacks, including pixel removal, rotation, and shearing. The purpose of the template is to enable resynchronization of the watermark payload spreading sequence. It is a key-dependent pattern of peaks, which is also embedded into DFT magnitude representation of the frame. Discrete Cosine Transform (DCT) is an important method for video watermarking. A lot of digital video watermarking algorithms embed the watermark into this domain. The usability of this transform is because that most of the video compression standards are based on DCT and some other related transforms. In this domain, some DCT coefficients of the video are selected and divided into groups, and then the watermark bits are embedded by doing the adjustment in each group.

V. ALGORITHMS

Algorithms for watermarking using DWT and PCA

A. Embedding Procedure:

Step 1: Convert the n × n binary watermark logo into a vector wm= { w1, w2,……, wn × n } of, 0's and, 1's.
Step 2: Divide the video (2N × 2N) into distinct frames.
Step 3: Convert each frame from RGB to HSV color format.
Step 4: Apply 1-level DWT to the value (V component) of each video frame to obtain four sub-bands LL, LH, HL and HH of size N x N.
Step 5: Divide the HH sub-band into k non-overlapping sub-blocks each of dimension n × n (of the same size as the watermark logo).
Step 6: The watermark bits are embedded with strength α into each sub-block by first obtaining the principal component scores by Algorithm 2. The embedding is carried out as equation 1.
$yi= yi + \alpha(wm) \ldots\ldots\ldots(1)$
Where yi represents the principal component matrix of the ith sub-block
Step 7: Apply inverse PCA on the modified PCA components of the sub-blocks of the HH sub-band to obtain the modified wavelet coefficients.
Step 8: Apply inverse DWT to obtain the watermarked value component of the frame. Then convert the video frame back to its RGB components.

B. Extraction Procedure:

Step 1: Divide the watermarked (and possibly attacked) video into distinct frames and convert them from RGB to HSV format.
Step 2: Choose the value (V) component of a frame and apply the DWT to decompose the V component into the four sub-bands LL, HL, LH, and HH of size N×N.
Step 3: Divide the HH sub-band into n × n non-overlapping sub-blocks.
Step 4: Apply PCA to each block in the chosen subband HH by using Algorithm 2.
Step 5: From the HH subband, the watermark bits are extracted from the principal components of each sub-block as in equation 2.
$y1=abs(y-y1)/\alpha\ldots$ (2) Where y1 is the watermark extracted.

VI. CONCLUSION

We focused on the discrete wavelet transform (DWT) approach or mechanism for video watermarking. We proposed a mechanism at the ground root on DWT to implant a watermark image which will be as large as the cover image.
We can change the singular value of the cover image. As the video is a sequence of a number of frames. We applied watermark on every frame and created a watermarked video. By using IDWT watermark is extracted from the watermarked video. Extracted watermark compared with an original watermark.
High PSNR and correlation coefficient of the watermarked image is important to check watermarked video quality and data integrity of video respectively.

REFERENCES

[1] Sadik. A. M.Al-Taweel, Putra Sumari, "Robust Video Watermarking Based On 3D-DWT Domain," IEEE Communications, 2009.

[2] Min-Jeong Lee, Dong-Hyun Kim, Hae-Yeoun Lee, Kyung-Su Kim, Heung-Kyu Lee, "Real-time video watermarking system on the compressed domain for high-definition video contents: Practical issues", Digital Signal Processing 22 (2012) 190–198.

[3] Suppat Rungraungsilp, Mahasak Ketcham, Tanee Wiputtikul, Kanchana Phonphak , and Sartid Vongpradhip," Data Hiding Method for QR Code Based on Watermark by comparing DFT with DWT Domain", International Conference on Computer and Communication Technologies (ICCCT'2012) May 26-27, 2012 Phuket.

[4] Shanjun Zhang, Kazuyoshi Yoshino," DWT -Based Watermarking Using QR Code", Science Journal of Kanagawa University 19: 3-6 (2008).

[5] G. Langelaar, I. Setyawan, and R. Lagendijk, "Watermarking digital image and video data: A state-of-the-art overview," IEEE Signal Process. Mag., vol. 17, no. 5, pp. 20–46, Sep. 2000.

[6] Sanjana Sinha, Prajnat Bardhan, Swarnali Pramanick, Ankul Jagatramka, Dipak K. Kole, Aruna Chakraborty, "Digital Video Watermarking using Discrete Wavelet Transform and Principal Component Analysis" International Journal of Wisdom Based Computing, Vol. 1 (2), August 2011.

[7] S. Nafees Ahmed, B. Sridhar, Dr. C. Arun, "Robust Video Watermarking based on Discrete Wavelet Transform" International Journal of Computer Network and Security (IJCNS) Vol 4. No 1. Jan-Mar 2012.

[8] M.D. Swanson, M. Kobayashi, and A.H. Tewfik, Multimedia data-embedding and watermarking technologies, Proc. of the IEEE, vol. 86, no. 6, pp. 1064 –1087, Jun. 1998.

[9] I. Cox, J. Killian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," IEEE Trans. Image Process., vol. 6, no.12, pp. 1673 –1687, Dec. 1997.

[10] F. Hartung and B. Girod, Digital Watermarking of MPEG-2 Coded Video in Bitstream Domain, Proc. IEEE International

[11] Conference on Acoustics, Speech and Signal Processing, ICASSP-97, Vol.4, No.4, pp.2621-2624

[12] F. Hartung and B. Girod, "Watermarking of uncompressed and compressed video," Signal Process., vol. 66, no. 5, pp. 283– 301, May 1998.

[13] G. Langelaar and R. Lagendijk, Optimal differential energy watermarking of DCT encoded images and video[J], IEEE Transaction on Image Processing, vol. 10, no. 1, pp. 148–158, Jan. 2001.

[14] Sun T F, Jiang X H, Shi S, et al. A novel differential energy video watermarking based on Watson visual model [C] International Symposium on Electronic Commerce and Security. Nanchang, China: Nanchang Hong Kong University, 2009: 179 – 183.

[15] International Journal of Wisdom Based Computing, Vol. 1 (2), August 2011" Digital Video Watermarking using Discrete Wavelet Transform and Principal Component Analysis" Sanjana Sinha, Prajnat Bardhan, Swarnali Pramanick, Ankul Jagatramka, Dipak K. Kole, Aruna Chakraborty. Department of Computer Science & Engineering St Thomas" College of Engineering and Technology Kolkata, India.{aruna.stcet@gmail.com, dipak.kole@gmail.com, sanjanasinha89@gmail.com}

[16] Palaiyappan, Raja Jeya Sekhar. "A Block-Based Novel Digital Video Watermarking Scheme Using DCT" IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) e-ISSN: 2278-2834, p- ISSN: 2278-8735. Volume 5, Issue 2 (Mar. - Apr. 2013), PP 34-44 www.iosrjournals.org www.iosrjournals.org

[17] Sonjoy Deb Roy, Xin Li, Yonatan Shoshan, Alexander Fish, Member, IEEE, and Orly Yadid Pecht."Hardware Implementation of a Digital Watermarking System for Video Authentication" IEEE transactions on circuits and systems for video technology, vol. 23, no. 2, Feb 2013.

[18] Vipula Madhukar Wajgade1, Dr. Suresh Kumar2 1MTech(CSE) IInd Year, SGT Institute of Technology And Management, Gurgaon, Haryana, India 2HOD, Department Of Computer Science, SGT Institute of Technology And Management, Gurgaon, Haryana, India."Enhancing Data Security Using Video Steganography "International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250 -2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 4, April 2013) 549

[19] Harshitha K M, Dr. P. A. Vijaya Abbas Cheddad, Joan Condell, Kevin Curran, Paul Kevitt, "Enhancing Steganography In Digital Images". Proc. Canadian Conference on Computer and Robot Vision. "Secure Data Hiding Algorithm Using Encrypted Secret message" International Journal of Scientific and Research Publications, Volume 2, Issue 6, June 2012 1 ISSN 2250-3153 www.ijsrp.org.

[20] S. Tripathi, R.C. Jain," Novel DCT and DWT based Watermarking Techniques for Digital Images", The 18th IEEE International Conference on Pattern Recognition, 2006

[21] F. Perez-Gonzalez and J.R. Hernandez, "A tutorial on Digital Watermarking", Dept. Technologies delas Communications, ETSI Telecom., Universidad de Vigo, 36200 Vigo, Spain.

[22] Darshana Mistry, "Comparison of Digital Water Marking methods", International Journal on Computer Science and Engineering, Vol. 02, No. 09, pp 2905-2909,2010.

[23] M. D. Swanson, M. Kobayashi, and A. H. Tewfik, "Multimedia data-embedding and watermarking technologies," Proc. IEEE, vol. 86, no. 6, pp. 1064–1087, Jun. 1998.

[24] "Special Issue on Identification and protection of multimedia information," Proc. IEEE, vol. 87, no. 7, Jul. 1999.

[25] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding: a survey," Proc. IEEE, vol. 87, no. 7, pp.1062–1078, Jul. 1999.

[26] D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," Proc. IEEE, vol. 87, no. 7, pp. 1167–1180, Jul. 1999.

[27] F. Hartung and M. Kutter, "Multimedia watermarking techniques, "Proc. IEEE, vol. 87, no. 7, pp. 1079–1107, Jul. 1999.

[28] Paul, R.T., "Review of robust video watermarking techniques. IJCA Special Issue on Computational Science", 3: p. 90-95,2011.

[29] Hartung, F. and B. Girod, "Watermarking of the uncompressed and compressed video. Signal processing", 66(3): p. 283-301,1998.

[30] Pereira, S., J.J. Ruanaidh, F. Deguillaume, G. Csurka, and T. Pun. "Template-based recovery of Fourier-based watermarks using log-polar and log-log maps". in Multimedia Computing and Systems, 1999. IEEE International Conference on.IEEE, 1999.

[31] Deguillaume, F., G. Csurka, J.J. O'Ruanaidh, and T. Pun. "Robust 3D DFT video watermarking".in Electronic Imaging'99. International Society for Optics and Photonics, 1999.

[32] Liu, Y. and J. Zhao, "A new video watermarking algorithm based on 1D DFT and Radon transform. Signal processing", 90(2): p. 626-639,2010.

[33] Elbasi, E., Robust video watermarking scheme in transform domains, City University of New York: NEW YORK. 2007.

[34] Cox, I.J., M.L. Miller, J.A. Bloom, and C. Honsinger, Digital watermarking. Vol. 53, 2002.

[35] Wu, C., Y. Zheng, W. Ip, C. Chan, K. Yung, and Z. Lu, "A flexible H. 264/AVC compressed video watermarking scheme using particle swarm optimization based dither modulation. AEU-International Journal of Electronics and Communications", 65(1): p. 27-36,2011.

[36] Elrowayati Ali A., Z.Z.S., Abdulali Mohmed A., "Copyright Protecting Using Secure Watermarking Images in DWT Domain. 2012.

[37] El'Arbi, M., C.B. Amar, and H. Nicolas. "Video Watermarking Based on Neural Networks". in Multimedia and Expo, 2006 IEEE International Conference on. IEEE, 2006.

[38] Gaobo, Y., S. Xingming, and W. Xiaojing. "A genetic algorithm based video watermarking in the DWT domain". in Computational Intelligence and Security, 2006 International Conference on. IEEE, 2006.

[39] Wang, L., H. Ling, F. Zou, and Z. Lu, "Real-time compressed-domain video watermarking resistance to geometric distortions. IEEE Multi-Media", 19(1): p. 0070,2012.

[40] Ansari, R., M. Devanalamath, M. Hussain, and V.P. Prabha, Robust BPSO and Scene Change Based Digital Video Watermarking Algorithm in DWT-DFT-SVD Domain, in Security in Computing and Communications, Springer, 2013.

[41] Dolatabadi, Z.S.S., A.B.A. Manaf, and M. Zamani, "Using Three Levels DWT to Increase Robustness against Geometrical Attacks. International Journal of Advancements in Computing Technology (IJACT). ISSN": p. 2233-9337,2013.

[42] Fridrich, J., M. Goljan, Q. Chen, and V. Pathak. "Lossless data embedding with file size preservation". in Electronic Imaging 2004. International Society for Optics and Photonics, 2004.

Implementation of Agricultural Field Monitoring and Control of Irrigation System using IOT

Dhaval Desai, LakhanKabra, Mohit Jain, Mr. Shailesh Sangle

Computer Department, Thakur College of Engineering & Technology, Kandivali (East), Mumbai-400101, India dhavalrox16@gmail.com, jmohit083@gmail.com, lakhan.kabra.431997@gmail.com,shailesh.sangle@thakureducation.org

ABSTRACT—Agriculture is the primary occupation in our country for ages. But now due to migration of people from rural to urban there is hindrance in agriculture. So to overcome this problem a smart agriculture technique is used with the help of IOT. This project includes various features like GPS based remote controlled monitoring, moisture & temperature sensing, and proper irrigation facilities. It makes use of wireless sensor networks for noting the soil properties and environmental factors continuously. Various sensor nodes are deployed at different locations in the farm. Controlling these parameters are through any remote device or internet services and the operations are performed by interfacing sensors, Wi-Fi, with microcontroller.

Keywords Smart Agriculture, IOT, GPS, Internet, Wi-Fi, micro-controller.

1. INTRODUCTION

The world is trending towards new technologies and implementations and in order to be at par it is necessary to trend up in agriculture. Many researches are researching in the field of agriculture and most of them signify the use of wireless sensor network that collect data from different sensors deployed at various nodes and send it through the wireless protocol. The collected data provide the information about the various environmental factors. Monitoring the environmental factors is not the complete solution to increase the yield of crops. There are number of other factors that decrease the productivity. Hence, automation must be implemented in agriculture to overcome these problems. In order to provide solution to such problems, it is necessary to develop an integrated system which wills improve productivity in every stage. Hence, project deals about developing smart agriculture using IOT which will help in welfare of the farmers.

2. LITERATURE SURVEY

Title: IOT based Agricultural Field Monitoring and Controlling of Irrigation System

Author: Jeevanandam Jotheeswaran, Dr. S. Koteeswaran

Description: The collected data provide the information about the various environmental factors. Monitoring the environmental factors is not the complete solution to increase the yield of crops.

Development: .Controlling of all these operations will be through any remote smart device or computer connected to Internet and the operations will be performed by interfacing sensors, Wi-Fi and actuators with Arduino

Feature: Controlling of all these operations will be through any remote smart device or computer connected to Internet and the operations will be performed by interfacing sensors, Wi-Fi and actuators with Arduino.

3. PROPOSED WORK

This proposed work includes an embedded system for automatic control of irrigation. This project has wireless sensor network for real-time sensing of an irrigation system. When the moisture level in the soil reaches below threshold value then system automatically switch ON the motor. When the water level reaches normal level the motor automatically switch OFF. The sensed parameters and current status of the motor will be displayed on user's android application.

In this proposed system we are using various sensors like temperature, humidity, soil moisture sensors which senses the various parameters of the soil and based on soil moisture value land gets automatically irrigated by ON/OFF of the motor. These sensed parameters and motor status will be displayed on user android application.

The rate of water system, physical variables and so on are constantly transferred in the server. The manual mode offers alternative to choose the rate of discharging water by pumps, length of water system and so on .In the principal stage the self-ruling mode chooses the rate of water system as indicated by the present physical parameters by the examination of past standard studies transferred at first in server. The following period of computerization will perceive the perfect rate of water system by utilizing machine realizing where the physical components, rate of water system and the rate of development in the primary stage are utilized as preparing information. The pumps can likewise be controlled from an inaccessible place by means of electronic applications or versatile applications.

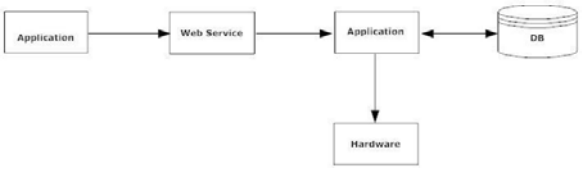


FIG.1 Lifecycle

4. SYSTEM ARCHITECTURE

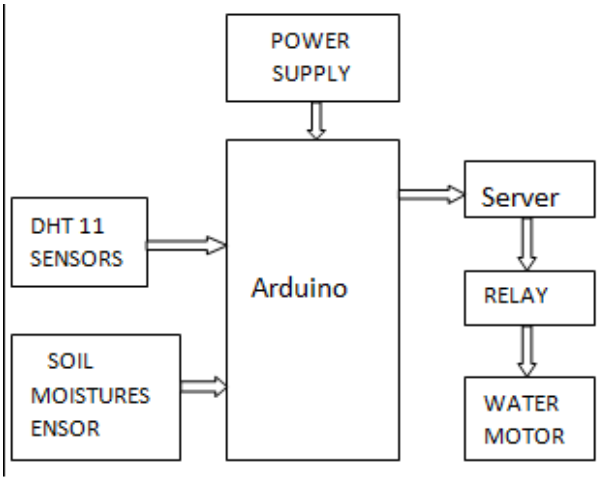


FIG.2 SYSTEM ARCHITECTURE

5. FEASIBILITY STUDY

A **feasibility study** is a preliminary study which investigates the information of prospective users and determines the resources requirements, costs, benefits and feasibility of proposed system. A feasibility study takes into account various constraints within which the system should be implemented and operated.

Thus we evaluated the feasibility of the system in terms of the following categories:

TECHNICAL FEASIBILITY:

Technology used at front end: CSS, HTML

Technology used at Back end: PHP, Mysql

Resources Required: Manpower, Programmers, testers, debuggers

Software required: Testing Tools (to perform black box and white box testing), Github

Editor: Sublime Text

Hardware required: PC for development, server for deployment

Infrastructure: For scaling up the project

ECONOMIC FEASIBILITY:

The Cost Benefit Analysis summarizes the revenues and costs involved with the proposed project. As the proposed system will be used for the benefits of the vagrants, no additional cost will be paid by them. Hardware system is included in our project, so the software cost gets minimized. Only minimal amount will be required by certain softwares. Hence, our system is Cost Efficient

LEGAL FEASIBILITY:

Data on social media which is publicly will be extracted so that we can apply machine learning algorithms on it and classify it according to it's sentiment. The proposed was to make a simplified web application. It is simpler to operate and can be used in any webpages. It is free and not costly to operate.

SCHEDULING FEASIBILITY:

Report about the researched work will be submitted by 26 March 2019.

A project will fail if it takes too long to be completed before it is useful. Typically, this means estimating how long the system will take to develop, and if it can be completed in a given period of time using some methods like payback period.

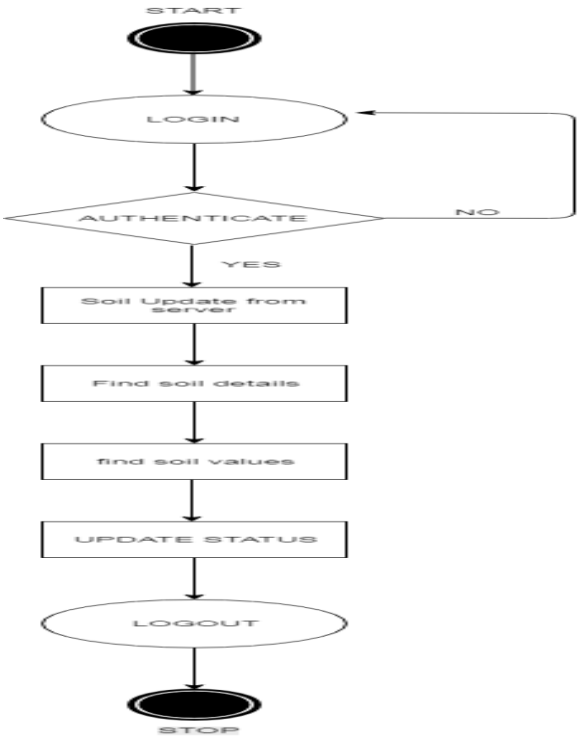


FIG.3 FLOWCHART

6. RESULT AND DISCUSSION

IOT is a shared Network of objects where these objects interact through Internet. One of the important applications of IOT is Smart Agriculture. Smart Agriculture reduces wastage of water and increases the crop yield. Here a system is proposed to monitor crop-field using sensors for soil moisture, humidity and temperature. By monitoring these parameters the irrigation system can be automated if soil moisture is low. Agriculture plays vital role in the development of agricultural country. In India about 70% of population depends upon farming and one third of the nation's capital comes from farming. Issues concerning agriculture have been always hindering the development of the country. The only solution to this problem is smart agriculture by modernizing the current traditional methods of agriculture. Hence the project aims at making agriculture smart using automation and IOT technologies. It includes smart irrigation with smart control and intelligent decision making based on accurate real time field data. Controlling of all these operations will be through any remote smart device or computer connected to Int.

7. CONCLUSION

Thus we can conclude it very well that it may be upgraded by building up this framework for extensive sections of land of land. Likewise, the framework can be incorporated to check the nature of the dirt and the development of harvest in each dirt. Additionally, the framework can be additionally enhanced by calculator learning calculations, which can learn and comprehend the prerequisites of the harvest, this would enable the field to be a programmed framework. The perceptions and results disclose to us that this arrangement can be executed for decrease of water misfortune and lessen the labor required for a field.

8. FUTURE SCOPE

For future developments, it can be enhanced by developing this system for large acres of land. Also, the system can be integrated to check the quality of the soil and the growth of crop in each soil. The sensors and microcontroller are successfully interfaced and wireless communication is achieved between various nodes. Also, the system can be further improved by adding machine learning algorithms, which are able to learn and understand the requirements of the crop, this would help the field be an automatic system. The observations and results tell us that this solution can be implemented for reduction of water loss and reduce the man power required for a field.

9. ACKNOWLEDGEMENT

We sincerely thank our guide Mr. Shailesh Sangle for his guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

We thank the HOD, Dr. Sheetal Rathi, Dean Academic, Dr. R.R Sedamkar, the Principal, Dr. B. K. Mishra and the college management for their support.

REFERENCES

[1]K. Lakshmisudha, Swathi Hegde, Neha Kale, Shruti Iyer, "Smart Precision Based Agriculture Using Sensors",

[2]International Journal of Computer Applications (0975- 3. 8887), Volume 146-No.11, July 2011

[3]Nikesh Gondchawar, Dr. R.S. Kawitkar, "IoT Based Smart Agriculture", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE),

[4] Issue 6, June 2016. 5. M.K. Gayatri, J. Jayasakthi, Dr.G.S. Anandhamala.

[5] "Providing Smart Agriculture Solutions to Farmers for Better Yielding Using IoT", IEEE International Conference on Technological Innovations in ICT for Agriculture and Rural Development (TIAR 2015).

[6] Chetan Dwarkani M, Ganesh Ram R, Jagannathan S, R.

[7] Priyatharshini, "Smart Farming System Using Sensors for Agricultural Task Automation", IEEE International Conference on Technological Innovations in ICT for Agriculture and Rural Development (TIAR 2015).

[8]S muthunpandian, S. Vigneshwaran, R. C Ranjitsabarinath, Y. Manoj kumar reddy, IOT Based Crop-Field Monitoring And Irrigation Automation, vol. 4, no. 19, April 2017.

[9]Joaquín Gutiérrez, Juan Francisco Villa-Medina, Alejandra Nieto-Garibay, Miguel Angel Porta-Gándara, "Automated Irrigation System Using a Wireless Sensor Network and GPRS Module", IEEE Transactions On Instrumentation And Measurement, vol. 17, 2017.

[10]Mohanraj, Ashokumarb Kirthika, J Naren, Field Monitoring and Automation using IOT in Agriculture Domain, vol. IJCSNS, no. 6, June 2015.

Review on Deep Reinforced Learning Based Text Summarization Models

P. G. Magdum, Sheetal Rathi
Department of Computer Engineering, Rajendra Mane College of Engineering and Technology, Ratnagiri, India
Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India
magadumpg@rmcet.com, sheetal.rathi@thakureducation.org

***Abstract*—The huge amount of information is growing day by day, due to development in information technology. It becomes very difficult to refer to such vast data and create concise and relevant information to make the right decision. The summarization is the best technique to extract relevant and concise information from huge data. The extractive and abstractive text summarization methods help to relevant information from text documents. The number of techniques is used from over the year like heuristic, fuzzy based techniques, machine learning, supervised and unsupervised techniques. But still, there are many issues in text summarization. The reinforced deep learning is an important field of artificial intelligence. It is capable of building a strong model to handle any application. The reinforcement learning is useful to resolve some issues in said techniques. In this paper, we have reviewed different reinforcement learning model and analysis performance.**

Keywords-- Reinforced learning, supervised learning, unsupervised learning

I. INTRODUCTION

The automatic summarization widely used in Natural Language processing for extracting potential relevant information from text documents. It is useful in precise and concise information access application. These application inputs data from input sources like news, social media, e-books, blogs, websites etc. The single document summarization is extracting shorter version of input document without changing meaning.it is very basic task using from many years. To find a summary for text documents traditional methods are used. But the better accuracy is generated when neural based architectures are used. The neural based model trains the model by using strong datasets.

For both extractive and abstractive summarization kind of summaries can be obtained by neural based methods [1]. The extractive text summarization extracts some relevant sentences
text document as a summary. The extractive summarization decides which sentence should be included in summary by checking sentence score. This sentence score is obtained by using some extracting features of the original text.

The abstractive text summarization produces new phrases using the words which are not in the input text document. Neural network model based on the encoder-decoder model with attention mechanism having the capability to produce better abstractive text summary [2]. This model produces a good ROUGE score. The quality of output may vary according to the use of dataset used. The use of traditional, machine learning, supervised and unsupervised learning techniques for text summarization still causes many issues like redundancy, semantic similarity, handling long sentences etc. The reinforcement learning achieves good performance of various dataset of summarization.

II. REINFORCEMENT LEARNING

The deep reinforcement learning is goal oriented technique. The idea behind the reinforcement learning is the agent will learn the current situation by communicating with it and receive the rewards signals for actions [3]. The learner discover which actions gives the most rewards by trying them. The trial and error and delayed rewards are two important rewards characteristics of reinforcement learning. The deep learning enables reinforcement learning to find solution for decision making problems.

There are two reinforcement learning algorithms.

i. Deep Q- learning
ii. A3C learning

The state diagram of deep reinforcement learning is as shown in fig. 1

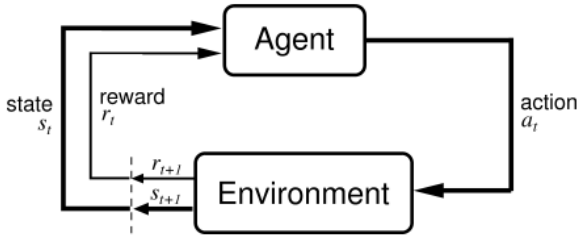


Fig. 1 State Transition Diagram for Reinforcement Learning (Source: Analyticsvidhya [4])

At time t, the agent receive the state S_t from the environment. The agent uses its own rule to choose respective action a_t. When the action is executed then environment changes step and provide next state S_{t+1} along with rewards r_{t+1}. The agent uses knowledge of state transition of form (St, at, S_{t+1}, a_{t+1}). Agent learn using these transition and improve its rules.

III. EVALUATION METHODS OF TEXT SUMMARIZATION

In text summarization, the main critical task is an evaluation of generated summary by different algorithms. There are intrinsic and extrinsic methods used for evaluation [5]. Since deep reinforcement learning systems require a large amount of data for training. Now a day commonly used datasets are DUC 2001, DUC 2002, DUC 2003, DUC 2004, DUC 2005, DUC 2006, and DUC 2007 also ACL-ARC, Wikipedia, CNN and Daily Mail news datasets developed an as good source for training and testing purposes also many authors utilized Gigaword dataset for abstractive as well as extractive summarization.

ROUGE

ROUGE is popular tool used to evaluate the text summary generated by text summarization methods. The ROUGE stands for Recall Orientated Understanding Gisting Evaluation. There are two version of ROUGE tools ROUGE 1.0 and ROUGE 2.0. The ROUGE having three measure i.e. Precision, Recall and F-Measure.

Precision (p) is given as follows:

$$Precision(P) = \frac{No.of\ relevant\ items\ retrieved}{No.of\ retrieved\ items}$$

Recall(R) is given as follows:

$$Recall(R) = \frac{No.of\ relevant\ items\ retrieved}{No.of\ relevant\ items}$$

F-measure is given as follows:

$$F1 = \frac{P * R}{P + R}$$

The other versions of ROUGE are ROUGE-N, ROUGE- W, ROUGE-L, ROUGE- S and ROUGE- SU.

IV. Deep Reinforcement Learning Based Summarization Techniques

A. REFRESH [6]

In this research, the model is purposed to find extractive text summary by using sentence ranking techniques. This model optimizes reward function in reinforcement learning. It produces concise extractive and abstractive summary. This model uses CNN in sentence encoder to identify named entity, LSTM+RNN in document encoder and sentence extractor to remove problem of vanishing gradient during training long sequence.

Rewards calculation:

The reward can be calculated as

$$\nabla L(\theta) = -E_{y\sim p\theta}[r(y)] \quad (01)$$

Where, θ is model parameter, p is number of pages.

Agent updation:

The agent interact with environment consisting of documents. At the starting agent initialized randomly. It reads Document D and predict score for each sentence using certain policy. The updation of agent is done using following mathematical formulation.

$$\nabla L(\theta) = -E_{y\sim p\theta}[r(y)\nabla logp(y|D,\theta)] \quad (02)$$

Deep learning techniques used: CNN, LSTM+RNN

B. RES-QFR [7]

This research proposed question focused reward function which create concise, fluent and informative summary. It convert the human abstracts to set of cloze-style comprehensive question. It helps in producing summary.

Rewards calculation:

The reward can be calculated as

$$Ra(Y) = \frac{1}{K}\sum_{k=1}^{k} \log P(e_k^*\ |Y, Qk) \quad (03)$$

Where, θ is model parameter, p is number of pages.

Agent updation:

$$a_{k,i} = \frac{\exp(qk\ W^a h_i^s)}{\sum_{i=1}^{|s|}\exp(qk\ W^a h_i^s)} \quad (04)$$

Deep learning techniques used: Bidirectional LSTM and attention mechanism.

C. RLEF [8]

This research proposes the reinforcement learning based system that uses deep Q- network learning algorithm. It estimate the Q value to provide text summarization for given input document. This model embeds many features like content embedding vector and position embedding in deep reinforcement learning. The use of such embedding feature provide meaning and position of sentences.

Rewards calculation:

$Reward(state_t, state_{t+1}) = score(state_{t+1,Hd}) - score(state_{t,Hd})$ (05)

Where score measures the quality of partial summary, Hd human reference.

Agent updation:

The role of agent is produce summary by using sentence selector. The updation of agent is done using following mathematical formulation.

$DA_w = sigmoid(Wda_{out}, Hidden_{da})$ (06)

$Hidden_{da} = tanh(wda_{hidden,}\ INPUT_{da})$ (07)

$INPUT_{da} = concat(DocVec, ContVec\ of\ At)$ (08)

Deep learning techniques used:

The model use RNN in Q- network to find next state.

D. $DQN_{rnn\text{-}rnn}$ [9]

This approach based on Deep Q network which can reduces redundancy of sentences in Q value approximation. It produces good ROUGE score. This approach consists of two hierarchical network architectures. One generate informative features from the documents to represent state of Deep Q network. And other produce a list of potential actions from sentences in the document for the Deep Q network. This model used to generate extractive kind of text summary.

Rewards calculation:

The reward can be calculated as

$$rt = \sigma(W\ rx\ x\ t + W\ rh\ h\ t{-}1 + b\ r) \quad (09)$$

Where, Xt is input vector, rt is reset gate vector and ht id hidden gate output vector. W and b are parameters of Bidirectional GRU-RNN.

Deep learning techniques used: RNN-CNN

E. RL [10]

This research proposes attentional, RNN based Encoder-decoder model for abstractive text summarization. It reduces the redundancy and inconsistent phrases. This model combines supervised word prediction and deep reinforcement learning for better accuracy.

Rewards calculation:

The reward can be calculated as

$$R(Y) = softmax(Wout[h_t^d||c_t^e||c_t^d] + bout) \quad (10)$$

Where R(y) is reward, c content vector, h hidden input state.

Agent updation:

$$\alpha_{tt'}^d = \frac{\exp(e_{tt'}^d)}{\sum_{j=1}^{t-1}\exp(e_{tt'}^d)} \quad (11)$$

Where t is decoding step, d is document, e attention score of hidden layer.

Deep learning techniques used: RNN, bidirectional LSTM.

V. COMPARATIVE PERFORMANCE OF REINFORCEMENT LEARNING MODEL

All above discuss reinforcement model are compared considering some point as shown in table 1.

TABLE I COMPARATIVE PERFORMANCE OF MODELS

Model	Deep model used	Extractive/ abstractive	Single / multi document	Dataset	ROUGE Performance for dataset		
					ROUGE-1	ROUGE - 2	ROUGE -L
REFRESH [6]	CNN, RNN+ LSTM	Extra-ctive	single	CNN and DailyMail	40.0	18.2	36.6
RES-QFR [7]	Bidirectional LSTM	Extra-ctive	single	CNN	31.7	11.6	21.5
RLEF [8]	RNN	Extra-ctive	single	DUC 2001, 2002, Wikipedia, ACL-ARC data	21.63		
DQN_{RNN} [9]	RNN,CNN	Extra-ctive	single	CNN/Daily corpus, DUC 2002, 2004	43.4	14.1	39.2
RL[10]	RNN+LSTM	Abstra-ctive	single	CNN Daily Mail, New York Times	47.2	30.5	43.3

VI. CONCLUSION

The abstractive and extractive text summarization method using deep reinforcement learning can be optimized using ROUGE metrics. There is great improvement in performance of text summarization using deep reinforcement learning techniques. It produce concise, informative text summary for input text document. The techniques like Q network learning with different deep neural techniques provides solution to critical problem where many other basic techniques fails. It reduces inconsistency and redundancy of word. The use of reinforcement learning techniques may improve significant result in automatic text summarization. Using deep reinforcement learning, it is possible to propose hybrid model for abstractive and extractive text summarization. Also, it is possible for single and multi-document text summarization.

REFERENCES

[1]. Wan-Ting Hsu , Chieh-Kai Lin , Ming-Ying Lee , Kerui Min , Jing Tang , Min Sun,"A Unified Model for Extractive and Abstractive Summarization using Inconsistency Loss", Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers), pages 132–141 Melbourne, Australia, July 15 - 20, 2018. Association for Computational Linguistics

[2]. Ramesh Nallapati, Bowen Zhou, Cicero dos Santos,Caglar Gulcehre, and Bing Xiang, "Abstractive text summarization using sequence-tosequence RNNs and beyond", In Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL) 2016.

[3]. Xingxing Zhang and Mirella Lapata, "Sentence simplification with deep reinforcement learning". In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) 2017.

[4]. https://www.analyticsvidhya.com/blog/2017/01/introduction-to-reinforcement-learning- implementation/

[5]. Chin-Yew Lin, "ROUGE: A Package for Automatic Evaluation of Summaries", In Proceedings of Workshop on Text Summarization Branches Out, Post-Conference Workshop of ACL 2004, Barcelona, Spain.

[6]. Shashi Narayan, Shay B. Cohen, Mirella Lapata, "Ranking Sentences for Extractive Summarization with Reinforcement Learning", NAACL 2018.

[7]. Kristjan Arumae Fei Liu, "Reinforced Extractive Summarization with Question-Focused Rewards", Proceedings of ACL 2018, Student Research Workshop, pages 105–111 Melbourne, Australia,
July 15 - 20, 2018. Association for Computational Linguistics.

[8]. Gyoung Ho Lee, Kong Joo Lee ,"Automatic Text Summarization Using Reinforcement Learning with Embedding Features", Proceedings of the The 8th International Joint Conference on Natural Language Processing, pages 193–197, Taipei, Taiwan,
November 27 – December 1, 2017c 2017 AFNLP.

[9]. Yao, K., Zhang, L., Luo, T., & Wu, Y. (2018). Deep reinforcement learning for extractive document summarization. Neurocomputing, 284, 52–62.doi:10.1016/j.neucom.2018.01.020.

[10]. Romain Paulus, Caiming Xiong & Richard Socher, "A Deep Reinforced Model for Abstractive Summarization", ICLR 2018

Implementation of Home Automation System Using Smart Mirror

Prasad Pol, Vijay Borkar, Mahesh Bakve, Manil Kapadia, Mr. Vikas Singh (Guide)

Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India

prasadpol18@gmail.com, vip14327@gmail.com, mahesh26895@gmail.com, kapadiamanil@gmail.com

***Abstract*— Our way of life has evolved in such a way that time optimization is the most important thing. Every day we see more and more connected devices. There is no end to objects that can be made "smarter", some are better suited to this than others. Mirrors, for example, provide a great surface for displaying information and interacting with it. Most people have mirrors at home, the concept of smart mirror that you can interact with is attractive and has been fantasized in many futuristic movies.**

We present the development of a different application integrating interactive information services, offered via a user interface on the surface of a mirror. Our work is based on the idea that we all look in the mirror when we go out, so why does not the mirror become intelligent? Smart mirrors, such as Magic Mirror and Home Mirror, have recently begun to be developed by members of the Maker community, with varying degrees of interactivity. However, until now, the characteristics of these mirrors have been limited. This article explains how a smart mirror was created from scratch using a Raspberry Pi for hardware and custom software created on Raspbian, a Linux distribution.

The goal of the project was to create a Smart Mirror device with which users could interact, but also to continue developing technologies to enable them to install and develop their own applications. The framework will provide basic services, such as customized meteorological data presentation, time and date, and incorporate additional features, such as mobile reminder service and social networking. The data or information includes news, notifications, etc.

Home automation has been around for a long time and it is about turning the house into a smart unit in order to increase the comfort and efficiency of the home. Some of the typical applications are automatic lights, smart thermostats, alarms, blinds. In this we will describe how we can use home automation and the intelligent mirror together, that is, the automation of the home through the intelligent mirror.

Keywords — Smart Mirror, IOT, Raspberry pi , Wi-Fi devices, Home Automation

I. INTRODUCTION

This project was inspired by a most popular Raspberry Pi-powered voice-activated home automation center, as well as other devices discovered while browsing the DIY section of the popular "Smart Mirror By" website. Evan Cohen ". The device has a web page that displays weather information and has a very futuristic look. We really liked this project and we thought we could do it.

Improve it by adding interaction means to the device. We also discovered a similar project built using a Raspberry Pi minicomputer, but it was still a static panel without interaction. This is a new version of Smart Mirror with an operating system that would allow you to install applications that can develop as Android or iOS. The project covers a very wide range of domains in the areas of Internet, Maker culture and home automation.

In our project we can focus on the automation of the home, we can make a small model of home automation, we can make an intelligent mirror in it, we can make a small model of "Intelligent mirror". However, it would be very easy to write an application to turn on and off the lights in the room. These examples are only the tip of the iceberg, since they are new connected devices that emerge every day and that could interact with the mirror.

The objectives of the smart mirror should focus on the user's daily routine and provide a better understanding of the user and technology. The functionality must comply with these descriptions in the design. We will design a futuristic intelligent mirror that provides a natural interaction between the users and the environmental services of the home. The core of the mirror is based on a home automation system, which we developed to demonstrate the various functions provided by the mirror. The mirror screen is provided by a flat monitor, which transmits live continuously from a webcam connected to the mirror to a traditional mirror function.

The hardware requirements for this project are the usual bidirectional mirror, screen and Raspberry Pi configuration, but with a USB microphone to support voice commands. We use the webcam microphone because the audio signals would be sent through USB and we will not need an external sound card to process the audio. Although it is recommended to run it on a Raspberry Pi, it can also be run on other systems. This intelligent mirror is more advanced and more complete than most others, however, the software has been considered as a unique application that can do many things and this makes it difficult to write your own applications for the system.

A web application is created using JavaScript, which is then hosted on a server on the local host for local use. This web application has many other APIs installed on it, such as Google maps, Sound Cloud, traffic rates, Fit Bit and much more, which provides a perfect data analysis for the user of the current day. All functions work entirely in the voice, which guarantees that the user can use it while doing multiple tasks with other things, which makes it more useful in conditions of a lot of work. When

the web application is hosted on a web server, the user can control it from anywhere through the Internet.

So, with this we can use the smart mirror in our daily life and we can save our time doing many jobs through the voice command and also through the application.

The main objective was not to make it look fresh or futuristic, but to help disabled or elderly people who may or may not have the ability to manually use appliances using voice navigation. Voice recognition is done by the Google speech recognition engine, which guarantees the most accurate recognition that is experienced, which makes it compatible with the user and the ease of access with less chance of failure in the system.

Another point was to ensure that the technology of home automation with the most up-to-date technology to date should be delivered to users at an affordable price. This would ensure that the advantages of its characteristics and power are for everyone.

How this system can be reprogrammed completely according to the needs of the user due to the reduced complexity (Python programming language) and the powerful open source forum (Raspberry pi forum) to be affordable even after the installation of the smart mirror or after the sale.

Using the GPIO pins of the raspberry pi, you can control many electronic devices and circuits and control them from anywhere. This not only makes it efficient in the consumption of resources, but also makes its use intelligent.

II. Problem Identification and area

Our lifestyle has evolved in such a way that optimizing time is the most important thing. Every day we see more and more connected devices. There is no end of objects that could be made “smarter”, some being more suited to this than others. Mirrors, for example, provide a large surface ideal for displaying information and interacting with. Most people have mirrors at home so the concept of a smart mirror that you can interact with is attractive and has been fantasized in many futuristic movies. We present the development of a different application that integrates interactive services of information, offered through a user interface on the surface of a mirror.

III. Objectives and expected outcome

The goal of the project was to create a Smart Mirror device that people could interact with but also to further develop the technology so that it would let you install and develop your own applications for it. The framework will offer basic services, like the presentation of personalized weather data, time, date and will incorporate some additional functionality, like reminder service by mobile synchronization and through social media. The data or information includes news feed, notifications and so on.

IV. Applications

Application of the project are as follows :

1. Perform Home Automation using voice commands
2. Displaying things on normal mirror
3. Displaying web based applications on smart mirror

V. Impact Analysis

The development is going on even if we add thing like hand gesture, voice authentication, user authentication the system will work without any error as the system is capable to handle all this changes and perform its regular task and make it adaptable to new environment

VI. Requirement Specification

The requirement for the project is as follows:

Hardware Requirement:

1. 2 Way mirror
2. Wooden Frame
3. Microphone
4. Raspberry pie
5. Relay module
6. Led Lights
7. Dc Motor
8. Battery

Software Requirement:

1. Node js
2. Annyang(Voice Recognition)
3. Python
4. Java Script
5. Kodi(OSMC)
6. STT engine
7. TTS engine

VII. FLOWCHART



VIII. RESEARCH / PROJECT WORK / TECHNICAL

LITERATURE SURVEY

Title : Smart Mirror For Smart Life

Publication year : 2017

Author :

1. Muhammad Mu'izzudeen
2. Yusri Shahreen Kasim

Description :

IoT is known for its advantage that can help simplify peoples everyday rouine. Hence the reserchers comes with proposed system called smart mirror.

Title :

Smart Reflect – A Modular Smart Mirror Application Platform

Publication year : 2016

Author :

1. Derrick Gold

2. David Sollinger

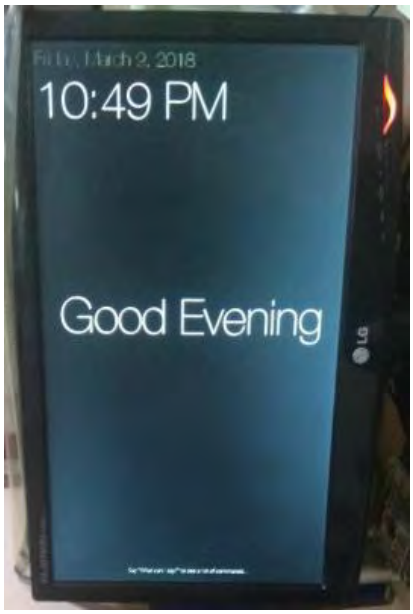
3. Indratmo

Description :

A smart mirror is a device that functions as a mirror with additional capability of displaying multimedia data such as text, images, videos.

IX.SAMPLE OUTPUT





X.FUTURE SCOPE

The speech recognition system can be used for entry of the data in data forms or admission forms, for the people who are not literate to write or Handicapped.

People who are on wheel chairs or old people who cannot stand up and switch the buttons on or are not literate enough to use Android phones or any high tech devices, can easily

It can save much time as we can give commands from long distance and hence keeping our homes safe and secure.

Due to automation it makes the whole system efficient and making a difference in Electricity Bill and Resources too.

XI.CONCLUSION

From this we can achieve comfort and efficiency at home and it can also help full for physically challenged people

One of the most needed thing for disabled/elderly people is to perform their daily thing when they are alone at home.

This will help them to do their daily activity without getting help form anyone and become self in dependant.

ACKNOWLEDGMENT

We sincerely thank our guide “Mr. Vikas Singh” for his guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

We thank the HOD, “Dr. Sheetal Rathi”, Dean Academic, “Dr. R.R Sedamkar”, the Principal, “Dr. B. K. Mishra” and the college management for their support.

REFERENCES

[1] BarisYuksekkaya, A. AlperKayalar, M. BilgehanTosun, M. KaanOzcan, and Ali ZiyaAlkar “A GSM, Internet and Speech Controlled Wireless Interactive Home Automation System”, 2006, IEEE Transactions on Consumer Electronics, Vol. 52(3) , pp. 837 - 843.

[2] RozitaTeymourzadeh,SalahAddinAhmed,KokWai Chan and MokVeeHoong , “Smart GSM Based Home Automation System”, 2013, IEEE Conference on Systems, Process & Control, Kuala Lumpur, Malaysia.

[3] A. Alheraish, “Design and Implementation of Home Automation System”, 2004, IEEE Transactions on Consumer Electronics ,Vol. 50(4) , pp. 1087-1092.

[4] M.Van Der Werff, X. Gui and W.L. Xu, “A Mobile based Home Automation System, Applications and Systems”, 2005, 2nd International Conference on Mobile Technology, Guangzhou, pp.5.

[5] Mahesh.N.Jivani, “GSM Based Home Automation System Using App-Inventor for Android Mobile Phone”, 2014, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 3(9), pp. 12121-12128.

[6] Faisal Baig, SairaBaig, Muhammad Fahad Khan, “Controlling Home Appliance Remotely through Voice Command”, 2012, International Journal of Computer Applications , Vol. 48(17), pp.1 - 5.

[7] S.R.Bharanialankar,C.S.ManikandaBabu, “Intelligent HomeApliance Status Intimation Control and System Using GSM”, 2014, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4(4), pp.554 - 556.

[8] RozitaTeymourzadef, Salah Addin Ahmed, KokWai Chan, MokVeeHoong, “Smart GSM Based Home Automation System”, 2013, IEEE Conference on Systems, Process & Control, Kuala Lumpur, Malaysia, pp.306 - 309.

[9] R.Piyare,M.Tazil, “ Bluetooth Based Home Automation System Using Cell Phone”, 2011 ,IEEE 15th International Symposium on Consumer Electronics, Singapore, pp. 192 - 195.

[10] Home Automation System via Bluetooth Home Network”, 2003, SICE Annual Conference, Fukui, Vol. 3, pp. 2824 - 2829.

[11] H. Brooke Stauffer “Smart Enabling System for Home automation”, 1991, IEEE Transactions on Consumer Electronics, Vol. 37(2) , pp. 29-35.

[12] Eddie M C Wong, “A Phone Based Remote Controller for Home and Office Automation”, 1994, IEEE Transactions on Consumer Electronics, Vol. 40(1), pp.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation

[13] G. Eason, B. Noble, and I.N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil

[14] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73

[15] I.S. Jacobs and C.P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[16] K. Elissa, “Title of paper if known,” unpublished.

[17] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[18] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[19] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

Android Based Heart Monitoring Systems

[1]Shraddha More, [2]Dr.Rajesh Bansode, [3]Anita Chaudhari, [4]Aruna Pavate, [5]Ms.Brinzel Rodrigues
[1,3,5] IT Department St. John College of Engineering & Management, Mumbai, India, [2,4] IT department
Thakur college of Engineering, Mumbai, India
[1]shraddham@sjcet.co.in,[2] rajesh.bansode1977@gmail.com, [3]anitac@sjcet.co.in, [4]arunapavate@ atharvacoe.ac.in, [5]brinzelr@sjcet.co.in

***Abstract*—Steady advances in wireless technology, medical sensors and interoperability of software creates exciting ways in improving the ways in which we provide emergency care. Nowadays Healthcare Atmosphere has developed technology oriented. People are facing a problem of unpredicted death due to the cause of a heart attack which is because of absence of medical care to patient at the right time. So proposed system designed to avoid such sudden death rates by using Heartbeat sensor (HBS) technology.**

In this proposed system a patient will be carrying hardware having sensors and android phone will be contained an application, the sensors will sense the heart rate of the patient and these data is transformed to android smart phone via GSM modem. The device even allows the patient to move freely and can be monitored continuously. The android phone will be contained an application which will receive heart rate, according to the received data respectively, and if any abnormalities are found, then the message contains patient's heart rate will be sent to the patient's doctor.

Keywords—Heart rate, patient, Heartbeat sensor, GSM modem, Body temperature, Remote Monitoring.

I. INTRODUCTION

In today's world, the maximum use of resource is always complimented. So, the use of wireless technology is enhanced to meet the need of remote control and monitoring. Remote patient monitoring (RPM) is a technology that enables us to monitor patient outside of clinic or hospital without having to visit a patient. It may increase access to health services and facilities while decreasing cost. Remote Patient Monitoring saves time of both patient and doctor, hence increasing efficiency and reliability of health services. Heartbeat and body temperature are the major signs that are routinely measured by physicians after the arrival of a patient [1]. Heart rate refers to how many times a heart contracts and relaxes in a unit of time (usually per minute). Heart rate varies for different age groups. For a human adult of age 18 or more years, a normal resting heart rate is around 72 beats per minute (bpm). A lower heart rate at rest implies more efficient heart function and better cardiovascular fitness. Babies have a much higher rate than adults around 120 bpm and older children have heart rate around 90 bpm [2].This paper describes the system which is based on telemedicine which uses telecommunication and IT to provide efficient medical facility to the patients. This application is very useful for the medical staff which is not qualified enough to take critical decisions of the patients during emergencies in the physical absence of the doctor. In this application the medical information of the critical patient is fed into the server and this information is automatically synchronized into the Database. The information which is fed into the database is acquired by the doctor who is on the move and can monitor the status of the patient. The heart rate of the critical patient are fed into the database. In case of an emergency these reports of the patients who require an immediate medical attention are provided to the doctor. All this information about the critical patient is accessed by the doctor using an android application on his/her smartphone and take necessary steps.Previously patients suffering from critical heart illness had to go to their respective doctors at their clinic or hospitals for their routine checkup for their corresponding treatment.

If the doctors were not available in the hospitals then the patients were not able to get their appropriate treatment. Due to this the death rate had increased by approximately 25%.Therefore inorder to provide better medical healthcare facilities to the patients, we have built the project based on telemedicine which allows the medical staff to provide the patients critical parameters to the doctor on his android enabled smart phone. Following problems faces by peoples in daily life ,People find it difficult to seek medical help at the time of emergencies, especially for the patients who are located in remote places.The manual procedure to seek help from the doctors present in other hospitals is time consuming.Due to lack of medical facility in remote areas it becomes difficult to treat the critical patients without the guidance of the doctor.Thus we have developed this application to provide an immediate response to the patients at the time of emergencies.This application is developed so that the patient is able to seek treatment even in the physical absence of the doctor and irrespective of the patients location. In our Proposed scheme is founded on Heart beat device(HBS). Sensors are linked to the hardware and attached to the patient's body. Using sensors values of heart rate sensed and these values are sent to the android mobile phone. Android application collects heart rate and these values are analysed for guessing heart attack.

II. RELATED WORK

In 2014, C. K. Das, M. W Alam and M.I.Hoque [3] describes the development of a wireless heartbeat and temperature monitoring system based on a microcontroller at a reasonable cost with great effect. Most monitoring systems that are in use in today's world works in offline mode but it is of great need that a system must be designed so that patient can be monitored remotely in real time. They have used sensors which measures heartbeat and body temperature of a patient which is controlled by the microcontroller. Both the readings are displayed in LCD monitor. Wireless system is used to transmit the measured data to a remote location. The heartbeat sensor counts the heartbeat for specific interval of time and estimates Beats per Minute while the temperature sensor measures the temperature and both the data are sent to the microcontroller for transmission to receiving end. Finally, the data are displayed in the LCD at the receiving end.

In 2016, Sharanabasappa Sali1, Pooja Durge, Monika Pokar and Namrata kasge [4] designed to measure heart beat (pulse count), by using embedded technology. ALSO simultaneously it CAN measure and monitor the patient's condition. IT describes the design of a simple, low-cost controller based wireless patient monitoring system. Heart rate of the patient is measured from the thumb finger using IRD (INFRA-RED Device sensor).Pulse counting sensor is arranged to check whether the heart rate is normal or not. So that a SMS is sent to the mobile number using

GSM module interfaced to the controller in case of abnormal condition. A buzzer alert is also given. The heart rate can be measured by monitoring one's pulse using specialized medical devices such as an electrocardiograph (ECG), portable device e.g. The patient heart beat monitoring systems is one of the major wrist strap watch, or any other commercial heart rate monitors which normally consisting of a chest strap with electrodes. Despite of its accuracy, somehow it is costly, involve many clinical settings and patient must be attended by medical experts for continuous monitoring**.**

In 2014, Paola[9] Describes Importance Of Heart Monitoring not only in sports but in medical research and diagnosis purpose. Observation of heart function, accurate identification of cardiac states, and interference of heart sickness are essential goals. The electrocardiogram (ECG) is the gilt standardized method to find cardiovascular normality. In 2017,Maria Islam[10] described the design of an integrated portable device that can monitor heart rate (HR) continuously and send notifications through short message service (SMS) over the cellular network using Android application is presented.In situations where there is an absence of doctor or clinic nearby (e.g., rural area) and where the patient cannot realize their actual poor heart condition, is where our implemented system is of paramount importance.

In 2017,Satavant Kumar[11] Described the heart rate monitor, body temperature measurement, displaying and data-collecting at android mobile device. Heart rate, heart rate variability and temperature of the body are considered as vital signs for health of the human beings, which are crucial parameters that may be measured regularly every time at the time of illness. It makes heart rate a significant property of cardiovascular system.

Liangliang CHENG[12] Described The incidence of cardio-cerebrovascular disease is high, if the abnormal heart rate can be detected in time, the patients can be quickly rescued at the time of onset. Based on Bluetooth Low Energy (BLE) technology, this paper develops a portable, low-power heart rate meter. The heart rate monitoring system contains two parts: the heart rate data collection end and the mobile phone App. After the heart rate is acquired by the sensor, processed through the hardware filter and software filter, it is sent to the Android mobile phone by Bluetooth 4.0 wireless transmission technology

III. PROPOSED SYSTEM

If the doctors were not available in the hospitals then the patients were not able to get their appropriate treatment. Due to this the death rate had increased by approximately 25%. Therefore, in order to provide better medical healthcare facilities to the patients, we have developed this application to provide an immediate response to the patients at the time of emergencies.

The proposed system is based on Heartbeat sensor (HBS). Sensors are connected to the hardware and attached to the patient's body. Using sensor values of heart rate sensed and these values are sent to the android mobile phone. The Android application takes the values and these values are analysed for predicting heart attack. This application is developed so that the patient is able to seek treatment even in the physical absence of the doctor and irrespective of the patient's location.

In Our system,The android application is connected to the hardware via Web service,data is received from hardware,Data received displayed on application i.e. heart rate. This data is analysed for predicting heart disease.For guess of heart illness threshold values for heart rate is set. When the heart rate will be lower or overhead the threshold value, communication is sent to doctor's recorded number. A message is sent which contains patient's heart rate count with the time and date when it is measured.

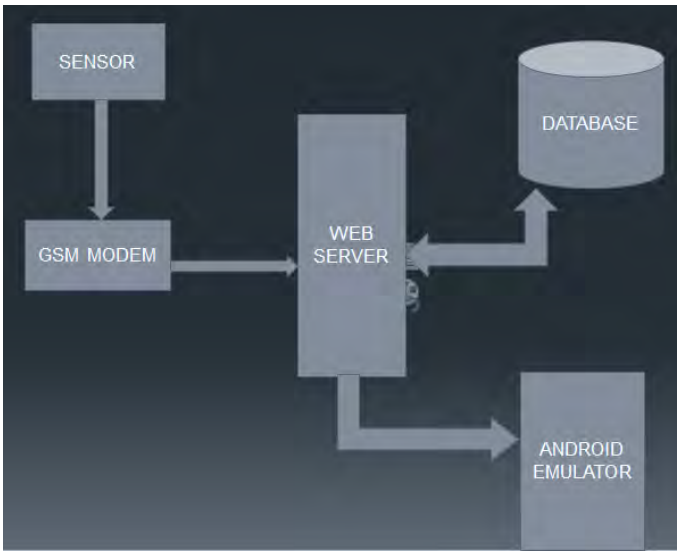


Fig 1: Block Diagram

In the proposed system, first the doctor has to register into the application. After getting the registered user have to enter the user id and password into the Android application which is installed on the mobile device and submit it to the server if it's valid then he can see the heart rate of the patient and then it will send to the doctor to the mobile device using SMS. Then after analysis of data accordingly feedback prescription will be sent to the user.

Data Description:

dbo.Admin_Login

Column name	Data type
Admin_id	Int
Username	Varchar(50)
Password	Varchar(50)

Dbo.Threshold_Master:

Column name	Data type
id	Int
MinValue	Varchar(50)
MaxValue	Varchar(50)

HeartRateMonitor.Readings:

Column name	Data type
id	Int
Reading	Varchar(100)
DateAdded	Datetime

IV.EXPERIMENTAL RESULTS

The system implementation is divided into following phases:

1. The android application is connected to the hardware via Web services.

First the doctor will visit the application then to login he have to enter the user id and password on mobile device, if its valid then the unique then he can move further or else he has to register himself and then stored in database.

After getting the registered user ,they have to enter the user id and password into the Android application which is installed on the mobile device and submit it to the server if its valid then he can see the heart rate of the patient and then it will send to the doctor to the mobile device using SMS. Then after analysis of data accordingly feedback prescription will be send to the user.

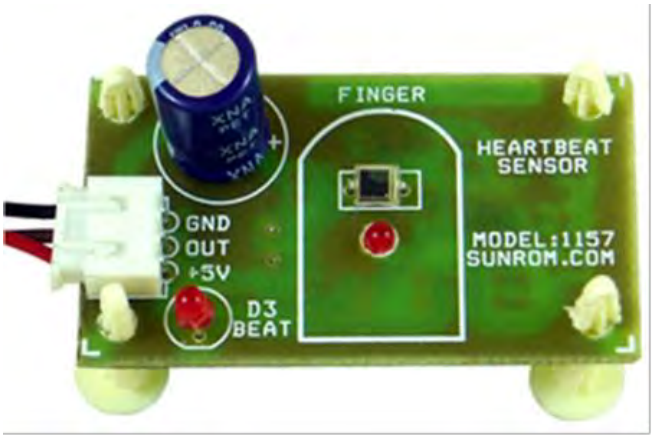


Fig 2: Heartbeat sensor

2. Data received from the hardware is displayed on the application which is shown in fig. 3 i.e. heart rate pulse. This data is analyzed for predicting heart attack.

For prediction of heart attack threshold values for heart rate is set. When the heart rate will be below or upper the threshold value Notification is sent to the patient's relatives, doctor's and hospital's registered number. The Notification contains patient's heart attack parameters and location. Location is tracked via GPS. The application contains the IP address of server to which we want to send the data for graph plotting. The server comprises database also doctor can checked chart of heart rate for their suitability. The server also contains a database for telemedicine part.

Product perspective ,In this we investigates the feasibility of heart rate (HR) & breath rate(BR) using a cell phone camera without using external sources. This works by placing index finger over camera thus phone acquires PPG signal & from that HR is estimated using peak detection algorithm. Each user and doctor will be having id which can be used for the login and for entering details etc. whenever user or doctor login to system, authentication process has done by server. If doctor is not registered then register they can enter detail and register. The details store in database. In case of retrieval of records and ordering made, much of human intervention can be eliminated.

By using Android Based Heart Rate Monitoring System , we are able to provide the patients from remote location to seek medical help. The patient is able to measure his heart rate with the help of hardware , if the patients heart rate goes below or above the threshold value set by the doctor, the doctor will get a notification and then provide appropriate feedback to the medical staff.

The software provides good graphical interface for the user any Operator can operate on the smartphone, performing the required task such as sending, receiving, update, viewing the data information.

- User can send or received data from other devices
- Heart rate analysis and search for doctor facility based on different criteria.
- Provide Security

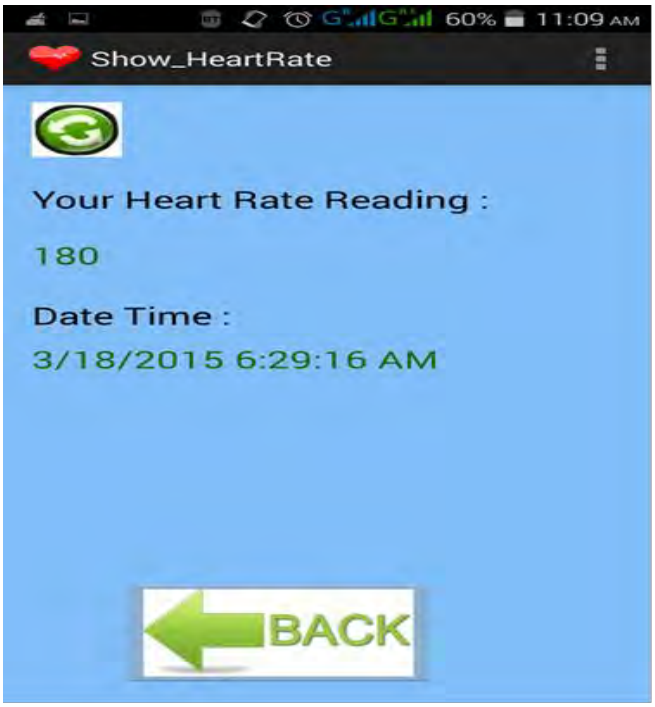


Fig 3: Hear rate reading

The constrain of our research is each user and doctor will be having id which can be used for the login and for entering details etc. whenever user or doctor login to system, authentication process has done by server.



Fig 4: Setting threshold

If doctor is not registered then register they can enter detail and register. The details store in database. In case of retrieval of

records and ordering made, much of human intervention can be eliminated.

IV. CONCLUSION AND FUTURE WORK

Proposed system 'Android Based Heart Monitoring System' is developed with the help of the Android Open Source Platform. The system is very usable for the patient as the patient can go wherever he wants along with the system. Also the parents and doctors will ensure the patient's safeties as if any problem occurred, the system will immediately inform them. Also, it is useful for the doctor as he can check each detail of patient whenever necessary. As the system is movable there is no longer need to stay in hospitals for patient. So he will be in real-life. We are developed model of this application using the constant monitoring of parameters to confirm and expect the heart attack and generate Notification alerts. Because of the availability of portable monitoring devices this application can be used in very great extent. The further addition to this system, would include collecting other critical parameters of the patient, such as ECG, Temperature, Blood pressure, etc. Also the location of the patient will be traced and accordingly a list of nearest hospitals, medical centres will be provided. Also the application will be made platform independent. In future we will use different prediction algorithm .

V. REFERENCES

[1]. http://www.healthline.com/galecontent/physical-examination. (August 31, 2013).

[2]. D. Mukherjee, K. Gupta, M. Pandey, and A. Agrawal, "Microcontroller Based Cardiac Counter System", IJEAM, Vol. 02, Issue 01, April 2013 ISSN (Online):2320-6608.

[3]. C. K. Das, M. W. Alam and M. I. Hoque "Wireless Heartbeat and Temperature Monitoring System for Remote Patients," Proceedings of the International Conference on Mechanical Engineering and Renewable Energy (ICMERE2013) 1-3 May 2014, Chittagong, Bangladesh.

[4]. Sharanabasappa Sali, Pooja Durge, Monika Pokar, Namrata Kasge, "Microcontroller Based Heart Rate Monitor", in International Journal of Science and Research (IJSR), May 2016.

[5]. R. Trivedi, G. Mathur, A. Mathur, "A Survey on Platinum Temperature Sensor", IJSCE, ISSN: 2231-2307, Volume-1, Issue-NCAI2011, June 2011

[6]. J. Prasath, "Wireless monitoring of Heart Rate using Microcontroller," International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE), vol. 2, pp. pp: 214-219, 2013.

[7]. Ibrahim and K. Buruncuk, "Heart Rate Measurement from the Finger Using a LowCost Microcontroller," Near East University, Faculty Of Engineering, TRN, 2014.

[8]. Online:
[9]. https://paarl.wikispaces.com/file/view/Standards+and+Procedure+in+Digitization+and+Digital+Preservation.pdf

[10]. Paola Pierleoni, Luca Pernini, Alberto Belli, and Lorenzo Palma, "An Android-Based Heart Monitoring System for the Elderly and for Patients with Heart Disease," International Journal of Telemedicine and Applications, vol. 2014,

[10]M. Islam *et al*., "Android based heart rate monitoring and automatic notification system," *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, Dhaka, 2017, pp. 436-439.

[11]Satavant Kumar Singh Arya, Chakshu Bhatia, Akhilesh Tiwari, et al. Android Based Heart Rate Variability and Body Temperature Monitoring System. *Research & Reviews Journal of Medical Science and Technology.* 2017; 6(3): 11–15p.

[12]Liangliang CHENG, Chenghua FAN, Yuanyuan ZHOU, Bingbing CHEN,"Design of a Wireless ECG Monitoring System using Android and Bluetooth Low Energy Technology" ISSN: 1473-804x online, 1473-8031 print

Automated PCB Disassembly System

Sneha L.Gaikwad, Pragati K.Bachhav, Shraddha B.Gangurde, Tejaswini S.Deshmukh
Electronics & Telecommunication, NDMVPS'S, KBTCOE Nashik
gaikwadsneha008@gmail.com, pragatib2411@gmail.com, shraddha6126@gmail.com, deshmukh.tejaswin@kbtcoe.org

***Abstract*— We are living in an era of electronic revolution. Our daily lives can't be imagined without various electronic devices and gadgets. With the ever fast changing technology, old gadget gets replaced with new and more advanced ones. We all move with the revolution and discard and replace our old electronic gadgets and devices but lesser we think of what will happen of the E-Waste generated and how it will be managed. Some of the common E- wastes include: home appliances such as televisions, air conditioners, electric cookers and fans, DVDs, Radios and microwaves among others; information technology equipment's such as computers, mobile phones, laptops, circuit boards, hard disks, and monitors among others; and other electronic utilities such as leisure, lighting, and sporting equipment's. In many developing countries improper E-Waste management using informal methods leads to major environmental problems. It affects the living lives also. A major part of this E-Waste is discarded Printed Circuit Boards. In E-Waste recycling process initial few steps are manual and are time consuming. In this proposed system we have automated one of the manual methods: We have proposed an automated system to dis-assemble components from PCB and replace the traditional manual method.**

Keywords—E-waste management, discarded PCB management, wastes printed circuit boards (WPCB)

I. INTRODUCTION

The e-waste recycling process is highly labour intensive and requires several steps. Various steps are included in the recycling process like picking shed, disassembly, first size reduction process, second size reduction process, over-band magnet, non-metallic and metallic components separation, water separation [2].

Picking shed involves manual removal of items that are hazardous like batteries. Disassembly step involves a serious labour intensive process of manual dismantling. The disassembled components are sorted into different categories. In first category size reduction process shredding of PCB is done to pieces less than 2 inches in diameter. Second category size reduction process consists of further crushing of shredded pieces of previous step. Over-band magnet involves removal of all the magnetic materials. The next step is the separation of metals and non-metallic components. Metals like copper, aluminium, and brass are separated from the debris to only leave behind non-metallic materials. The metals are either sold as raw materials or re-used for fresh manufacture. As the last step, plastic material is separated from glass by use of water. Once separated, all the materials retrieved can then be resold as raw materials for re-use. The products retrived include plastic, glass, copper, iron, steel, shredded circuit boards, and valuable metal mix [2]. The second step in above process is disassembly. In It is done traditionally by manual method. Our proposed system will help in making this step more efficient by automation.

II. MOTIVATION

Increasing levels of electronic waste and its unsafe and improper treatment and disposal through open burning or in dumpsites, pose significant risks to the human health and environment. They also present several challenges to sustainable development, as well as to the achievement of the Sustainable Development Goals (SDGs). By 2017, close to half the world's population uses the internet and most people in the world have access to internet networks and services. Annually generation of E-waste has grown to 44.7 Million Metric Tonnes which is Equivalent to Almost 4,500 Eiffel Towers. All the countries in the world together generated a staggering 44.7 million metric tonnes (Mt), or an equivalent of 6.1 kilogram per inhabitant (kg/inh), of e-waste per year in 2016, compared to the 5.8 kg/inh generated in 2014[1]. This is close to 4,500 Eiffel Towers each year [1].

III. OBJECTIVES

- Effective method to assist E-waste management.
- Minimize the pollution.
- Reduce human efforts.
- Safety for the operator.

IV. ANALYSIS AND DESIGN APPROACH

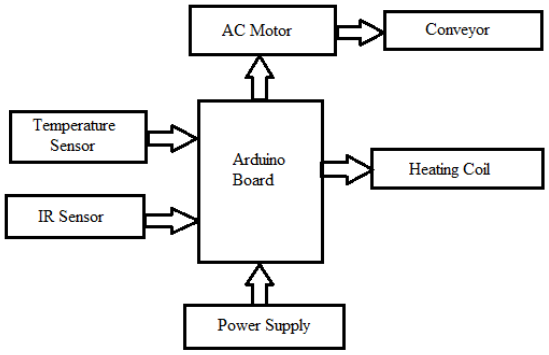


Fig.1 Block Diagram of system

Block Diagram Explanation:

- Power Supply: This block is regulated 9 V power supply. It is used to give regulated 9 V power to Arduino.
- Temperature sensor: This block is to sense the temperature of heating chamber.PT100 is used as the temperature sensor. The required process temperature is 250 degree Celsius. This block is essential to keep check on the temperature of heating chamber.
- IR Sensor: This is used to sense the presence of PCB on conveyor. Sensor used is IR sensor.
- Motor: Motor is used to drive the conveyor chain. According to the inputs feed by sensors to the arduino, operation of motor is controlled.

- Heating coil: Heating coil is used to heat the heating chamber. Turning ON and OFF of motor is also controlled by arduino based on the inputs it gets from temperature sensor block.
- Arduino: Arduino block is the controlling block. It is feed with inputs from sensors and controls the motor and heating coils. Sensors like IR sensor and PT100 are interfaced to the arduino. IR sensor senses the presence of PCB on the conveyor belt and feeds it to arduino. If PCB is present on conveyor IR sensors senses it and turns the heating coil ON. Once the required temperature is reached, temperature is sensed by PT100 and arduino turns the motor ON. On turning motor ON the conveyor belt starts rotating. If there is no PCB on conveyor, IR sensor sense it and motor is turned OFF. Whenever the temperature goes below the required temperature, PT100 senses it and arduino turns the motor OFF.

A. Arduino Uno Board

The Arduino Uno is used for control mechanism .The board is equipped with sets of analog and digital (I/O) pins that may be interfaced to various expansion boards (shields) and other circuits.

The ATmega328 on the Arduino Uno comes pre-programmed with a boot loader that allows uploading new code to it without the need of using an external hardware programmer. It uses the original STK500 protocol for communication. The Uno does not use the FTDI USB-to-serial driver chip unlike its all preceding boards.

Features:

1] Microcontroller: Microchip ATmega328P
2] Operating Voltage: 5 Volt
3] Input Voltage: 7 to 20 Volts
4] Digital I/O Pins: 14 (6 provide PWM output)
5] DC Current per I/O Pin: 20 mA
6] DC Current for 3.3V Pin: 50 mA
7] Flash Memory: 32 KB in which 0.5 KB used by bootloader
8] SRAM: 2 KB
9] EEPROM: 1 KB
10] Clock Speed: 16 MHz

B. Temperature Sensor

Platinum resistance thermometers (PRTs) offer accuracy a wide temperature range of 200 to +850 degree C. The principle of operation is for PT100 to measure the resistance of a platinum element. The most common type (PT100) has a resistance of 100 ohms at 0 degree C and 138.4 ohms at 100 degree C. The relationship between resistance and temperature is approximately linear over a small temperature range.

For a PT100 sensor, a 1 degreeC temperature change will cause a change in resistance by 0.384 ohm, hence even a small error in measurement of the resistance can cause a large error in measurement of the temperature.

For precision work, sensors have four wires - two to carry the sense current, and another two to measure the voltage across the sensor element [3].

C. IR sensor

Proximity Sensor is used to detect objects in front of sensor. Sensor consists of a transmitter and receiver. Transmitter keeps transmitting infrared light and when any object comes near, it is detected by the sensor by monitoring the reflected light from the object on the receiver.

Features:

- Ambient light protected IR receiver.
- 3 pin easy interface connectors.
- Power LED Indicator LED.
- Operate up to distance from 2cm to 30cm.
- Can differentiate between light and dark colours.
- Active Low on object detection.
- 3.3 to 5V operation.

D. Nichrome Wire

Here Nichrome wire is used because of its high heating property as compared to other good conductors (metal wires). Nichrome wire is a heating element that allows the electricity to pass through it and offer certain resistance that produces heat. The amount of heat produced depends on the magnitude of current passing through the heating element and the magnitude of resistance offered by the element wire.

Most heating elements use nichrome (80% nickel-20% chromium) ribbon, wire, or strip. Nichrome 80/20 is an ideal material, because it has high resistance and forms an adherent layer of chromium oxide when it is heated for the first time. Material beneath this layer will not oxidize, preventing the wire from burning out or breaking [4].

Features:

1] High melting point
2] High resistance
3] Chemical inertness
4] Non-magnetic
5] Good mechanical properties.

E. Relay

Relay is used for switching. It has one common terminal and two contacts. Contacts are in two different configurations: one can be Normally Closed and other one is Normally Open. So basically we can see the SPDT relay as a way of switching between 2 circuits: when there is no voltage applied to the coil one circuit receives current, the other one doesn't and when the coil gets energized the opposite is happening.

Terminal Descriptions:

1] NO: This is Normally Open switch. This is the terminal where we can connect the device that you want the relay to power when the relay is powered, meaning when the COIL receives sufficient voltage. The device connected to NO when the relay has no power and will turn on when the relay receives power.

2] NC: This is the Normally Closed Switch. This is the terminal used to connect the device that you want powered when the relay receives no power. The device connected to NC will be on when the relay has no power and will turn o when the relay receives power.

3] COM: This is the common of the relay. If the relay is powered and the switch is closed then ,COM and NO have continuity. If the relay doesn't powered and the switch is open, COM and NC have continuity. This is the terminal of the relay where you connect the rest part of your circuit.

F. Transistor

Bipolar junction transistor is used as switch. A bipolar junction transistor (BJT) can be used in many circuit configurations such as an amplifier, filter, oscillator and rectifier or just used as an on-off switch. If the transistor is biased into the linear region then, it will operate as an

amplifier or linear circuit. If biased alternately in the saturation and cut-off regions, then it is being used as a switch, it either allows current to flow or not to flow in other parts of the circuit.

G. Motor

An AC motor is an electric motor which is driven by an alternating current (AC). The AC motor commonly consists of two basic parts, an outer part known as stator having coils supplied with alternating current to produce a rotating magnetic field, and an inner rotor attached to the output shaft producing a second rotating magnetic field. Motor is used to drive the conveyor on which printed circuit board is present. This printed circuit board is passes through heating chamber using conveyor. Motor is continuously ON. If for the long delay printed circuit board is not detected motor will turn OFF.

Features:

1] 230V power supply

2] 0.5HP Motor

3] Speed of motor is 60 rpm

Proposed Model:

Initially Coil is turned on. Proximity sensor checks for presence of PCB on conveyor. If PCB board is sensed on conveyor and the required temperature is attained in the chamber, motor turns ON. PCB passes through heat chamber, where the solder metal melts. Once the solder metal melts, PCB passes over the metal brush and components fall off the PCB.

Flowchart:

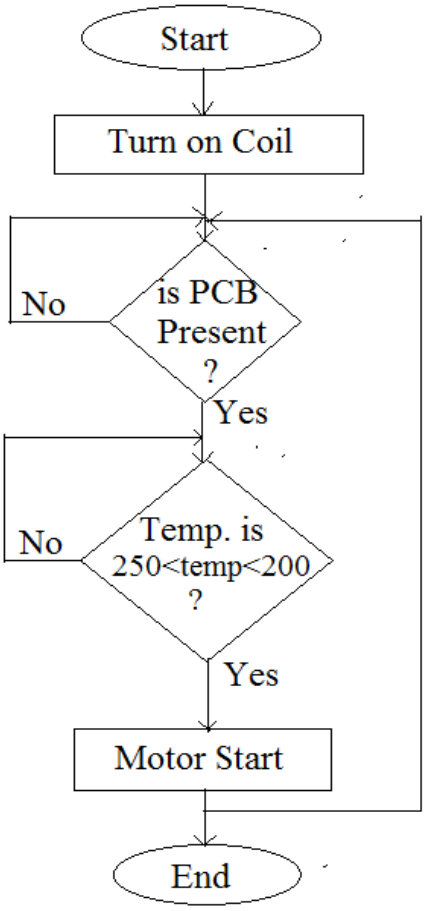


Fig.2 Flow chart for system

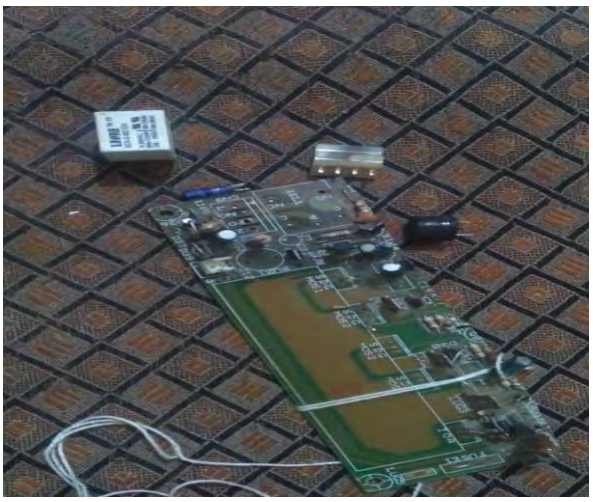


Fig.3 Components separated from board

CONCLUSION

In this project, we have made a system which will reduce manual work and replace it with automated system. It will be more efficient than the traditional manual method used.
This system will simplify initial steps of E-waste recycling process. It can be integrated and used with current E-waste recycling process assembly.

REFERENCES

[1] www.itu.int
[2] www.conserve-energy-future.com
[3] www.picotech.com
[4] Wikipedia.org
[5] Shubham Gupta , Gaurav Modi, Rahul Saini , Vijaya Agarwala, " A review on various electronic waste recycling techniques and hazards due to its improper handling",2014.
[6] Xianlai Zeng et al., *Current status and future perspective of waste printed circuit boards recycling,*2012.
[7] S.Deno, D. Landis,P. Hulina and S.Balasubramanian,"A Rapid Prototyping Methodology for Reverse Engineering of Legacy Electronic System, IEEE International Workshop on Rapid System Prototyping", 1999.

Implementation of Sentimental Analysis of Social Media using Naïve Bayes

Kshtiij Poojary, Prashant Sharma, Varun Toshniwal, Mrs.Lydia Suganya Mrs.Lydia Suganya

Department of Computer Engineering, Thakur College of Engineering & Technology, Mumbai, India
kshitijpoojary@gmail.com, prashant.dsv@gmail.com, varuntoshniwal97@gmail.com

***Abstract*--Sentiments, evaluations, attitudes, and emotions are the subjects of study of sentiment analysis and opinion mining. The growth of this field coincides with the widespread use of social media and other applications on web, e.g. review forums, blogs, twitter, facebook and many more. The data from these applications can be collected for research purposes, as they are available in huge volume. Since early 2000, sentiment analysis has grown to be one of the most active research areas in natural language processing. It is also widely studied in data mining, Web mining, and text mining. Truth be told, it has spread from software engineering to the executives sciences and sociologies because of its significance to business and society all in all. In recent years, industrial activities surrounding sentiment analysis have also thrived. We proposed a sentiment analysis system using modified means and Naïve Bayes algorithm.**

I. INTRODUCTION

A. Background

Humans have an ability to see, adapt, learn and judge. Humans like to give their opinion over everything weather it is positive or negative. They give their opinions over people, places, products and events. These types of opinions are consider as 'Sentiments'. Sentiment analysis is the study of automated techniques for extracting sentiments from written languages. With digitalization at its peak, social media websites have contributed tremendously as a source of data and this growth has resulted in an explosion of publicly available and user generated text data on the internet. As a platform for businesses to expand and get better. Twitter (microblogging website) has evolved over the years, with over 600 million users and over 500 million tweets are sent each day. Tweets which includes opinions, reviews of products and discussion on current affairs. These data and information can potentially be utilized to provide real-time insights into the sentiments of people. Even these information could be used by companies, where they could start a poll and get an overall response and review about their products. Retrieving these "tweets", creating a report for the companies (positive/negative/neutral responses) and helping them improve their product is everything that our project is about.

B. Problem Identification

Retrieving data/tweets manually is time consuming and can be inaccurate. Handling thousands of tweets, classifying the tweets in positive/negative/neutral accurately, and generating reports accordingly could be tough and unmanageable at times. There are many tools like TweetFeel, Crawler etc. which is being used to find out the sentiments of the people. But its comparatively easy to find out the sentiments of famous people and celebrities by using these tools. We will use different approaches to find the sentiments and our problems could be solved by using algorithm like Naïve Bayes, and Natural language processing (NLP).

C. Title Justification and Purpose

Sentiment is a demeanor, thought, or judgment provoked by feeling. Sentiment analysis, which is otherwise called opinion mining, examines individual's feelings towards specific substances. Web is a creative place as for assumption data. From a Twitter user's perspective, people tweet about their own content, give their opinion on the ongoing affairs, or give individual's review about a particular product. From a developer's perspective, we could use Twitter's APIs for data collection and analysis. As of now Twitter has 3 APIs, namely the REST API, which is used for data collection, the Search API, which could be used to gather specific content, and the Streaming API which is used to live stream data. This helps us understand the working of APIs, and how publicly available data could be stored, used and analyzed. These APIs could be merged for creating a particular application. One of the other reason for developing this application is exploring one of the most important sectors of computer science, Natural Language Processing. Learning about that permits investigating more possibilities in the data mining field, specifically in the developing nations where openings are tremendous and simple to catch.

II. OBJECTIVES & OUTCOME

The main ideology of developing this project will be retrieving data/tweets about a particular event (e.g. a political event) or about any newly launched product. Later analyzing these tweets and identifying the polarity of the tweet (i.e. positive, negative, neutral or maybe irrelevant). Once the analysis part is done, the results will be displayed in graphical format (a pie chart/bar graph) making the intended user understand better how the polarities change over a period of time. Our goal is to achieve the most accurate results, and hosting a user-friendly website where clients be able to get their desired results.

III. RESEARCH/TECHNICAL/PROJECT WORK

A. Review

Our everyday life has dependably been affected by what individuals think. Thoughts and conclusions of others have constantly influenced our very own suppositions. The progressions in web has prompted expanded action in

podcasting, blogging, labeling, adding to RSS, social bookmarking, and informal communication. Accordingly, there has been an ejection of enthusiasm for individuals to mine these huge assets of information for sentiments. Assessment investigation is the computational treatment of feelings, conclusions and subjectivity of content. We will likewise examine some particular themes in assessment investigation and the contemporary works in those zones.

IV. METHODS AND TOOLS

A. *Naïve Bayes Classifier*

The Bayesian classification is used as a probabilistic learning method. In naive Bayes classifiers, every feature gets a say in determining which label should be assigned to a given input value. To choose a label for an input value, the naive Bayes classifier begins by calculating the prior probability of each label, which is determined by checking frequency of each label in the training set. The contribution from each feature is then combined with this prior probability, to arrive at a likelihood estimate for each label. The label whose likelihood estimate is the highest is then assigned to the input value.

Following equations are used in calculating label likelihoods
P(label|features) = P(features, label) / P(features)
= P(label) × P(features|label)) / P(features)
= P(label) × Prodf in| featuresP(f|label)` / P(features)
(Since features are independent)

In this classification method, the naïve Bayes assumption or independence assumption which is independence of the features makes it easy to combine the contributions of different features since it need not worry about how they should interact with one another.

B. *Apache Spark*

It's an open source cluster computing platform to retrieve data. Apache spark is an in-memory fast processing system used for large scale data processing. It has come up as an advanced version of Hadoop. It has the capability to process data 100x times faster by memory partioning and 10x times faster on disk across different nodes. Apache Spark structure is based on RDD which is read only. It overcomes the limitation of MapReduce in which data after reducing was stored into a disk by implementing iterative algorithms who fetch data from multiple datasets in a loop thereby implementing repeated database-style querying of data. RDD is basically an abstraction feature which before data processing lays down the execution plan and then later depicts computation using Direct Acyclic Graph(DAG).The generated DAG acts as a framework to carry out the pattern analysis and processing and task segregation. Further, it has a better edge over other technologies as it is quite easy to implement due to multiple available APIs. Also, the other benefits include high level libraries.
Scala: Scala supports both High Level Functionality and Object Oriented Programming language model. This provides scala an edge over Java. You can write short codes in scala as compare to java. The major importance of Scala is its inbuilt Apache Spark feature. Scala supports all the libraries of Apache Spark.

C. *Twitter*

Twitter is a social networking platform where users can post their views which can be reached to lacs of people in a matter of seconds. Twitter have many benefits such as the amount of data on twitter is much larger as compare to blogs or review websites. Responses for a given topic is very huge on twitter and they are both positive and negative. Facebook does not provide much data in its API that's we didn't used it.

D. *IntelliJ Idea*

IntelliJ Idea is the IDE which we used to build, run and test our code. The community edition of Intellij Idea is free of cost and it supports all the SBT plugin which is required to import Apache Spark dependencies and build the project.

E. *Google Charts/Excel Chart*

Google Charts is a pure JavaScript based charting library meant to enhance web applications by adding interactive charting capability. Because of the advantages of the Google charts we use it to represent the output in the forms of bar charts and pie charts. They are dynamic and easy to make and change the data with a simple change in JSON array from the background.



V. RESULT AND DISCUSSION

A. *Observations*

In the wake of confronting various limitations or problems, effective disposal of those mistakes we have finished our task with persistent exertion. Toward the finish of the task the outcomes can be combined as:

- An easy to use online application.
- No aptitude is required for utilizing the application.
- Organizations can utilize the application to imagine item or brand survey.
- Google chart tools are used to display the sentiments in graphical representation.
- Google chart tools provide a better way to host any data. We used the pie and bar chart to represent complex output.

Naïve Bayes method have some limitations which can be resolve for the better output. Some of the findings which can be use for the improvement are as follows:

- If the output is not normally distributed then we can use transformation and different methods to convert it in normal distribution.
- If test data set has zero frequency issue, apply smoothing techniques “Laplace Correction” to predict the class of test data set.
- Remove correlated features, as the highly correlated features are voted twice in the model and it can lead to over inflating importance.

B. *Validation and Verification*

The testing procedure is a piece of more extensive subject alluding to check and approval. We need to recognize the framework determinations and attempt to meet the client's prerequisites and for this sole reason, we need to confirm and approve the item to ensure everything is set up. Confirmation and approval are two unique things. One is performed to guarantee that the product accurately executes a particular usefulness and other is done to guarantee if the client prerequisites are appropriately met or not by the finished result. Confirmation is progressively similar to 'would we say we are building the item right?' and approval is increasingly similar to 'would we say we are building the correct item?'.

VI. FINDINGS AND FUTURE DEVELOPMENT

Despite the fact that we have made a stage towards the examination and expectation of Naive Bayes execution, there is still considerably more work that needs more and more profound research.

(1) The gathered example information is constrained.

(2) For the parallel Naive Bayes usage in this task, it is not exclusively can be connected to content characterization or estimation examination. It can likewise be reached out to different territories like picture handling, or quality examination and so forth.

(3) We have finished up the execution forecast model identified with the extent of input information and the quantity of hubs in a bunch, we likewise have a certain comprehension of the connection between the execution and the worldwide counters, and the equipment design. For the last mentioned, regardless we require more information and examinations to have more subtleties.

To come up with the better solutions we need Data Mining techniques which help us gather the proper and accurate data of the event so we can put a better approach to find out the desired output. In Data Mining when implementing the product profiling, it uses the decision tree after comparing it with the clustering technique and for the trend analysis and forecasting it used Holt Winters method which is capable of analyzing seasonal data and predict proper values for the future because the Sentiment Analysis is based not only on the negative or positive polarity of words and concepts, but also on the syntactical tree of the sentence being analyzed. The system tries to read between lines, identifying idiomatic or colloquial expressions, giving interpretation to negations, modifying polarity of words basing on the related adverbs, adjectives, conjunctions or verbs, taking in account specific functional-logic complements.

A. *Future Scope*

One of the aspects is to determine whether the attributes identified by us for Naïve Bayes algorithm is sufficient to get the accurate results. We incorporated these attributes for additional support for the last Naïve Bayes model, yet further experimentation could be possible to test whether the predictions exactness level is really improving. The objective is to decrease "noisy data" while safeguarding the subtleties of natural language. Alternately, there are other characteristics that we have not thought about, for example, the utilization of emoji's; further research should be possible to check whether such would be a helpful expansion. The data sets available for any particular topic is limited to get the most accurate results. The more the data, the accurate the prediction algorithm will be. Collecting data for more period of time will assure us better results and probably the most accurate once.

We have concluded the performance prediction model related to the size of input data and the number of nodes in a cluster, we also have a certain understanding of the relationship between the performance and the global counters, and the hardware configuration. For the latter, we still need more data and experiments to have more details, such as how many global counters we should use for a certain use case, or under what situation that upgrading the hardware is worth the improvements it brings.

VII. CASE STUDY

Our project involves the usage of Apache Spark to analyze real time tweets. The objective of our case study is to find the polarity of the words (in tweets) retrieved. Each step in the framework involves several sub-tasks.

1. Data collection: Data in the form of raw tweets is retrieved by using the Scala library "Twitter4j" which provides a package for real time twitter streaming API. The API requires us to register a developer account with Twitter and fill in parameters such as consumerKey, consumerSecret, accessTokenaccess, and TokenSecret. This API allows to get all random tweets or filter data by using keywords. Filters supports to retrieve tweets which match a specific criterion defined by the developer. We used this to retrieve tweets related to specific keywords which are taken as input from users. Initially, we set at least set an application name and mode. We execute the program in local mode instead of cluster. Then, input array of keywords is provided as an argument to Streaming Context "ssc" using "sc" where "sc" is spark context. For example, on inputting multiple keywords like, 'Canada', 'Trump', 'Toronto', the output we obtained from 15 seconds' window time was the live stream of tweets associated with these keywords. Only caveat of using filters is that famous keywords like "India" have more tweets compared to niche words like "Focusrite" which makes it difficult to get data for niche specific keywords.

2. Data Processing: Data processing involves Tokenization which is the process of splitting the tweets into individual words called tokens. Tokens can be split using whitespace or punctuation characters. It can be unigram or bigram depending on the classification model used. The bag-of-words model is one of the most extensively used model for classification. It is based on the fact of assuming text to be classified as a bag or collection of individual words with no link or interdependence. The simplest way to incorporate this model in our project is by using unigrams as features. It is just a collection of individual words in the text to be classified, so, we split each tweet using whitespace. For example, the tweet "Met aziz today !!" is split from each whitespace as follows. { Met Aziz !!" } The next step in data processing is normalization by conversion of tweet into lowercase. Tweets are normalized by converting it to lowercase which makes its comparison with an dictionary easier.

3. Data Filtering

A tweet acquired after data processing still has a portion of raw information in it which we may or may not find useful for our application. Thus, these tweets are further filtered by removing stop words, numbers and punctuations.

Stop words: For example, tweets contain stop words which holds no additional information. These words serve no purpose and this feature is implemented using a list stored in stopfile.dat. We then compare each word in a tweet with this list and delete the

words matching the stop list Removing non-alphabetical characters: Symbols such as analysis and are removed using pattern matching.

Stemming: It is the process of reducing derived words to their roots. Examples includes words like “fish” which has same root as fishing and fishes. The library to use stemming is Stanford NLP which also provides various algorithms such as porter stemming. In our case, we have not employed any stemming algorithm due to time constraints.

Conclusion

The undertaking of supposition examination, particularly in the space of microblogging, is still in the creating stage and a long way from complete. So we propose two or three thoughts which we feel merit investigating later on and may result in additionally made strides execution. At the present time we have worked with just the least difficult unigram models; we can enhance those models by including additional data like closeness of the word with a refutation word. We could determine a window before the word (a window could for model be of 2 or 3 words) under thought and the impact of invalidation might be fused into the model on the off chance that it exists in that window. The closer the invalidation word is to the unigram word whose earlier extremity is to be determined, the more it ought to influence the extremity. For instance, if the negation is directly alongside the word, it might just invert the extremity of that word and more remote the nullification is from the word the more limited uncertainties impact ought to be.

VIII. REFERENCES

[1] G. O. Young, “Synthetic structure of industrial plastics (Book style with paper title and editor),” in Plastics, 2nd vol. ed. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

[2] Apoorv Agarwal, Fadi Biadsy, and Kathleen Mckeown. 2009. Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009), pages 24–32, March.

[3] https://www.researchgate.net/publication/268817500_Sentiment_Analysis_for_Social_Media

[4] https://www.researchgate.net/publication/230758119_Sentiment_Analysis_on_Social_Media

[5] B. Smith, “An approach to graphs of linear forms (Unpublished work style),” unpublished.

[6] C. Fellbaum. 1998. Wordnet, an electronic lexical database. MIT Press.

Implementation of Online Product Recommendation System using Collaborative Filtering

Shivani Patel, Pooja Songadkar, Mr. Nitin Harane

Thakur College of Engineering and Technology University of Mumbai

ABSTRACT-A recommender system is an Information Retrieval [1] technology that improves access and dynamically recommends relevant items to users by considering their precisely mentioned preferences and objective behavior. Recommender system is one of the major techniques that deal with the problem of information overload by suggesting users with relevant and appropriate items. They basically direct users towards particularly those items which can meet their needs cutting down large database of Information in the process. This paper identifies the issues faced by the current systems and tries to solve them by using a hybrid system of content based and collaborative filtering technique to achieve better and more accurate performance.

I .INTRODUCTION

Along with the rapid development of the Internet, the number of the servers connected to Internet and the Webs show a rise in the exponential growth. The rapid development of the Internet presents a mass of abundant information to us at the same time, for example, there are tens of thousands movies on Netflix, millions of books on Amazon, more than 10 billion page collection on Del.icio.us, so much of information, not to mention find some interesting content in all that, it is impossible to give all of that information a once-over. The traditional search algorithm only presents the same ordered results to all of the users; it fails to provide different service to different users according to their different interests. Personalized recommendation has been thought as one of the most effective tools to resolve the problem of information overload. Radically, the recommendation problem is to substitute user to evaluate the products, which include books, movies, CD, web and so on, it is a process from known to unknown. [2].

Recommendation is the type of information filtering which involves prediction of rating and user preferences, which would help a user to buy items according to their needs and interest. The books suggestion at Amazon or Flipkart is the best example of a recommendation system.

Recommendation system uses various technologies to recommend products viz Content filtering and Collaborative filtering. Content based filtering recommends item which is based on user's profile, which the user has liked in the past. It serves as finding relations/correlation among items in a large database. It works on the condition where X and Y are two sets of items. It finds correlation between X and Y i.e. if we buy an item X, then it finds the probability to buy the item Y as well. Collaborative filtering is a method to analyze the user's behaviour by predicting the user's taste to that of similar to another user.

II. PROPOSED SYSTEM:-

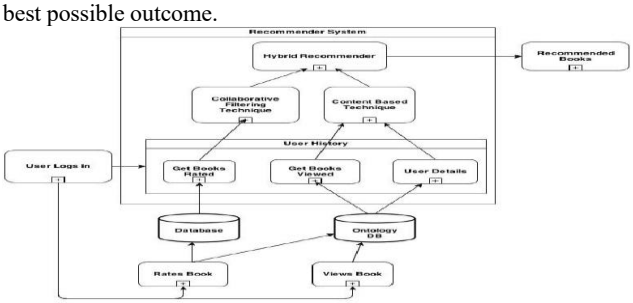
The proposed system uses a combination of collaborative filtering and content based recommendation system also known as personality based approach. It is a hybrid system and uses two methods in a broad light.

1. Collaborative filtering
2. Content based recommender

The difference between the two approaches can be highlighted by comparing - last.fm and Pandora, the two popular songs recommender system.
Last.fm – It uses collaborative filtering by creating a 'station' that recommend songs on the basis of what band or singer does a user listen to compares that to a similar user with similar interests and recommends the song that does not appear on UI of first user but is often played by second user. To be more specific this is an example of user based collaborative filtering.

2.1. Data for training:

There are two types of dataset we can use to train and then for prediction purpose namely implicit and explicit data. Explicit data requires effort from the users and they are seldom ready to take in any effort, it can be collected in the form of ratings, information provided in surveys etc. Implicit data is not collected on face value rather the model infers it on based on interaction of user with the system such as click through rate, watch time etc.
The asset of these two can be pooled to feed to a model and get best possible outcome.



Architecture

2.2. Model:

The hybrid model in this paper is a combination of content based filtering and item based collaborative filtering. CB uses principal component analysis for dimensionality reduction similar to UV decomposition for UV matrix. Content based filtering focuses on content of item rather than its similarity with other products. The content of each product is represented as set of descriptors in a vector.

The main ideology behind using a hybrid system is to use all the available data optimally. While using CF we do not consider its content related data such as actors likewise while using CB we don't consider correlation with other users. Since the paper is for developing recommendation engine for small businesses where data is sparse in the start we don't need to face problems of big organisation like amazon to process abundant data. But when the business grows the system should be robust enough to perform well in those circumstances hence can use clustering or categorization in early stage before using a learning model for recommendation thus dividing the users into several groups and then applying learning model for those groups.

Companies like amazon use multiple GPUs to meet time and space constraint [10] and data generation is often done on CPU along with analysis which is often snubbed. That's one of many ways for tackling big data. Even when the data is not massive enough to be called big data it may still be a huge load for a single processor to process in that we use alternating least square method which we are using in our paper.

Now in hybrid system we use data provided by user implicitly like CTR or explicitly by rating products and content of items which describes it and create a user profile. The weight of descriptors define how important that term is for recommendation. For a movie i for user u on the recommendation is [11]

$$p_{ui_{HF}} = w_{CF} * p_{ui_{CF}} + w_{CB} * p_{ui_{CB}}.$$

For example if the results for CB and CF are,

p_{CF}	Movie title
1	Terminator
2	The Apartment
3	Justice League

p_{CB}	Movie title
1	Justice League
2	The Prisoner
3	Spartan

Then recommendation can be calculated as,

p_{HF}	Movie title	Calculation
1	Justice League	$0.4 * 3 + 0.6 * 1 = 1.8$
2	Terminator	$0.4 * 1 + 0.6 * 4 = 2.8$
3	Spartan	$0.4 * 4 + 0.6 * 3 = 3.4$

III. LITERATURE SURVEY:-

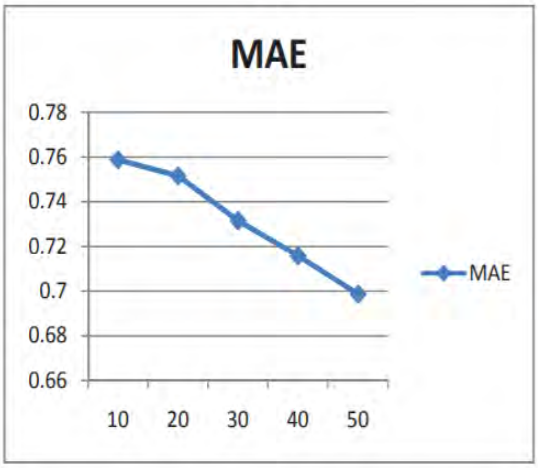
Abundant data [10] is available on the internet today and so in e-commerce. Sites like amazon, flipkart etc. have tons of data to maintain around and process them appropriately.

Sarwar and others [9] have showed the pragmatic results of different similarity algorithms and found out with the help of MAE graph that size of neighbourhood effects the prediction quality. p is the difference between the predicted ratings and actual ratings. High MAE indicates lower accuracy and vice versa for low MAE. Collaborative filtering [11] is the most widely used algorithm for recommending items.

Two types of collaborative filtering methods are proposed i.e. item based CF and user based CF out of which user based is preferred over other.

Content based filtering is used to fill the vacant ratings in the matrix. Hence the proposed amalgamation of collaborative filtering and content based recommendation system will help us to achieve the problem of data sparsity.

Hengsong Tan, et.al, have proposed a new technique to tackle data sparsity by combining item classification and item based CF. This approach classified the item through attributes and then produces prediction for items whose ratings are not available.



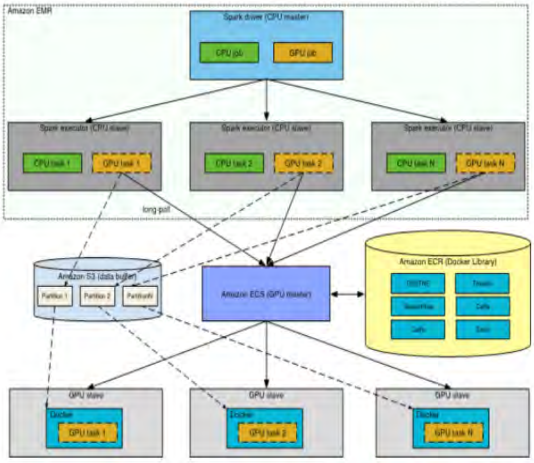
Mean Absolute Error

3.1. Amazon's Personalization System:

Amazon has a patent for item-to-item collaborative filtering. It is widely used by many giants for developing recommendations for their product. Amazon uses neural networks to generate personalized recommendations for its customer. With millions of users using their product they have distributed NN across multiple GPUs to meet space and time constraint [10]. Deep scalable sparse tensor neural engine is the model they have created and open sourced and is used for training and prediction.

Data generation and analysis the often overlooked area of training is processed on CPU. Now a hybrid of CPU and GPU is tough to manage since cluster managers do not inherently support GPU. They do have a solution where they run both CPU and GPU independently and drive end-to-end modelling process from Apache Spark.

Now amazon has millions of input and output nodes for which they easily reach trillions of weights for a fully connected network even if it is a shallow network and hence a single GPU is not suffice.



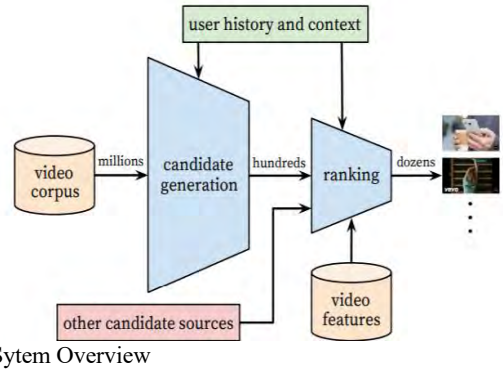
DSSTNE Architecture overview

They open sourced NN i.e. DDSTNE supports model parallel training in which the model is circulated across N-GPUs and the dataset is imitated to all nodes [10]. Each GPU then trains the model like in batches and shares the weight with rest of the nodes.

After training, prediction can be generated for each customer. This model can be scaled up by linearly adding more GPUs.

3.2. YouTube:

YouTube's system is built on Google brain open sourced as tensorflow [12]. This system consist two deep NN models, one for candidate generation and another for ranking.



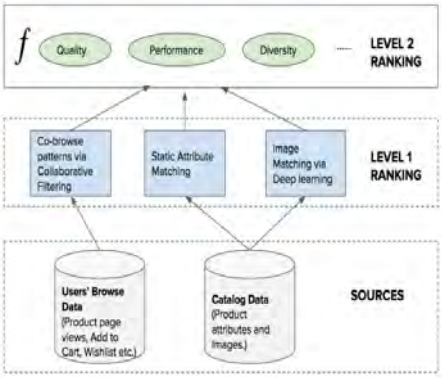
Sytem Overview

Former is used to select top 100 videos from an inventory of millions which is most relevant to a user. A high precision metric is considered for extracting these top videos. It only provides a wide perspective via collaborative filtering. They treat candidate generation model as a multiclass classification problem of a particular video.

Latter is a bit more personalized for a user and is required to have a more detailed level representation to discriminate among relative importance with high recall. It uses a similar deep neural network architecture to assign a score to each of those 100 videos using logistic regression. Eventual ranking of videos is persistently adjusted to results of live A/B testing and is usually a function of watch time per impression.

3.3 Flipkart:

Flipkart uses a hybrid of content based and collaborative filtering for recommending similar products. Users' watch history is used for training and predicting in collaborative filtering to find the most recurrent searched or looked product and product details and image catalogues is used in content based. Ranking of products is obtained by eliciting from these sources.



Flipkart's similar product recommendation ML model overview

System architecture of Flipkart's similar product recommendation consist of two levels.
Level 1 consumes of hundreds of millions of data of all users and products and is optimized to derive correlation between items and relative similarity.
Level 2 ranking [13] is optimized for performance and it uses point-wise logistic regression that can predict CTR and product to purchase conversion rate and rank them as per objective function.
This model is used by flipkart to recommend in batch and in real time too.

IV. FEASIBILITY ANALYSIS:-

Economic Feasibility: Our application will achieve accurate analysis, faster provision of recommendations, which will reduce human efforts, which in turn will reduce cost.
Technical Feasibility: Aim to achieve accurate recommendation in various fields using collaborative filtering and content based in python language.

Operational Feasibility: Data on which recommendation system is based is provided by the user itself and the system will analyze the pattern between the various items in the system.

V. APPLICATIONS

1. Movies Recommendation: Movies of various genres, languages, ratings etc will be recommended to the user.
2. News, Video searches: When a user follows one particular type of news or videos, the further recommendation list will show similar type of news and videos.
3. Music apps: The app will compile the songs played by the user and create a playlist according to the user's listening habits.
4. Online shopping (Clothes, electronics etc): Depending on the previous purchases of a user, the system will recommend products of various brands, price, colors etc. A new user will have

VI. FUTURE SCOPE

Web services in particular suffer from producing recommendations of innumerable items to millions of users. The time and computational power can restrain the performance of even the best hybrid systems available. For a larger dataset we can focus on scalability problems of a recommendation system

The Prediction approach can also be tried on numerous datasets to test harmony performance of system scalability problems of recommendation systems.

VII. RESULT AND DISCUSSION:-

Our system will be a web-based application which will provide the user with accurate and unerring recommendations based on the user's preferences and interests.

For the frequent users: The system will show them recommendations according to their previous searches and purchases. Similar items will be displayed to them which will conveniently match their previous choices.

For the new users: Our system will give them the top recommendations which will be based on the data collected from the frequent user.

VIII. CONCLUSION:-

Recommendation system provides user with the most accurate and efficient recommendation according to their interests and previous choices. For building the system we used the algorithm Collaborative filtering and Content based filtering. Both of which helped in attaining the desired results. Content based filtering recommended the item which is based on user's profile, which the user has liked in the past. It served as finding relations/correlation among items in a large database. And Collaborative filtering was used to analyze the user's behavior by predicting the users' taste to that of similar to another user.

REFERENCES:-

[1]https://www.researchgate.net/publication/294860665_Recommender_Systems_Issues_Challenges_and_Research_Opportunities
[2]https://pdfs.semanticscholar.org/0b91/5f2a711b1e2696563b0e148c5805d72b0148.pdf
[3] E. Uko Oko, B. O. Eke and P.O Asagba "An Improved Online Book Recommender System using Collaborative Filtering Algorithm" in proceedings of International Journal of Computer Applications, 2018
[4]https://www.studysevt.com/science/article/pii/S1110866515000341
[5] https://www.techemergence.com/use-cases-recommendation-systems/
[6]https://www.sciencedirect.com/science/article/pii/S0167923615000627
[7] Gediminas Adomavicius ,Jesse Bockstedt ,Shawn Curley and Jingjing "Impact of Recommender Systems on Consumer Preferences: A Study of Anchoring Effects",California,2014,pp 241-261
[8] W. Kowalczyk, Z. Szlávik And M. C. Schut "The Impact of Recommender Systems on Item-, User-, and Rating-Diversity", in proceedings of International Workshop on Agents and Data Mining Interaction, Web,Amsterdam,2011, pp 261-287
[9] Sarwar, B., Karypis, G., Konstan, J., And Reidl, J.2"Item Based Collaborative Filtering Recommendation Algorithms", in proceedings of 10th International Conference on World Wide Web, ACM, New York, 2001, pp 285-295.
[10]https://aws.amazon.com/blogs/big-data/generating-recommendations-at-amazon-scale-with-apache-spark-and-amazon-dsstne/
[11]https://beta.vu.nl/nl/Images/werkstuk-postmus_tcm235-877824.pdf
[12]https://storage.googleapis.com/pub-tools-public-publication-data/pdf/45530.pdf
[13]https://tech.flipkart.com/e-commerce-recommendations-using-machine-learning-5002526e5

Search Nearest Keyword Set Using Projection and Multiscale Hashing

Nayana Agwan, Vaishali Nirgude
Department of Information Technology,Thakur College of Engineering and Technology, Mumbai, India
Department of Information Technology,Thakur College of Engineering and Technology, Mumbai, India
agwan3nayana@gmail.com, vaishali.nirgude@thakureducation.org

Abstract— **In this paper, objects such as images ,documents are tagged with keywords and described by gathering of important features and usually represented as points in multidimensional space.In multidimensional dataset to place each information point need a set of keywords.The presence of keywords in feature space allows for the development of new tools to query.In multidimensional dataset images are represented using color feature vectors.NKS query is set of user given keywords and the outcome of the query include k sets of data points each of which include all the query keywords.Such as existing strategies utilizing tree based indexes that give answers for NKS queries with respect to multidimensional datasets but performance decreases with the expansion of size of dataset.To this motivation in this paper uses short for projection and multi-scale hashing known as exact ProMiSH that gives optimal outcomes.It utilize random projection and hash built list structures to achieves scalability.Exact Promish used for quick processing for NKS queries and also it give fast access to multidimensional objects.**

Index Terms— multi-dimensional data, indexing, hashing.

I. INTRODUCTION

Keyword based hunt over text rich multidimensional datasets facilitates number of novel provisions and instruments. In multidimensional datasets to place each information point need a set of keywords. The presence of keywords in characteristic space allows for the development of new tools to query and investigate these multidimensional datasets.Objects like images,documents and chemical compounds are described by an accumulation of important features and usually represented as points in multidimensional characteristic space.Examples such as pictures are represented utilizing color characteristic vectors and more typically have spellbinding data such as labels or keywords connected with them.Regard multidimensional database place each information point need an accumulation of keywords.For a query group of points include every inquiry keywords and manifestations the tightest bunch compared with another group of points which include every inquiry keywords for achieve this it create ProMiSH. ProMiSH is short for Projection and Multi-Scale Hashing will empower quick processing for NKS queries.To perform a restricted scan it utilize hashtables and inverted indexes.The hashing procedure will be propelled toward Locality Sensitive Hashing (LSH) which will be a state-of-the-art strategy for closest neighbor hunt in high-dimensional spaces.Dissimilar to LSH-built techniques that permit just estimated look for probabilistic guarantees,those list structure in ProMiSH-E helps exact hunt. It makes hashtables at different bin-widths, known as list levels. An solitary round about scan over an hashtable get subsets of points that hold inquiry outcome, furthermore it investigates every subset utilizing an Search Nearest Keyword Set in Multi-Dimensional Datasets Using Promish quick pruning-based algorithm. Framework employments an flicker images to an exploration around multi-dimensional dataset. This system is work on search keyword query and display result.

II. LITERATURE REVIEW

In paper[1] author explain about nearest keyword set known as NKS queries lookinginto text-rich multidimensional datasets. NKS inquiry will be a set of user given keywords and outcome of the inquiry incorporate k sets of data points which holds every inquiry keywords and gives top-k tightest bunch in the multi-dimensional space.NKS queries need aid advantageous to numerous applications, such as photo-sharing in social networks, geolocation search in GIS systems. In paper[2] author explain about a novel spatial keyword inquiry known as m-closest keywords inquiry. Provided for a database of spatial objects, every tuple will be connected with a portion of spellbinding data quell in the structure of keywords. Those mCK inquiry means with discover the spatially nearest tuples which match m client given keywords. Provided for an set about keywords from a document, mCK inquiry useful in geo-tagging the record by comparing the keywords on different geo-tagged documents in a database. To address mCK inquiry effectively, they present another list known as bR*-tree, which is an development of the R*-tree. Depend upon bR*-tree, they misuse a priori-built hunt methodologies to adequately diminish the hunt space.Professor. Wanda Stryla et al [2], the main goal of the article is to attempt to preliminarily analyze the impact of physiotherapy in virtual reality on improvement in motor performance of the paretic upper limb in patients after ischemic stroke. An additional aim of this work is to determine the effect of exercises in a virtual environment on nonpareil's upper limb in the same group of patients.

In paper[3] author describes multidimensional spatial information are got when a amount of data acquisition units are deployed throughout various zones on measure an certain group about qualities of the study subject. In this paper, it concentrate on characterizing 3d spatial operations and connections to 3d spatial components (points, tetrahedrons)and then applying these operations on 3d spatial objects, where every object is made of a group of tetrahedrons. In paper[4] author investigates there is a important business and research enthusiasm toward area built web search engines. Provided a amount of hunt keywords and one or more areas that a client may be intrigued in, a location-built web scan retrieves and ranks most textually and spatially pertinent web pages. In this kind for hunt, indexed both spatial and textual data. To handle location-built web hunt in an effective way in this paper they proposes another list known as Spatial-Keyword Inverted File.

They develops another separation measure called spatial tf-idf for to seamlessly discover furthermore rank important documents.

In paper[5] author explain about collective spatial keyword query which is used to discover a group of objects in the dataset which contains a group of provided keywords collectively and has littlest cosset. In paper[6] author specifies spatial web objects that have both an geological area and also a printed depiction need aid picking up in pervasiveness also spatial keyword queries that misuse both area and printed depiction would picking up previously, noticeable quality. However, the queries concentrated on in this way by concentrate on finding distinct objects that every fulfills an inquiry instead of discovering set of objects where the objects in a set collectively fulfill a inquiry.In paper[7] author describes mapping mashups need aid developing Web 2. 0 provisions for which information objects such as blogs, photographs are combined from distinct sources furthermore checked in a map utilizing APIs that need aid discharged toward web mapping results for example Yahoo Maps. These objects are commonly connected with an situated from claiming tags catching the inserted semantic and a situated of coordinates demonstrating their geological areas. In this paper, they concentrates on the basic provision from claiming placing geological assets and proposes an productive tag centric query transforming system. Main goal to discover group of closest co-located objects which together match those inquiry labels.

In paper[8 author describes pictures with GPS coordinates are a rich source of data for an geographic area. Inventive client administrations and provisions are being constructed utilizing geo labeled pictures taken starting with group keeping contributed storehouses such as Flickr. Best an little subset of the pictures in these storehouses is geo tagged, restricting their investigation and powerful usage. They recommend to utilize discretionary meta-data alongside picture substance to geo-cluster every pictures in a partly geo labeled dataset.In paper[9] author recommended SIMP for r-NN inquiry in a high dimensional space. Dissimilar to state-of-the-art techniques SIMP gives 100 percent correctness and also effectiveness to at whatever query extent. To obtain a high rate of pruning and best performance SIMP utilize projection and spatial intersection.In paper[10] author tended to the issue of querying huge subregions in raster pictures.It developed generic scoring scheme to measure similitude between an inquiry picture and an picture area.In paper[11] to project the points author utilizes random vectors built from p-stable distributions.

III. PROBLEM DEFINITION

The main problem is existing strategies utilizing tree based indexes that provide solutions to NKS queries with respect to multidimensional datasets but performance decreases with the increase of size of dataset. Results show that these algorithms may take hours to terminate for a multi-dimensional dataset of millions of points.So for this purpose promish algorithm used that scales with dataset dimension and yields practical query efficiency on large datasets.To search nearest keyword set in Multi-Dimensional Datasets by using ProMiSH that uses random projection and hash-based index structures which gives high scalability and speedup .The main goal is searching nearest keyword set with appropriate information as per given query keyword.

IV. PROPOSED SYSTEM

Dataset Construction:First to construct the dataset of images and documents.Admin crawl the images online from flickr with different tag and store into database so every image store with tagged and view count and the url of images.Also dataset for text images admin upload the text image and store it into database.Document which containing text file. Then dataset for document admin first upload the text file and store in database with file name,file description and file use.

HI Construction: HI construction it comprises of various hash tables and inverted indexes known as HI.It contains two components first is inverted index and second is hashtable inverted index pair.In inverted index treat keyword as keys and every keyword points to a set of data points that are associated with keyword.It having three paremeters index level,Number of unit random vectors and hashtable size.Project every points in dataset on a unit random vector and segment the projected values into overlapping bins of binwidth.

Image preprocessing and feature extraction:To acquire the data points first images are converted into grayscale. After that extracting its color histogram and convert every image into a d-dimensional point and associate every data point with a set of keywords.

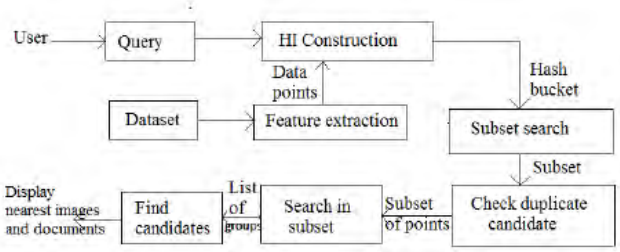


Fig 1. System Architecture

Search NKS: Subset search: ProMiSH-E retrieves every lists of hash bucket ids comparing to keywords in Q from the inverted index. Intersection of these lists get a set of hash buckets which holds every inquiry keywords.For each chosen hash bucket to obtain a subset of points ProMiSH-E retrieves every points in the bucket and filters these points utilizing bitset.Subset holds those points which are labeled with at least one inquiry keyword and investigated further.

Check duplicate candidate::Then depend upon identifiers points are sorted.After that applying two distinct standard hash functions to produce two hash values and then concatenated these hash values to obtain a hash key h for the subset. To decrease hash collisions it utilize various hash functions. Then to performed an element-wise match if hashtable of subset already has a list of subsets at h.

Search in subset:Then search in subset kth littlest diameter rk is retrieved from the priority queue.For provided subset and a query every points are bunch utilizing query keywords.Then performed a pairwise inner join of the bunch.The distance between points stores in adjacency list which fulfill the distance threshold rk and the count pairs between groups stores in an adjacency list M .

Find candidates: An intermediate tuple is verified against every point of list of bunch then controlled it utilizing adjacency list if the separation between the last point in curSet and a point in list of bunch is at most rk.Then,update diameter of curSet. NewCurSet is obtained if a point fulfill the distance predicate with every point of curSet.A candidate is found if curSet has a point from all bunch.Then display images and documents nearest to query.

V. CONCLUSION

In this paper ,system provide a fundamental evaluation of exact promish algorithms in the searching of nearest keyword result.In this evaluation admin crawl images online from flickr and store it into database.Also admin add text images and add documents in database.After loading the dataset perform image preprocessing and extracting the feature of dataset images like color. From that feature construct histogram and obtain data points. Then search point having at least one query keyword which gives subset. Then check duplicate candidate in the subset.After removing duplicate candidate search in subset which gives list of groups. From that subset find the nearest result to the query.To search the nearest images as per query it display all the related images which stored in database.So in database total 148 images are stored and in result it display those images which are nearest to query.Also keywords per query increases response time of algorithm also increases.In database total 35 documents and 12 text images are stored.System extracted nearest documents and images as per given query.This system gives the solutions to the problem of searching nearest keyword set in multi-dimensional datasets. So ,ProMiSH-E that finds nearest result related to the query.

In future, we plan to explore different scoring schemes for ranking the result sets techniques like tf-idf.

REFERENCES

[1] Vishwakarma Singh, Bo Zong, and Ambuj K. Singh, “Nearest Keyword Set Search in Multi-Dimensional Datasets”, VOL. 28, NO. 3, MARCH 2016.

[2] D. Zhang, Y. M. Chee, A. Mondal, A. K. H. Tung, and M. Kitsuregawa, “Keyword search in spatial databases: Towards searching by document”, in Proc. IEEE 25th nt. Conf. Data Eng., 2009, pp. 688–699.

[3] W. Li and C. X. Chen, “Efficient data modeling and querying system for multidimensional spatial data ”, in Proc. 16th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst., 2008, pp. 58:1– 58:4.

[4] Y. Zhou, X. Xie, C. Wang, Y. Gong, and W.-Y. Ma, “Hybrid index structures for location-based web search ”, in Proc. 14th ACM Int. Conf. Inf. Knowl. Manage., 2005, pp. 155–162.

[5] C. Long, R. C.-W. Wong, K. Wang, and A. W.-C. Fu, “Collective spatial keyword queries: A distance owner-driven approach”, in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2013, pp. 689–700.

[6] X. Cao, G. Cong, C. S. Jensen, and B. C. Ooi, “Collective spatial keyword querying”, in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2011, pp. 373–384.

[7] D. Zhang, B. C. Ooi, and A. K. H. Tung, “Locating mapped resources in web 2.0 ”in Proc. IEEE 26th Int. Conf. Data Eng., 2010, pp. 521–532.

[8] V. Singh, S. Venkatesha, and A. K. Singh, “Geo-clustering of images with missing geotags ”in Proc. IEEE Int. Conf. Granular Comput., 2010, pp. 420–425.

[9] V. Singh and A. K. Singh, “SIMP: Accurate and efficient near neighbor search in high dimensional spaces”, in Proc. 15th Int. Conf. Extending Database Technol., 2012, pp. 492–503.

[10] V. Singh, A. Bhattacharya, and A. K. Singh, “Querying spatial patterns”in Proc. 13th Int. Conf. Extending Database Technol.: Adv. Database Technol., 2010, pp. 418–429.

[11] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, “Locality sensitive hashing scheme based on p-stable distributions”in Proc. 20th Annu. Symp. Comput. Geometry, 2004, pp. 253–262.

Data Mining for Diagonisis of Diabetes

Mrs. Neha Kamble, Mrs. Rashmi Thakur
Department of Computer Engineering, Thakur college of Engineering &Technology, Mumbai
nehakamble19@gmail.com, thakurrashmik@gmail.com

Abstract— **Diabetes is a disease which is affecting many people now-a-days. Most of research is happening in this area. In this project, we proposed a model to solve the problems in existing system in applying data mining techniques namely clustering and classifications which are applied to diagnose the type of diabetes and its severity level for every patient from the data collected. The ongoing report of WHO demonstrates a exceptional climb in the quantity of diabetic patients and this will be in a similar example in the coming decades too. Early distinguishing proof of diabetes is an essential challenge. Information mining has assumed an essential job in diabetes inquire about. Information mining would be an important resource for diabetes specialists since it can uncover concealed learning from a gigantic measure of diabetes related information. Different information mining systems help diabetes investigate and eventually enhance the nature of social insurance for diabetes patients. This paper gives a review of information mining techniques that have been usually connected to Diabetes information investigation and forecast of the ailment.**

Keywords— Data mining, SVM, Naïve Bayes, Decision tree, KNN.

I. INTRODUCTION

Today the buzz word is "Health Care" all over the world. Early Prediction of diseases can reduce the fatal rate of human. There are very large and enormous data available in hospitals and medical related institutions. Information technology plays a vital role in Health Care. Diabetes is a chronic disease with the potential to cause a worldwide Health Care crisis.

Early prediction of diabetes is quite challenging task for medical practitioners due to complex interdependence on various factors. Diabetes affects human organs such as kidney, eye, heart, nerves, foot etc. Data mining is a process to extract useful information from large database. It is a multidisciplinary field of computer science which involves computational process, machine learning, statistical techniques, classification, clustering and discovering patterns. Data mining techniques has proved for early prediction of disease with higher accuracy in order to save human life and reduce the treatment cost.

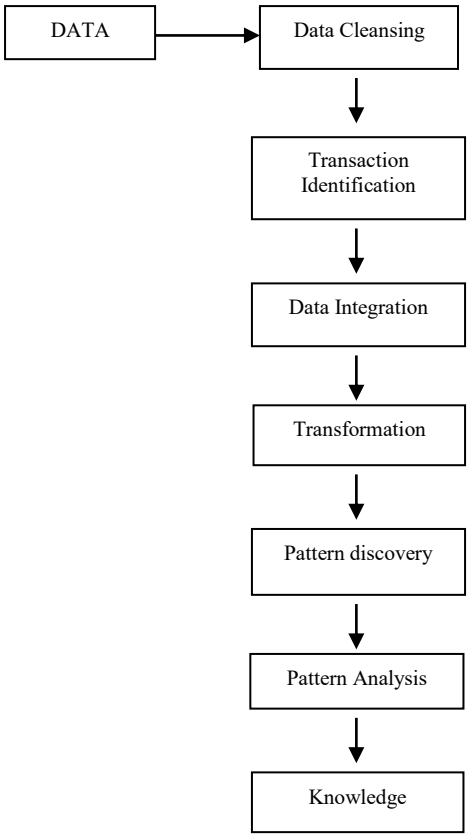


Fig. 1: Data Mining Process

Data Mining alludes to separating learning from a lot of information [1]. It empowers us to investigate the vast examples and investigate the equivalent by methods for Factual and Artificial Intelligence in expansive datasets [2]. The Data Mining system is utilized to foresee conceivable future patterns or to find concealed examples in the conduct of the information. Strategies, for example, Fake Neural Networks, Decision Trees, Arrangement, Clustering, Association rule calculations and so forth are generally used by specialists.

Data mining procedures is broadly being connected by analysts in Bioinformatics. Bioinformatics is the investigation of putting away, extricating, sorting out, deciphering what's more, using data from organic arrangements what's more, and atoms [3]. As of late, Knowledge Disclosure and Data Mining methods are generally utilized for removing the examples from the huge natural databases. The measure of organic information is developing quickly [4]. Examining these informational indexes requires comprehending the information by construing structure or speculations from the information. The cooperation between data mining and

bioinformatics assumes a fundamental job in analysis numerous sicknesses.

Diabetes is a state or on the other hand a condition in which the body can't create or on the other hand use the insulin appropriately which bringing about creating maladies influencing kidney, eyesight ,nerve framework, veins and furthermore heart related issues [6]. There are three sorts of diabetes. Type-1diabetes used to be called adolescent beginning diabetes. It is generally brought about by an auto-resistant response where the body's guard framework assaults the cells that produce insulin. Type-1 is likewise called insulin subordinate. Practically 90% of the diabetes experiences Type II diabetes which is on the other hand called as non-insulin subordinate diabetes or Adult-beginning diabetes. This is portrayed by protection from insulin and insufficiency of relative insulin which may either or both together be present at the season of which diabetes is analyzed. In the instance of Gestational Diabetes (GDM) which is frequently seen amid pregnancy coming about because of high glucose levels. This is regularly cultivated by renal intricacies, cardiovascular infections and fringe vascular illnesses. Early recognizable proof of patients with undiscovered sort 2 diabetes or those at an expanded danger of creating type-2 diabetes is an imperative challenge in the field of medicinal. The accessibility of immense measures of medicinal information prompts the requirement for incredible information investigation devices to remove helpful learning. Scientists have for quite some time been worried about applying factual what's more, information mining apparatuses to enhance information examination on vast informational indexes. Sickness conclusion, for example, diabetics is one of the applications where information mining devices are demonstrating victories in the ongoing years.

II. LITERATURE SURVEY

According to V. Anuja Kumari, R.Chitra [7], utilized datasets for diabetes infection from the machine learning research center at University of California, Irvine. All the patients" data are prepared by utilizing SVM. The decision of best estimation of parameters for specific piece is basic for a given measure of information SVM approach can be effectively used to recognize a typical illness with basic clinical estimations, without research facility tests. In the proposed work, SVM with Radial premise work portion is utilized for arrangement. The execution parameters, for example, the characterization precision, affectability, and particularity of the SVM and RBF have observed to be high along these lines making it a decent alternative for the arrangement procedure. In their research, they were using SVM and for various calculations they had used Matlab which is giving accuracy of 78%.

According to Zahed Soltani and Ahmad Jafarian[8], using artificial neural networks, they can design and implement the complex medical processes as software. These software systems in turn are effective for different fields of medicine sciences such as diagnosis, treatment and to help Surgeons, physicians and the public. These systems can be implemented in different scales in a parallel and distributed manner. In general, ANNs are parallel processing systems which are used for identifying complex patterns among data. Therefore, in there paper PNNs are applied to identifying diabetes disease type 2. they implemented the PNN model in MATLAB. The Pima Indians Diabetes data set was used for diagnosing diabetes type 2, which consists 768 data samples with 9 features. 90% of these 768 samples are used as training set and 10% used as testing set. The method achieved 89.56% of diagnosis accuracy in training phase, and 81.49% in test phase. Both training and testing measures could identify the diabetes disease type 2 with a good accuracy.

In paper [9] the authors have developed an expert healthcare predictive decision support system that predicts diabetes. This model is trained on Pima diabetes dataset. Decision tree and K-nearest neighbor algorithms are used to develop the model and found that C4.5 algorithm has achieved 90.43% accuracy.

In paper [10] authors developed a prediction model that would predict whether a person would develop diabetes by considering daily lifestyle activities. To build prediction

Model PIMA diabetes data set was used and CART Classification and Regression Trees) machine learning classifier was applied. The proposed model could provide an accuracy of 75%.

Chaudhari et al[11] Disease diagnosis is one of the most important applications of such system as it is one of the leading causes of deaths all over the world. Predict the human use the inputs from complex tests conducted in labs

and also predict the disease based on risk factors such as tobacco smoking, alcohol intake, age, family history, diabetes, hypertension, high cholesterol, physical inactivity, obesity. Researchers have been using several data mining techniques to help health care professionals in the diagnosis of heart disease. K-Nearest-Neighbor (KNN) is one of the successful data mining techniques used in classification problems. Recently, researchers are showing that combining different classifiers through voting is outperforming other single classifiers. This paper investigates applying KNN to help healthcare professionals in the diagnosis of disease specially heart disease. It also investigates if integrating voting with KNN can enhance its accuracy in the diagnosis of heart disease patients. The results show that applying KNN could achieve higher accuracy than neural network ensemble in the diagnosis of heart disease patients. The results also show that applying voting could not enhance the KNN accuracy in the diagnosis of heart disease.

Thangaraju et al [12] Data mining is the act of looking at huge previous databases so as to produce new information. There are various types of information mining methods are accessible. Order, Clustering, Affiliation Rule and Neural Network are the absolute most noteworthy methods in information mining. In Health care businesses, Data mining plays a noteworthy job. Most oftentimes the information mining is utilized in human services enterprises for the way toward anticipating ailments. Diabetes is a ceaseless condition. This implies is goes on for quite a while, frequently for somebody's entire life [13]. This paper contemplates the examination of diabetes estimating approaches utilizing grouping procedures. Here we are utilizing three various types of bunching strategies named as Hierarchical grouping; Thickness based bunching, and Simple K-Means grouping. Weka is utilized as a device.

Berina Et Al. [3]: Abstract: the overview of machine learning techniques in classification of diabetes and cardiovascular diseases (CVD) using Artificial Neural Networks (ANNs) and Bayesian

Networks (BNs). The comparative analysis was performed on selected papers that are published in the period from 2008 to 2017. The most commonly used type of ANN in selected papers is multilayer feed forward neural network with Leven berg-Marquardt learning algorithm. On the other hand, the most commonly used type of BN is Naive Bayesian network which shown the highest accuracy values for classification of diabetes and CVD, 99.51% and 97.92% retrospectively. Moreover, the calculation of mean accuracy of observed networks has shown better results using ANN, which indicates that higher possibility to obtain more accurate results in diabetes and/or CVD classification is when it is applied to ANN.

III. METHODOLOGY

A. *Random forest*

Random Forest algorithm is capable of performing each classification and regression tasks. The fundamental principle of RF is that a group of weak learners' come together to create a robust learner. Random forest rule uses bagging approach to form the bunch of decision trees with random set of the information. The model is trained few times on random sample of the dataset to attain best prediction performance from the RF rule. In this ensemble technique of learning, the output of all decision trees within the RF is combined to form a final prediction. The final prediction of the RF rule is derived after polling the results of every decision tree. Suppose there are N cases within the training set. Then these N samples are taken randomly however, with replacement. These samples are training set for growth of tree. If m < M is specific. The simplest split of this m is employed to separate the node. The value of m is constant whereas growing the forest.

B. *Decision tree*

Decision tree is a tree structure, which is in the form of a flowchart. It is used as a method for classification and prediction with representation using nodes and internodes. The root and internal nodes are the test cases that are used to separate the instances with different features. Internal nodes themselves are the result of attribute test cases. Leaf nodes denote the class variable. Figure4.2 shows a sample decision tree structure.

Decision tree provides a powerful technique for classification and prediction in Diabetes diagnosis problem. Various decision tree algorithms are available to classify the data, including ID3, C4.5, C5, J48, CART and CHAID. In this paper, J48 decision tree algorithm [10] has been chosen to establish the model. Each node for the decision tree is found by calculating the highest information gain for all attributes and if a specific attribute gives an unambiguous end product (explicit classification of class attribute), the branch of this attribute is terminated and target value is assigned to it.

C. *K-Nearest Neighbor (K-NN)*

It is the closest neighbor calculation. The k-closest neighbor's calculation is a procedure for grouping objects dependent on the following preparing information in the element space. It is the easiest among all instruments learning calculation. This calculation is introduced by choosing k focuses in kd as the underlying k group delegates or "centroids". Procedures for select the essential seeds incorporate inspecting indiscriminately from the dataset setting them as the arrangement of bunching a little subset of the information or bothering the worldwide mean of the information k-times. At that point the calculation emphasizes between two stages till intersection:

Step1: In Data Assignment every datum point is allocated to its connecting centroid, with ties broken subjectively. This outcomes in an apportioning of the information.

Step2: Relocation of "signifies" each gathering agent is moved to the focal point of all information directs appointed toward it. On the off chance that the information focuses accompany a plausibility measure, at that point the migration is to the desires for the information segments.

D. *SVM*

Support vector machine is a machine learning approach that can be used as classifier as well as for regression. SVM classifies the data into different classes by finding hyperplane (line) which separates training data into classes. SVM does not overfit the data and gives best classification performance in terms of precision and accuracy.SVM does not make any strong assumptions on data. It shows more efficiency for correctclassification of the future data. SVM is classified into 2 categories i.e. Linear and non-Linear. In linear approach, training data is separated by line i.e. hyperplane.

E. *Naïve Bayes*

Naïve Bayes classifier is based on Bayes theorem. This classifier uses conditional independence in which attribute value is independent of the values of other attributes.

The Bayes theorem is as follows:

Let X={x1, x2,……,xn} be a set of n attributes.

In Bayesian, X is considered as evidence and H be some hypothesis means, the data of X belongs to specific class C.

We have to determine P (H|X), the probability that the hypothesis H holds given evidence i.e. data sample X. According to Bayes theorem the P (H|X) is expressed as

$$P(H|X) = P(X|H)\, P(H) / P(X).$$

F. *Neural Network*

Artifical Neural Network is inspired from the brain that is considered as a data processing system. In this algorithm, many microprocessors are responsible for data processing and they are acting as an interconnected and parallel network with each other to solve a problem. By using programming science in this network, a data structure is designed that can act as a neuron and this data structure is called neuron. By setting a network between the neurons and applying a learning algorithm, the network is trained. In this Neural Network, neurons are divided into two enable (NO or 1) or disable (OFF or 0) modes and each edge (synapses or connections between nodes) has a weight. Edges with positive weight, stimulate or enable the next disable nodes and edges with negative weights, disable or inhibit the next connected nodes (if they are enabled)

IV. CONCLUSION

In this paper, the various techniques are discussed for predict the diagnosis of diabetes. Using the data mining technique the health care management predicts the disease and diagnosis of the diabetes and then the health care management can alert the human being regarding diabetes based upon this prediction. From this literature survey we conclude that C4.5 is giving maximum accuracy of 90% then various other techniques.

REFRENCES

[1] Jiawei Han and Micheline Kamber, "Data Mining :Concepts and Techniques", 2nd edition

[2] E. Papageorgiou, I. Kotsioni, A. Linos, "Data Mining:A New Technique In Medical Research", *HORMONES*,4(4):pp 189-91,2005.

[3] Mohammed J. Zaki, Jason T. L. Wang,Hannu T.T.Toivonen, "BIOKDD01: Workshop on Data Mining in Bioinformatics",www.kdd.org/sites/default/files/issues/3-2-2002-01/zaki.pdfK. Elissa, "Title of paper if known," unpublished.R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[4] Khalid Raza,"Application of data mining in bioinformatics", *Indian Journal of Computer Science and Engineering*, Vol 1 No 2, pp 114-118M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[5] Screening for type-2 diabetes: Report of a World Health Organization and International Diabetes Federation meeting.www.who.int/diabetes/publications/en/screening-mnc03.pdf.

[6] B. M. Patil, R. C. Joshi, Durga Toshniwal, "Association rule for classification of type -2 diabetic patients",*Proc. of the Second International Conference onMachine Learning and Computing*, pp 330-334, 2010

[7] V. Anuja Kumari, R.Chitra,"Classification Of Diabetes Disease Using Support Vector Machine", International Journal of Engineering Research and Applications (IJERA),2013.

[8] Zahed Soltani and Ahmad Jafarian," A New Artificial Neural Networks Approach for Diagnosing Diabetes Disease Type II", IJACSA,2016.

[9] Hashi EK, et al. An expert clinical decision support system to predict disease using classification techniques. In: International conference on electrical, computer and communication engineering (ECCE), ©2017 IEEE, February 16–18, 2017.

[10] Anand A. Prediction of diabetes based on personal lifestyle indicators. In: 2015 1st international conference on next generation computing technologies (Ngct-2015) Dehradun, India, 4–5 September 2015.

[11] Anand A. Chaudhari, Prof.S.P.Akarte, " Fuzzy and Data Mining based Disease Predection using K-NN Algorithm", International Journal of Innovations in Engineering and Technology, Vol. 3, Issue No. 4, April 2014.

[12] P. Thangaraju, B.Deepa, T.Karthikeyan, "Comparison of Data mining Techniques for Forecasting Diabetes Mellitus", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue No. 8, August 2014.

[13] Rakesh Motka, Viral Parmar, Balbindra Kumar, A. R. Verma, " Diabetes Mellitus Forecast Using Different Data Mining Techniques", nternational conference on computer and Communication Technology

Survey on MRI Image Segmentation for Cancer Detection

[1]Mrs. Ujwala V. Gaikwad, [2]Mr. Vishwajit B. Gaikwad, [3]Dr. Kamal Shah
[1,2]*Computer Engineering Department Terna Engineering College, Nerul, Navi Mumbai, India,*
[3]*Information Technology Thakur College of Engineering and Technology, Mumbai, India*
[1]ujwalagaikwad@ternaengg.ac.in, [2]vishwajitgaikwad@ternaengg.ac.in, [3]kamal.shah@thakureducation.org

***Abstract* – Image processing techniques have vast use in the medical field especially for detection of many diseases. Detection of cancerous tumors is challenging job in various types of images related to cancer. This paper studies the various cancer detection mechanism related to MR images. This process uses segmentation as one of the method. Depending on the image modality the features extracted may differ. Using these features the machine learning techniques generates the valuable information.**

Keywords*: *Cancer detection, MRI, segmentation.

I. INTRODUCTION

Cancer is the second cause of death in the world. Though there is decrease in mortality rate [1] but number of increasing cancer patient is prominent [2]. Cancer is disease where uncontrolled growth of cells happens which can be considered as tumors. Not all tumors are cancerous. Only malignant tumors are considered as cancerous. Detection of cancer is done through different tests like Mammography, Magnetic resonance imaging (MRI), Human papillomavirus (HPV) testing, Colonoscopy, Sigmoidoscopy, Fecal occult blood test (FOBT), computed tomography (CT) scan [3]. But screening tests may have risks, some screening tests can cause serious problems, False-positive and False-negative test results are possible, over diagnosis is also possible which may lead to overtreatment which may not be required. While MRI scans help in diagnosis of malignant tumors in breast, brain, prostate cancers where less possibility of risks is found.

Segmentation is the image processing technique which helps to detect objects. This is useful in many application areas. By applying the image processing techniques like segmentation can help in detection of tumors. Segmentation is divided in to mainly two categories discontinuity based and similarity based using pixel intensity. Thresholding, edge based and region based methods are used for segmentation.

II. LITERATURE SURVEY

There are many applications of segmentation found in literature but here only MRI images related cancer are considered. There are various types of cancer but which mostly observed are brain, breast, prostate which can be detected through MR images, CT images as well with ultrasound images.

a. Brain Cancer

The figure 1 shows a cancerous tumor in the brain

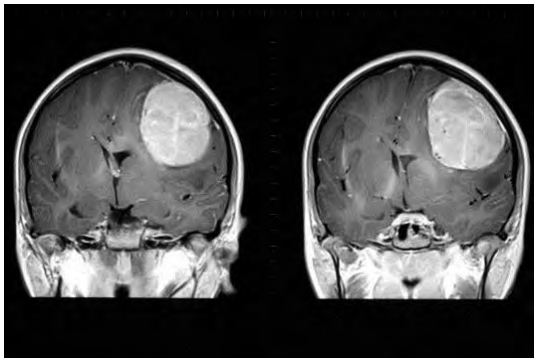


Fig 1: Brain MRI showing malignant tumors[4]

Saeid Fazli, Parisa Nadirkhanlou [5] worked on MRI images of brain tumor. For pre-processing noise is removed by anisotropic diffusion filter which preserves the required edges and structures. Using fast bounding box algorithm tumors are extracted. This algorithms considers the symmetry of brain as basis for finding the tumor region. Using SVM classifier the images are classified.

Anjali Gupta, Gunjan Pahuja[6] worked on brain tumor segmentation using hybrid clustering. The image is preprocessed using Gaussian filter. For clustering Fuzzy C-mean clustering with FCM thresholding is used. Finally level set method is used for calculating the exact tumor shape.

Jun Cheng et al. worked on brain tumors[7]. In this approach they constructed augmented tumor regions on the T1-weighted contrast enhanced MRI images as a first step. Then the augmented tumor regions splitted into sub-regions by using adaptive spatial division method which uses intensity orders. Within subregions image patches are extracted as local features. Fisher kernel vector per sub-region is created using local features of subregions and all the vectors are concatenated to form a single vector which is considered as image level signature. The vectors in query image are compared with the database images by using closed form metric learning algorithm for similarity.

Meiyan Huang et al[8] derived CBIR system using spatial layout information for brain tumor. This system uses T1 – weighted contrast enhanced MR images. The bag-of-visual-words model is created using tumor information from the known images where query images are compared. This model uses partition learning for extract informative features. An algorithm called the Rank Error-based Metric Learning (REML) which is distance metric learning based algorithm, reduces the semantic gap between low-level visual features and high-level semantic concepts.

Sérgio Pereira et al[9] worked on brain tumor segmentation using CNN on MR images. A small 3x3 kernel helps to create a CNN which segments the MR images. Before applying the segmentation the image is preprocessed using intensity normalisation and data augmentation which gives very effective results of segmentation.

b. Breast Cancer

Figure 2 shows the MRI of a breast having cancerous tumour

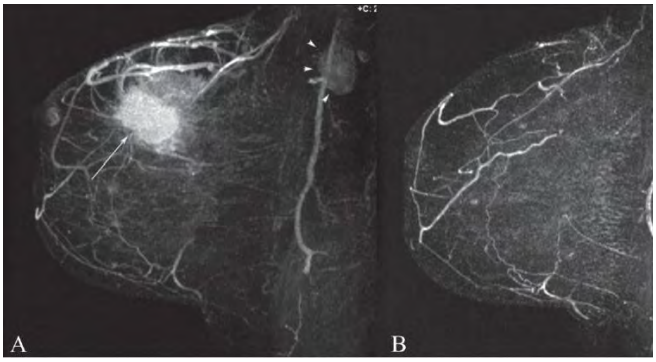


Fig 2: breast cancer tumour A) before B) after chemotherapy[10]

Liangbin Zhang et al.[11] used lesion segmentation on breast tumor MRI images of DWI and DCE modalities. This segmented information is processed using Fuzzy c-mean (FCM) combined with Dempster-Shafer (DS) method. This approach used Time –intensity curve for making the decision.

Aida FOOLADIVANDA et al.[12] worked on breast region segmentation. Wall boundaries are detected using intensity based operation. SVM is used to determine the connectivity of tissues in the features. Final chest wall boundaries obtained using Atlas-based method are refined using geometric shape.

Dr. H. B. Kekre, Ms. Tanuja K. Sarode, Ms. Saylee M. Gharge[13] considered the vector quantization method detection of tumor in breast cancer MRI images. A Linde Buzo-Gray algorithm is used for segmentation. This algorithm quantizes the vector into 128 codes and these codes are clustered into 8 clusters using LBG algorithm. These 8 clusters are the segmented images which has shown improved results in comparison with watershed and GLCM.

Snekha Thakran et al.[14] worked on segmentation of DCE MRI images for outer and inner breast tissues. This is done in 2 steps first step as outer segmentation done using profile screening, Otsu thresholding, morphological operations and empirical observation where breast tissues are separated from other body. The inner segmentation is done using similar operations. The inner breast tissues are categorised into fibroglandular(FG), fatty and tumor tissues. The computed results are verified with manual segmentation using average of Dice and Jaccard coefficients.

Yuanjie Zheng et al.[15] proposed segmentation and classification for breast tumors into malignant or benignant which used dynamic contrast-enhanced MR images. The new graph cut based segmentation method is proposed for fine identification of tumors. By using serial contrast enhanced images a model is created which shows spatio-temporal segmentation of tumors. This model helps to extract the spatio-temporal enhancement patterns. These patterns shows the dynamic enhancement and architectural features and also spatial variation in pixel. Temporal features extracted using Fourier transform and spatial features are extracted using moment and rotation invariant Gabor textures.

c. Prostate Cancer

Figure 3 shows the multi parametric MRI of prostate cancer

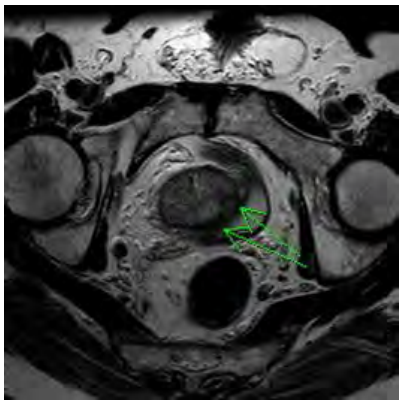


Fig 3: Multiparametric MRI of prostate cancer[16]

Yu Guo et al.[17] suggested automatic segmentation method for prostate cancer MRI. Cancerous tissue information is extracted using fuzzy c-mean clustering. Adaptive fuzzy fusion operator which is based on Bayesian model is created. This model calculates the membership degree map on the data of clusters. By using this membership degree values decision is taken for cancerous region. This method shown the improved result in peripheral as well transition zones of prostate cancer.

Navaei Lavasani S., Mostaar A., Ashtiyani M.[18] used dynamic contrast enhanced MRI for segmentation of prostate cancer. This method first pre-processed the images by registration which uses sum of squared intensity difference with reference image. The ROI is manually placed on the different modality images. The different kinetic parameters are used as semi quantitative features are calculated from the images like initial enhancement rate, maximal enhancement rate and amplitude as well as enhancement rate at various time points. For transient changes in images DWT is applied on it. Two decomposition of Haar wavelet is calculated. In addition mean, variance and detail coefficients are extracted. The fuzzy C-mean clustering is used on the MRI data.

Matthias Kirschner, Florian Jung, and Stefan Wesarg[19] discussed an automatic prostate segmentation model in MR images. This uses probabilistic active shapes. After pre-processing the image by normalising the intensities, a bounding box of features under ROI is used find the sub regions having prostate is applied. After determining the bounding box probabilistic active shape model is used for segmentation.

III. CONCLUSION

The study of various methods shows that many times Fuzzy C-mean Clustering algorithm is useful for finding the correct features which helps in segmentation for locating the tumor region. SVM is also effectively used for finding tumor region location. The work has been done various types of MRI but T2-weighted, DWI type are mostly used images which are giving more information about tumor. Dynamic contrast enhanced MR images are also useful which showing significant results. Also the study shows that the segmentation techniques along with

machine learning techniques improvise the result of cancer detection.

REFERENCES

[1] Siegel RL, Miller KD, Jemal A, Cancer statistics, 2018, CA Cancer J Clin. 2018 Jan;68(1):7-30. DOI: 10.3322/caac.21442. Epub 2018 Jan 4.

[2] https://www.cancer.gov/about-cancer/understanding/statistics referred on 22/08/2018

[3] https://www.medicinenet.com/cancer_detection/article.htm referred on 22/08/2018

[4] https://wtop.com/cancer-qa/2015/03/what-symptoms-are-typically-associated-with-brain-cancer/ referred on 22/08/2018

[5] Saeid Fazli, Parisa Nadirkhanlou, "A Novel Method for Automatic Segmentation of Brain Tumors in MRI Images", Research Institute of Modern Biological Techniques University of Zanjan, Iran, 2015. Available on https://arxiv.org/pdf/1312.7573

[6] Anjali Gupta and Gunjan Pahuja 2017 IOP Conf. Ser.: Mater. Sci. Eng. 225 012187 DOI:10.1088/1757-899X/225/1/012187

[7] Cheng J, Yang W, Huang M, Huang W, Jiang J, Zhou Y, et al. (2016) Retrieval of Brain Tumors by Adaptive Spatial Pooling and Fisher Vector Representation. PLoS ONE 11(6): e0157112. DOI: 10.1371 / journal.pone.0157112

[8] Huang M, Yang W, Wu Y, Jiang J, Gao Y, et al. (2014) Content-Based Image Retrieval Using Spatial Layout Information in Brain Tumor T1-Weighted Contrast-Enhanced MR Images. PLoS ONE 9(7): e102754. DOI:10.1371/journal.pone.0102754

[9] Sérgio Pereira*, Adriano Pinto, Victor Alves, and Carlos A. Silva Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 35, NO. 5, MAY 2016. Digital Object Identifier 10.1109/TMI.2016.2538465

[10] http://medind.nic.in/ibn/t09/i2/ibnt09i2p161.htm referred on 24/08/2018

[11] Liangbin Zhang, Wenjun Ma, Xing Shen, Yuehua Li, Yuemin Zhu, Li Chen, and Su Zhang, Research on the lesion segmentation of breast tumor MR images based on FCM-DS theory, AIP Conference Proceedings 1816, 080009 (2017); DOI: 10.1063/1.4976613

[12] Aida FOOLADIVANDA, Shahriar BARADARAN SHOKOUHI, Nasrin AHMADINEJAD, Breast-region segmentation in MRI using chest region atlas and SVM, January 2017, Turkish Journal of Electrical Engineering and Computer Sciences 25:4575-4592 DOI: 10.3906/elk-1512-40

[13] Dr. H. B. Kekre, Ms. Tanuja K. Sarode, Ms. Saylee M. Gharge, Detection and Demarcation of Tumor using Vector Quantization in MRI images, International Journal of Engineering Science and Technology, Vol.1(2), 2009, 59-66.

[14] Thakran S, Chatterjee S, Singhal M, Gupta RK, Singh A (2018) Automatic outer and inner breast tissue segmentation using multi-parametric MRI images of breast tumor patients. PLoS ONE 13(1): e0190348. https://doi.org/10.1371/journal.pone.0190348

[15] Yuanjie Zheng, Sajjad Baloch, Sarah Englander, Mitchell D. Schnall, and Dinggang Shen, Segmentation and Classification of Breast Tumor Using Dynamic Contrast-Enhanced MR Images, Med Image Comput Comput Assist Interv. 2007 ; 10(Pt 2): 393–401. available in PMC 2010 March 17.

[16] http://www.manchesterurology.org.uk/investigations/3t-multiparametric-mri referred on 24/08/2018

[17] Yu Guo, Su Ruan, Paul Walker, Yuanming Feng, Prostate Cancer Segmentation From Multiparametric MRI Based On Fuzzy Bayesian Model, 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI 2014) April 2014, DOI: 10.1109/ISBI.2014.6868008

[18] Navaei Lavasani S., Mostaar A., Ashtiyani M., Automatic Prostate Cancer Segmentation Using Kinetic Analysis in Dynamic Contrast-Enhanced MRI, J Biomed Phys Eng. 2018 Mar 1;8(1):107-116. eCollection 2018 Mar.

[19] M. Kirschner, F. Jung, S. Wesarg, "Automatic prostate segmentation in MR images with a probabilistic active shape model", Proc. Med. Image Comput. Comput.-Assisted Intervention Conf. Prostate Segment. Challenge, pp. 28-35, 2012.

E-Garage Management System

Er. Swati Ganar
Information Technology
M.H. Saboo Siddik College of Engineering
Mumbai,India
swati.ganar@mhssce.ac.in

Gulhasan Siddiquee
Information Technology
M.H. Saboo Siddik College of Engineering
Mumbai,India
gulhasansiddique13@gmail.com

Attaullah Khan
Information Technology
M.H. Saboo Siddik College of Engineering
Mumbai,India
khanattaullah6@gmail.com

Soyab Anwar
Information Technology
M.H. Saboo Siddik College of Engineering
Mumbai,India
1995soyabmohd@gmail.com

***Abstract:-*There are three panels which are enrolled in this service, first is user who want to take service, second is the garage who provide the requested services and the Admin who control and monitor all activity taking place between user and service provider (i.e. garage).With the help of this platform user first give the vehicle registration number ,after that they can select his/her willing garage notify problems and get the service(e.g. Tyre puncher, Break and clutch related, engine related etc) and estimated charges will also be generated at the time of placing order. The admin notifies the garage and send OTP to both user as well as service provider. Garage will provide the service and take the OTP in order to notify the admin, after completion of work. Toll-Free service can be used in case of emergency. User is able to give feedback to the service provider based on the performance which helps the admin to give rating to the garage.**

Keywords:-E-garage, Toll-free emergency service, Active garage, android app.

I. INTRODUCTION

In today's developing era, the numbers of vehicles are increasing in almost all over the world. So providing efficient service to each vehicle user is a challenging task for garages in future. And vehicle users have to stand in queue for getting service. Using this application the user can locate the nearby active garages /mechanics and communicate with them to get service in need.

The purpose of designing this application to facilitate the user as well as service provider (i.e. garage) user will be benefited because they not need to move the garage which ultimately saves the time, efforts and money. on the other hand (for garage) number of user increases which increase the turn over and give high return of investment.

II. LITERATURE REVIEW

In this era of technology everybody wants their work done very quick and on finger tip. The number of vehicle goes on increasing so it is difficult task for garage shop to provide quick and efficient services to their customers. The survey regarding this web application includes information

gathering from various sources. These sources include some of the garages service application, websites such as Ola app, MyMazda. Ola garage is web application provide the platform

through which user is able to get services which include all types of service. Although it is provided by Ola but it is not so popular.

Serviceko is also an android application basically meant for vehicle service purpose, it is not fully developed and currently not in use. The GUI is not simple hence difficult to use.

IEEE papers are used for knowing the advantages and disadvantages of previously proposed system. Example In paper titled as "Automobile Service Centre Management System":-It is a android based application which provide facility like Notify user for service, Next service instalment, EMI calculator but not provide service in out areas, no emergency service when net wont available and mechanics not go outside for service[2].In paper titled as" Online Management System for Automobile services": -It is a web based system using Angular JS, Mongo DB and Node-JS which provide the feature like find nearby garage service notifications but not provide Emergency service[1].For clearing the concepts and algorithms included in this project. Example Dijkstra's algorithm for finding shortest path algorithm to schedule the available mechanics and provide the services to customer in a very efficient way [3].

Advantages over existing system are:-

(1) Emergency toll-free number to get service in case if user is not accessible to internet.
(2) Service to the requested location on time and in affordable budget.
(3) Mechanics not associated with garage can also provide services.
(4) Estimated cost and time will provide to customer in advance.

III. PROPOSED SYSTEM

By keeping all the drawbacks of the existing system in the mind new proposed system have been designed. This consist of three panel User, Garage and Admin

[A]User:

1. User who wants to take service from this platform must have registered them self by providing necessary details. After registration login will be required.
2. They have to enter the vehicle registration number (Plate Number), validate the number for further processing,
3. They have to choose the types of vehicle (i.e. whether 2 wheeler or 4 wheeler), on the contrary they are able to use emergency option in case of rapid service.
4. User has to define the problem or issues in vehicle by choosing the problems. Whether the user know the problem or don't know (select accordingly).
5. Select the nearest active garage, listing of garage is based on location and the rating.

Estimated amount for the selected problem will be calculated and displayed to the user, User may continue for further processing or can change the garage. User may either come to garage or get the service to his location once agreed on price; request will be made to a respective garage.

[B]Garage:

1. Notification for a service is received and depending upon the condition (i.e. user is coming for service, or want service to that' location) action will be performed.
2. Mechanic will be assigned if service required to requested location and then mechanic details are send to user.
3. Keeping track of the mechanic.

[C]Admin:

1. Notify the user and garage
2. Send the OTP to user which help the garage owner to notify the assigned work is complete and your mechanic is free
3. Keeping track of all feedback as well the time stamp between requested to be made and granted.
4. Feedback helps to rate the garage and evaluate the list

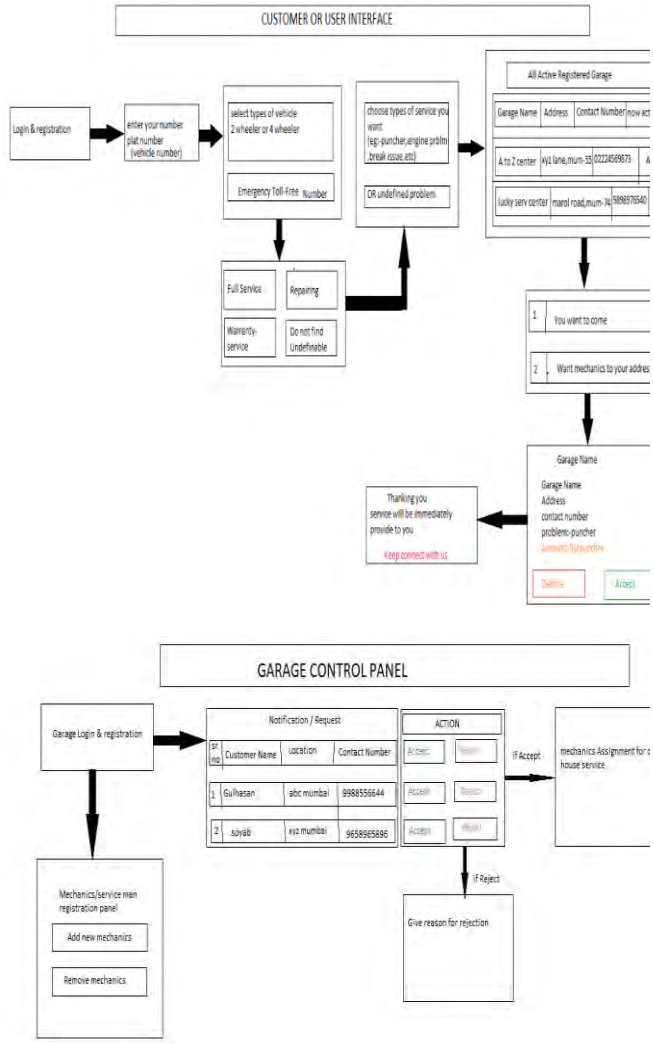
IV. SYSTEM FEATURES

Features such as:

1. User Registration: Garage shop, vehicle user and mechanics all will register.
2. Login: Both Garage shop, vehicle user and mechanics will have to Login.
3. Virtual Application Tour: One time tour about the Application.
4. Type of vehicle: People will allow choosing type of vehicle.
5. Vehicle validation: Validating vehicle by using vehicle number entered by user.
6. Type of service Required: Choosing from list of services
7. Display list of nearby active garage: List of nearby active garage.
8. Choose and send request to garage: By choosing garage from list send request to get service
9. Estimate cost and time: Cost and time required to provide the service..
10. Accept or reject response: User can accept or reject response on basis of cost and time mentioned in response.
11. Emergency toll free no. In case of user not getting access to internet or nearby garage are unable to provide service
12. Payment gateway: Online or offline
13. Track services: Tracking of both side activity

V. IMPLEMENTATION

In this section we have demonstrated the various stages such as user have to register himself/herself on this application either as garage or vehicular user or mechanics .After signing up user can login to application by providing the user name and password. Then user will provide their vehicle number which will be verified by using an API,then user have to choose the problem form given list on basis of it user will be provided with the list of nearby active garage and services offers by them along with cost .User will have freedom to choose any garage from list as per his convenience and will send request to garage to get service. Garage can either accept or reject the request. User will be provided service detail such as estimated time, cost on basis of it user also have option to reject the response of one garage and can choose other. After acceptance from both sides, service will be provided and after completion of one services admin will be notified by garage and customer will give feedback about the service.



VI. FUTURE SCOPE

Most of the people are having vehicle but they don't have time to spend for vehicle service. They can use this platform to service their vehicle by trusted service provider. User will be able to search the garage and request them for service. Easily select the problem and garage and take service from that. In emergency case user can use the Toll-Free number and get quick response. Also provide the service at location from where the request has been made whether it is remote area or area under coverage, this can be widely used to facilitate user in all over country since there is no solution for it. It also promotes the provider business and increase the number of user and return of investment also.

Future scope in enlarging the system we can add one more module where user will be provided a rent based vehicle in case he/she cannot wait till repairing. User can connect and inform each other wirelessly if they are passing close to each other so that user can also help each others in need.

VII. CONCLUSION

To overcome the drawbacks of the existing system of automobile servicing, this application will provide a platform which facilitate user who wish to take services at its location as well as in emergency also and will increase its business value .User friendly GUI and quick response will attract the user. It will increase the employee opportunities.

VIII. REFERENCES

(1) "Hanamant B. Sale , Dharmendra Bari , TanayDalvi , Yash Pandey", "Online Management System for Automobile Services", International Journal of Engineering Science and Computing (IJESC), Volume 8 Issue No.02, March-2018.

(2) "Prof. Shilpa Chavan Saket Adhav, Rushikesh Gujar, Mayur Jadhav, Tushar Limbore","Automobile Service Center Management System", International Journal of Scientific and Research Publications, Volume 4, Issue 3, March 2014

(3) "An improvement of the shortest path algorithm based on Dijkstra algorithm Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on (Volume:2). Ji-xian Xiao Coll. of Sci., Hebei Polytech. Univ., Tangshan, China FangLing Lu.

(4)"N.Shivshankaran,P.Senthilkumar","Scheduling of Mechanics in Automobile Repair shop", N. Shivasankaran et.al / Indian Journal of Computer Science and Engineering (IJCSE), Vol. 5 No.2 Apr-May 2014.

(5) "NehaSelokar, Vijay Masne, Roshani Pimpalkar, SrushtiPuranik, NidhiBhoyar", "24*7 Vehicle Management Systems for Automobile Industry" International Journal of Modern Engineering Research (IJMER)Vol.3, Issue.1, Jan-Feb. 2013 .

Hr Supportive Analysis Using Cv

Bhagyashree Mahajan, Prathamesh More, Prashant Gupta, Guide- Dr. Rekha Sharma

Department of Computer Science ThakurCollege of Engineering& Technology Kandivali, Mumbai

bhagyashreemahajan123@gmail.com, bhagyashreemahajan123@gmail.com , bhagyashreemahajan123@gmail.com

***Abstract*—This project is aimed at developing an application with web-based page and Recruitment Process System for the Private Company's and HR Consultants sector. Some features of this application will be creating aptitude test papers, Uploading CVs, Updating CVs, storing Candidates data, Interview process initiation, storing aptitude results for the candidates, and finally sending the notification of shortlisting to candidates. Using HR supportive Analysis like recruitment websites or job sites also play a role in simplifying the job curriculum vitae process. Such websites have facilities where prospective candidates can upload their resumes and apply for jobs suited to them. These sites also make it possible for recruiters and companies to post their staffing requirements and view profiles of interested candidates.**

Keyword—Candidate, Shortlist, CV, Recruiter, Human Resource, Ranking, Aptitude, Profile, Resume.

I. INTRODUCTION

Human Resource (HR) is an operation in companies designed to maximize employee performance in order to meet the employer's strategic goals and objectives. More precisely, HR focuses on management of people within companies, emphasizing on policies and systems.In short, HR is the process of recruiting, selecting employees, providing proper orientation and induction, imparting proper training and developing skills.HR supportive Analysis using cv also includes admin and candidate assessment like Applying for the job, Scanning the CVs, Forwarding CVs to the company, Adding the aptitude question, Uploading the CVs, Updating the CVs, Notifying the candidate, Giving aptitude etc.Recruitment, or the process of selecting the right candidates from a vast pool of candidates, has always been a very fundamental issue as far as employers are concerned. Traditional techniques included conducting personality and other technical eligibility evaluation tests, interviews and group discussions. Of late, with the advent of technology around us, we have seen a shift in the way recruitments are being conducted. Online recruit systems are knowledge management systems to employ job seekers, by utilizing the scope of Web 2.0 and other social networking sites. The use of social media as a recruitment tool presents newer opportunities and challenges for employers. Social media evidently offers speed, efficiency and the ability to target and attract specific, particularly pertinent job seekers from the vast pool of candidates.

RELATED WORK

HR supportive analysis using CV is a web-based system in which the candidate will have to register in the system and login using the login credentials. The candidate will initiate the process by filing the form which will have different fields like, the number of papers published, the journals in which the papers are published, experience at the previous organization, marks in 10th, 12th and graduation, technical skills, participation in extra-curricular activities. Based on the information provided in the form a score will be generated by taking into consideration the level of expertise in particular field and by mapping the criterias entered by the HR recruiter the best fitting candidates will receive email confirmation.

FLAWS IN CURRENT SYSTEM

The existing system does not take into consideration different factors like the number of papers published, the journals in which the papers are published, work experience at previous organization, whether it is relevant or not to the requirements of the particular position. The wrong candidate gets selected because many times the system selects a particular candidate with more number of work experience which is actually not relevant to the field of the job he/she is applying for.

II. IV. FEATURES

[1] System uses the score generated for each candidate for finding the best suited candidate.

[2]For score generation different factors are taken into consideration like, the number of papers published, the journals in which the papers are published, experience at the previous organization, marks in 10th, 12th and graduation, technical skills, participation in extra-curricular activities..

[3] Shortlisted candidates will receive the notification by email.

[4] HR recruiter will be able to add and edit the questions in aptitude test.

[5] HR recruiter will have access to all the CVs uploaded by different candidates.

[6] HR recruiter will able to enter the requirements of the particular position and on the basis of the requirements of the HR recruiter best suited candidates will be shortlisted.

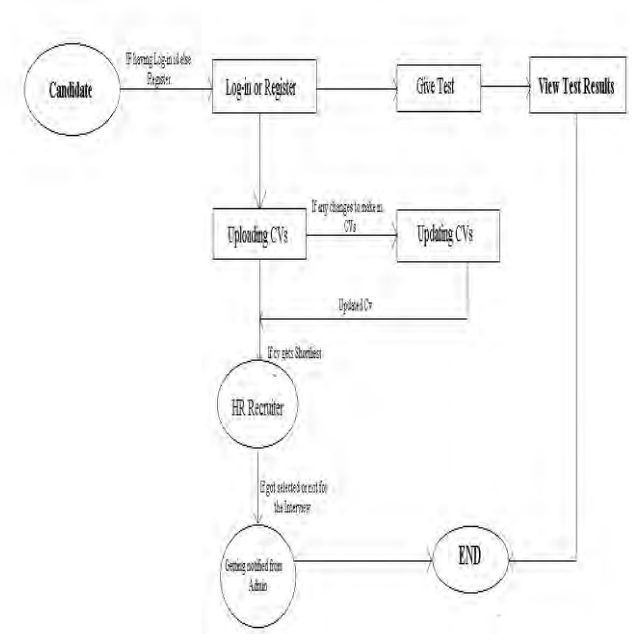
[7]Based on the performance of the candidate in aptitude test a bra graph will be generated highlighting the candidates performance in particular topics of the test.

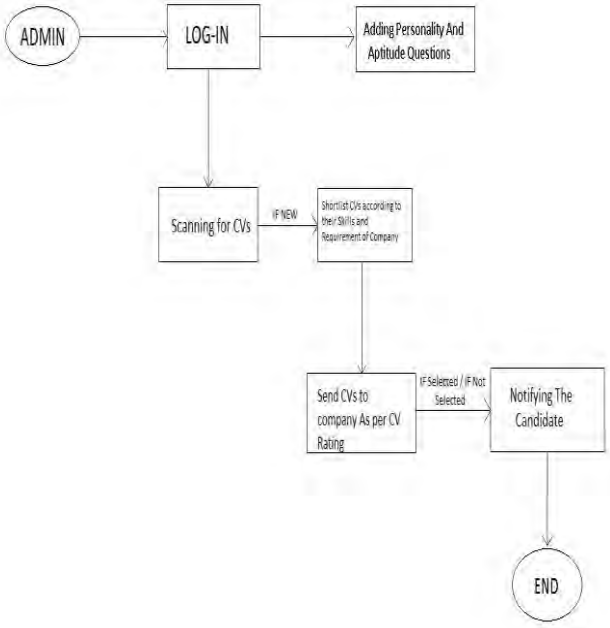
BLOCK DIAGRAM OF THE PROPOSED SYSTEM

First candidate will log -in to the system using his/her unique user id and password after logging-in to the account,the candidate can update and upload their revised resume or CV.After logging-in personal account they can give practise tests of aptitude and technical tests and analyse their performance and improve accordingly.Then the recruiter will get the notification.

The Admin can review the CVs of the candidates by loggingin to the system. After the candidate has updated the cv and uploaded into the system the recruiter will get a notification of the changes done by the candidate. Then the shortlisted list of the candidates on basis of their abilities will be mailed to the recruiter and at the same

time the candidate will get a notification of the same. The recruiter will get a score report of the tests given by the candidate.





DESIGN AND IMPLEMENTATION

This system will help the HR department to easily shortlist the candidate based on the CV ranking policy. This system will focus not only in qualification and in experience but also focuses on other important aspects, which are required for particular job position. This system will help the human resource department to select right candidate for particular job profile, which in turn provide expert workforce for the organization.

fid	first_name	middle_name	last_name	gender	contact	work_status
1	Bhalo	Singh	talwar	M	55555555	W
1	t	t	t	m	2147483647	w
2	e	e	e	f	2147483647	w
2	e	e	e	f	2147483647	w
1	tt	ttt	ttt	m	2147483647	w
1	tt	ttt	ttt	m	2147483647	w
1	name	name	name	m	2147483647	w

For candidate to be shortlisted different factors will be considered like the number of papers published, the journals in which the papers are published, experience at the previous organization, marks in 10th, 12th and graduation, technical skills, participation in extra-curricular activities. For each of these above factors certain points are assigned, which is shown in the diagram below. According to these parameters score will be generated for each candidate and on the basis of the score generated, the highest ranking candidate will come at first position and list will be generated. When the HR recruiter will enter the skills required for the particular position all the candidates with those skill sets will be shortlisted and they will receive email confirmation.

The HR recruiter will be able to add and edit the aptitude test questions and candidates will give the aptitude test, the score of that aptitude test will be available for the recruiter so that he/she can see who are the top rankers in aptitude test.

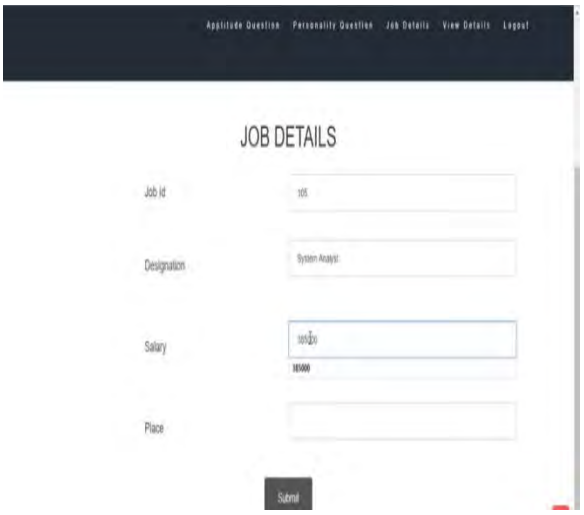
If the recruiter wishes to see the particular CV of the candidate he/she will be able to do so by clicking on the CV of the particular candidate. Different technologies are used for the development of this web-based system like C#, ASP.net, MySQL for maintaining the database.

Parameters	25 points	50 points	75 points	100 points
Number of papers published	1	2	3	More than 3
Type of journal in which paper is published	IJSER	IOSR	IJSR	IEEE
Experience in months at the previous organization	6 months or less	1 year	2 years	More than 2 years
Marks in 10th	Less than 60%	60% to 75%	75% to 90%	More than 90%
Marks in 12th	Less than 60%	60% to 75%	75% to 90%	More than 90%
Average % in graduation	Less than 60%	60% to 75%	75% to 90%	More than 90%
Technical skills known	C	C, JAVA	C, JAVA, HTML, CSS, PHP	C, JAVA, HTML, CSS, PHP, PYTHON
Participation in sports	School level	District level	State level	National leve.

III. VI. FUTURE WORK

The system collects information like test results of candidates,Uploaded CVs.The candidate can monitor his progress by giving aptitude test. This project focuses on single user environment which can be extended to multiuser environment. The Admin will notify to the individual candidate about shortlisting of the CV as well as the documents needed for the interview. From the evaluation and the result obtained from analysis the system is better forcandidates, who are in search for the job as it makes easier to upload cv and helps in improving aptitude skills.

This model can be later used in Private Company's and HR Consultants where candidate could get identify the different forms of diseases much efficiently and swiftly. The application can be used to replace doctors if enough processing elements and data set are used on a larger scale.



IV. VII. CONCLUSION

Thus we have designed and developed a system for short listing CV's on the basis of previous work experience, number of papers published, academic qualification, extra and co-curricular activities. This system will reduce the work of HR recruiter and save the valuable time of HR recruiter which they usually spend analyzing the CV's of the candidate and then shortlist them.

V. ACKNOWLEDGEMENT

We sincerely thank our guide Dr. Rekha Sharma for her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

We thank the HOD, Dr. Sheetal Rathi, Dean Academic, Dr. R.R Sedamkar, the Principal, Dr. B. K. Mishra and the college management for their support.

REFERENCES

VI. [1] Vishnu M Menon ,Rahulnath H A , A noval Approach to Evalutate and Rank candidates in A Recruitment Process by Estimating Emotional Intelligences through Social Media Data, 2016

VII. [2] QihangZhang ,xiaodong Xu , The Application of the Job Characteristics Model in E-HR System , 2010

VIII. [3] Wei Zhou , Xiao-Qian Zhu , Li-Xuan Ye , Design and Realization of Distinctive Data Inerface Based on SAP HR System, 2011

IX. [4]EvanthiaFaliagka, Kostas Ramantas, AthanasiosTsakalidis,"Application of machine learning algorithms to an online recruitmentsystem", ICIW 2012: The Seventh International Conference on Internet and Web Applications and Services.

X. [5] Andrea Broughton, Beth Foley, Stefanie Ledermaier, Annette Cox, "The use of social media in the recruitment process" The Institute for Employment Studies, 2013.

Plagiarism Detection:A Better Approach Towards Uniqueness of Data

[1]Dhekale Prasad,[2]Jamdade Pradip,[3]Rahinj Swapnil [4] Shantanu Rai [5]S V Athawal
[1,2,3,4,5]*Computer Department*
AISSMSCOE
Pune,India
[1]prasaddhekale911@gmail.com[2]pradipjamdade20@gmail.com[3]swapnilrahinj200@gmail.com[4]shantanurai1997@gmail.com[5]svathawale@gmail.com

***Abstract*—The major evaluation in technology is providing a way to access the big amount of data which arises the problem of plagiarism. In today's world plagiarism is increasing exponentially. There are so many tools available in the market to detect the plagiarism. They have their own advantages and disadvantages. Main problem with these softwares are its cost and word limit. As they are very expensive due to which most of the people can't afford these softwares. In this paper we are going to address different types of plagiarism tools, their issues and solutions. This work mainly focuses on increasing word limit and reducing its cost to detect plagiarism in text and CSV files efficiently and give more accurate result.**

Keywords:- Plagiarism, Tokenization, Stop word removal, CSV

I. INTRODUCTION

A wide vary of plagiarism and similarity detection systems was developed in recent years. Basically their task is to find similarities in files, which might indicate either plagiarism (in case of students' works) or code duplication (in case of a computer code project analysis). The quality of the system is primarily determined by the tactic of similarity calculation. The same method usually encompasses a high influence on the speed of the detection. Typically, the quicker is detection routine, the less precise results it provides. In our work we have a tendency to study the possibility of exploitation the combined approach: the faster (and less precise) algorithm performs the initial selection of suspicious files, then the additional reliable (and slower) routine calculates similarity ratios for file pairs. It should be noted that we have a tendency to focus on a thus referred to as "offline" (or "hermetic") plagiarism detection, which deals with analysis of local file collections. "Online" plagiarism detection (search for similar documents within the Internet) is expounded heaps of to information retrieval, thus coverage will turn out to be much more important than exactitude.

II. EXISTING SYSTEM

Plagiarism and similarity detection software system is well-known in universities for years. Despite the range of methods and approaches used in plagiarism detection, the typical trade-off between the speed and also the reliability of the algorithmic program still remains. We introduce a new ballroom dancing approach to plagiarism detection that mixes high recursive performance and also the quality of pairwise file comparison. Our system uses fast detection methodology to select suspicious files only, so invokes precise (and slower) algorithms to get reliable results. We show that the planned method does not noticeably reduce the quality of the pairwise comparison mechanism while providing better speed characteristics.

III. PROPOSED SYSTEM

Theproposed systemwill filer files very fast and quiteaccurate. It will discardsignificantly the files that cannotMach criteria of plagiarism.

A. Existing System Vs Proposed System

Parameter	Existing System	Proposed System
Word Limit	Low	High
Cost	High	Low
Speed	Moderate	High
Efficiency	Moderate	High
Ease of use	High	High
Security	Moderate	High
Diversity of text format	Moderate	Moderate

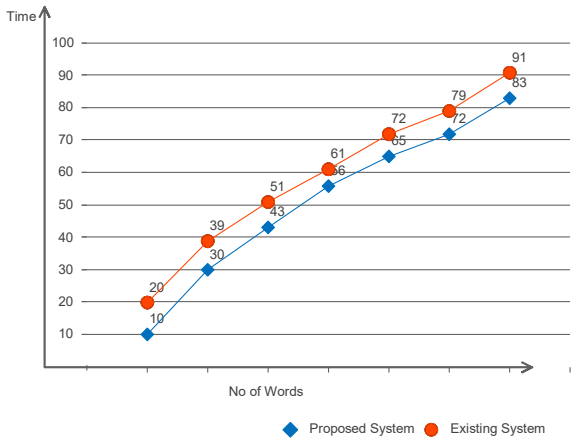


Fig 1. Compression Graph

IV. PROBLEM STATEMENT

A lot of copying of original documents are taking place, whether it's at an institutional level or at a development level. The main reason for a text similarity tool is to determine the similarity between submitted documents. Any person who wants to check the originality of a document can benefit from the system as well as the system can be used in search engines. Any fraudulent activity detect by the system. Natural language processing and machine learning have become such a wide field for building new techniques and systems which help a computer analyze various inputs and then give the desired outputs. Automated results decrease the workload of people.

V. SYSTEM REQUIREMENTS

A. Analysis model: SDLC model to be applied

An effective System Development Life Cycle (SDLC) should result in a high quality system that meets customer expectations, reaches completion within time and cost evaluations, and works effectively and expeditiously in the current and planned Information Technology infrastructure. System Development Life Cycle (SDLC) could be a conceptual model which incorporates policies and procedures for developing or altering systems throughout their life cycles. SDLC is used by Analysts to develop associate degree information system. SDLC includes the following activities:

- Requirements
- Design
- Implementation
- Testing
- Deployment
- Operations
- Maintenance

B. Phases of SDLC:

Systems Development Life Cycle could be a systematic approach which explicitly breaks down the work into phases that are required to implement either new or modified system.

a. Scope

To achieve unique result. To provide better accuracy without changing the meaning of the text.

VI. MATHEMATICAL MODEL

I = Input
I = F {f1,f2, f3,….fn}
F = file
Database file = { Df1, Df2, Df3……Dfn }
Stop words = S {s1, s2, s3,…sn }
Remove the content
S ϵ F
Generate token = T
T = {t1, t2, t3,…tn }
Perform lexical analysis = LA
Form chunks of file
C = (c1, c2, c3,… cn)
Perform chunks matching on Df
C ≡ Df
Find semantic matching for total no. of words & synonyms matched with documents text
Semantic ≡ w
Synonym ≡ c
words → w
Count of synonym → c
Output = O
Find PR (Percentage result of document)
PR = (w + c / total no. of words in F) * 100
O = PR

VII. USER MODULE

- Registration
- Login

-Upload text files of payment slip and back statement on system (mention the type of files i.e payslip/bank statement/offer letter)
-Get matching and unique percentage result for document.
-Get result whether provided documents are fraudulent or not.

VIII. SYSTEM MODULE

-First find out the company details and name from the provided document using text mining.
-Check is database that the provided company name is fake or real (if fake give result to user that documents are fraudulent, if real company name then proceed further steps)
-Remove stop words from the user file and the files to be compared from database.
-Perform Dependency parsing (Tokenization – generate tokens, perform lexical analysis).
-Form chunks of user file (form small groups of sentences)
-Perform chunk matching and final semantic matching for total no of words and synonyms matched with other documents texts.
-Find percentage result of unique andplagiarized data from user document.

-The documents with high percentage for plagiarized data, calculated attribute based matching of data for those documents.
-In attribute based matching find out the words and values for transaction id, amount, date, time, account number, cheque no. (If values are matching for these attributes then show result that documents are fraudulent, if values are not matching for these attributes then show result that documents are real and valid).

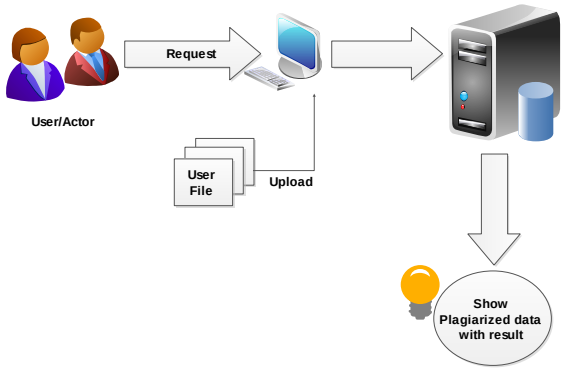


Fig 2. Process Flow Diagram

Algorithm:

- Text mining
- Tokenization
- Lexical analyzer

- SIMILARITY DETECTION SYSTEM

Modern plagiarism detection systems usually implementcertain content-comparison techniques. The most popularapproaches include heuristic string tiling (finding the jointcoverage for a pair of files), and parse trees. Usually these algorithms work for filepairs, so the comparison routine should be called for eachpossible file pair found in the input collection. It means thatO(f(n)N2) time is required to perform the detection. Here N isthe number of files in the collection, and f(n) is the timeneeded to compare two files of length n.

Fast Plagiarism Detection System (FPDS) tries toimprove the algorithmic performance of plagiarism detection by utilizing a special indexed data structure (suffix array) to store input collection files. A special heuristic search routine isused to compare any given file against the whole collection atonce. The complexity of the complete detection procedure isO(nNγ + N2), where N is the total number of files, n is the average file length, and γ is a special finely-tunable constant(γ = Ω(log nN) should be used for the best performance). In terms of quality, FPDS shows the results, which are close tothe ones, provided by other content-comparison systems Comparison.

- TOKENIZATION

Tokenization is a commonly-used technique that fightsagainst renaming variables and changing loop types in computer programs. Simple tokenization algorithms substitutethe elements of program code with single tokens. For example,all identifiers can be substituted with <IDT>, and all valueswith <VALUE> tokens.If another file has a variable appearing in the same context, itis treated as identical.All modern plagiarism detection systems, aimed atprogram code analysis, implement some tokenization-likeprocedure.

CONCLUSION

We have developed a reliable plagiarismdetection system by combining older Plague and FPDS projects. We have showed that the use of FPDS as a filter does not noticeably reduce the reliability of Plague, but providesmuch better algorithmic performance than naïve file-filecomparison techniques. The results were tested on a largeenough set of Java source files.As a main result, we can expect that the proposed systemwill filer files very fast and quite accurate. It will discardsignificantly the files that cannot Mach criteria of plagiarism.The FPDS part is responsible for this operation. The number of files sent for a detailed investigation to the Plague part issmall enough.

REFERENCES

[1]. Mozgovoy, M., Desktop tools for online plagiarism detection in computer programs, Informatics in Education, vol. 5(1), 2006, pp. 97- 112.
[2]. Lancaster, T., Culwin, F., Using freely available tools to produce a partially automated plagiarism detection process, Proc. of the 21st ASCILITE Conference, 2004, pp. 520-529.
[3]. Culwin, F., MacLeod, A., Lancaster, T., Source code plagiarism in UK HE computing schools, issues, attitudes and tools, South Bank University Technical Report, 2001.
[4]. Wiedermeier, P.D., Preventing plagiarism in computer literacy courses, Journal of Computing Sciences in Colleges, vol. 17(4), 2002, pp. 154-163.
[5]. Faidhi, J.A.W., Robinson, S.K., An empirical approach for detecting program similarity within a university programming environment, Computers and Education, vol. 11(1),1987, pp. 11-19.
[6]. Verco, K.L., Wise, M.J., Plagiarism la mode: a comparison of automated systems for detecting suspected plagiarism, The Computer Journal, vol. 39(9),1996, pp. 741-750.
[7]. Wise, M.J., YAP3: improved detection of similarities in computer program and other texts, Proc. of SIGCSE96 Technical Symposium, 1996, pp. 130-134.
[8]. Prechelt, L., Malpohl G., Philippsen, M., Finding plagiarisms among a set of programs with JPlag, Journal of Universal Computer Science, vol. 8(11), pp. 1016-1038.
[9]. Gitchell, D., Tran, N., Sim: a utility for detecting similarity in computer programs, Proc. of the 30th SIGCSE Technical Symposium on Computer Science Education, 1999, pp. 266-270.
[10]. Belkhouche, B., Nix, A., Hassell, J., Plagiarism detection in software designs, Proc. of the 42nd Annual Southeast Regional Conference, 2004, pp. 207-211.
[11]. Mozgovoy, M., Fredriksson, K., White, D., Joy, M., Sutinen, E., Fast plagiarism detection system, Lecture Notes in Computer Science, vol. 3772, 2005, pp. 267-270.
[12]. Joy, M., Luck, M., Plagiarism in programming assignments, IEEE Transactions on Education, vol. 42(2), 1999, pp. 129-133.
[13]. Baker, B. S., On ending duplication and near duplication in large software systems, Proc. of the 2nd IEEE Working Conference on Reverse Engineering, 1995, pp. 86-95.
[14]. Ahtiainen, A., Surakka, S., Rahikainen, M., Plaggie: GNU licensed source code plagiarism detection engine for Java exercises, Proc. of the 6th Baltic Sea Conference on Computing Education Research,2006, Uppsala, Sweden.
[15]. Wise, M.J., Running Rabin-Karp matching and greedy string tiling, Basser Department of Computer Science Technical Report, 1994.

[16]. Wessa, P., Free statistics software, Once for Research Development and Education, v. 1.1.17, URL http://www.wessa.net

Stereo Image Depth Mapping using FPGA

Yogesh Ghadigaonkar, Prof. Sunil Somani, Prof. Kalpana Bapat

Dept. of E&TC MIT COE , Pune, India

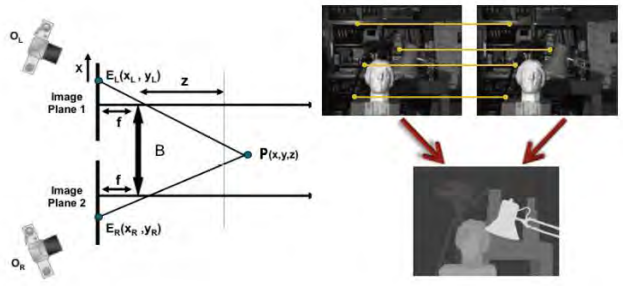
yog2ghadi@gmail.com, sunil.somani@mitcoe.edu.in, kalpana.bapat@mitcoe.edu.in

Abstract—**The human being has the ability to recognize the depth of an object in front of their eyes. They combine two images known as stereo images and creates a single image which helps to identify the distance of that object from eyes. The concept is known as stereo image depth mapping (disparity image) which is used for 3D image reconstruction. Disparity image generates from the stereo images is possible using MATLAB tools and embedded platform. The disadvantage of using this type of platform is a delay for disparity generation and low frame rate. To overcome this issue is a hardware platform like FPGA having high clock frequency, parallel processing and a large number of I/O's for external connection.**

Keywords—SAD (Sum of Absolute Difference), Disparity Map, FPGA.

I. INTRODUCTION

Computation of stereovision involves sensing of depth in the 3D form of a particular scene from multiple views of the same scene. Stereo matching obtains the depth information from pair of an images in which corresponding pixels have displacement [1] [2]. The disparity is called as displacement in a number of pixels. The establishment of correspondence between a pair of images to generate depth is known as stereo matching. Various stereo matching algorithms can be classified as area-based and pixel-based, feature-based.



(a) (b)

Fig. 1. (a) Stereo vision camera setup, (b) Disparity map from stereo image

The general process of extracting depth information from the stereo image is pixel shifting from corresponding images and calculate their disparity [see also Fig. 1(b) [6]. In traditional stereo image system, two cameras are places almost in the same horizontal line in order to obtain scene image, similar to human eye position. The camera setup in Fig. 1(a), the distance to a point P (x, y, z) can be estimated. Two points denoted as ER and EL on the same horizontal line. The point EL of the left image has shifted some distance as point ER of the right image in same horizontal line. The shifted point decides the disparity of that point or pixel. The disparity image having depth information of stereo image used to regenerate 3D image and video in robotic surveillance etc.

II. LITERATURE SURVEY

Stereo matching is the fundamental algorithm to estimate the 3-D depth from a pair of binocular 2-D stereo images. The depth is estimated from the disparity between a point in the right image and its corresponding point in the left image [6],[7],[8]. Finding a point in the target image which corresponds to a given point in the reference image is referred to as stereo matching. Stereo matching methods can be classified into three major classes, (1) pixel or window-based search method, (2) semi-global method such as the dynamic time warp (DTW) algorithm, and (3) global method based on the graph cut algorithm.

Window Based Stereo Matching Algorithm

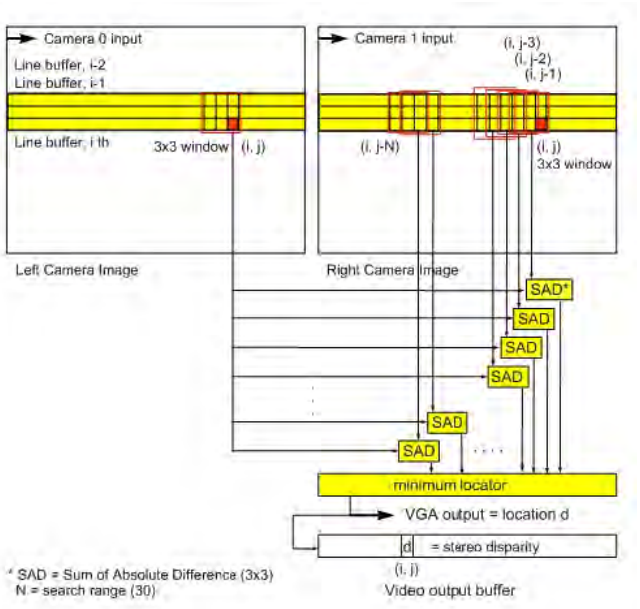


Fig. 2. FPGA architecture of stereo matching to calculate disparity 'd'

The window base algorithm of stereo matching is to search the best matching pixel block along a raster scan line in the target image (left image) to a given pixel block of the reference image (right image) of the same raster line6. Mathematically, the best match is

defined as the minimum of SAD given in Equation. (1). The sum of absolute differences (SAD) between two windows at (i, j − k) of the left image IL and at (i, j) the right image IR is calculated for the window size of 3×3.

$$s(i, j) = \min_{k \in R} r(i, j, k) \quad (1)$$

In Fig.2. left image from camera 0 and right image from camera 1 having same N x N resolution in size. The array of 3 x 3 window size of left image is scan on same raster line of right image for window matching. For each window matching through SAD module for minimum Sum of Absolute Difference (SAD) to minimum locator.

III. PROPOSED SYSTEM

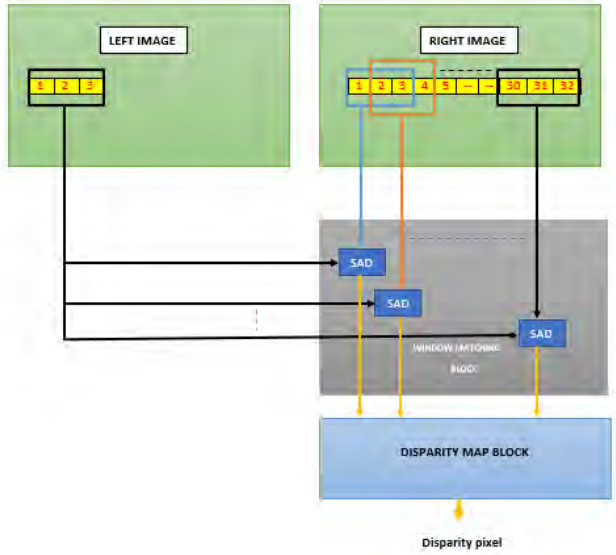


Fig. 3. FPGA architecture of calculating disparity pixel value

The window-based stereo matching is the best to find a matching pixel along a horizontal scale line. The target image (left image) is scanned horizontally on the same line along with a reference image (right image) for window matching. Mathematically the best way to find window matching is by calculating a minimum SAD value. This algorithm calculates a minimum Sum of Absolute Difference (SAD) value and gives the best matching window. The common way to find a window matching algorithm using N x N window size i.e. (N = 3,5,7,11). This matching window algorithm required a number of the clock cycle to load pixel value from RAM. If the size of the window increases will increase the number of a clock cycle for execution and increases delay per frame. This issue can be solved by using 1xN window size i.e. (N = 3,5,7,11).

Fig.3. shows that, a left image having 1x3 window size and a right image having 1x32 window size for window matching. Basically, from research papers, window matching is executed serially one by one which takes lots of clock cycle for execution. This paper gives the parallel window matching algorithm that reduces clock cycle required for window matching. On each clock cycle pixel value of both windows is shifted left side and that value is calculated for window matching. The disparity map block compares minimum SAD value and gives disparity pixel value for a respective matching window.

Hardware Acceleration of FPGA

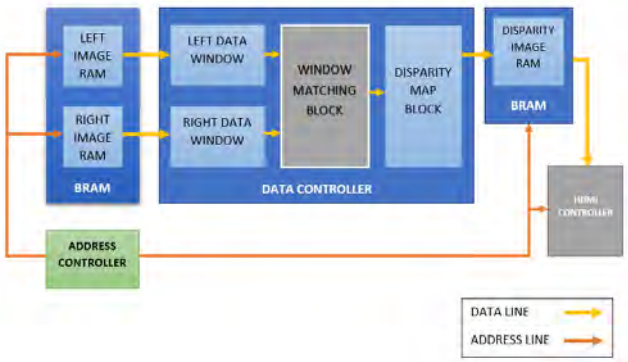


Fig. 4. Fig. 4. Architecture of proposed system

A hardware architecture is implemented on an FPGA in order to accelerate the calculation of disparity pixel value. In our implementation, the hardware platform is the Xilinx ZYBO Z7-20 development board. The board having on board SoC FPGA and ARM with 32 I/O interfaces with dedicated HDMI port for video processing.

The architecture is designed for parallel processing to calculate disparity value. The speed of the hardware system depends on a clock frequency of FPGA. The clock frequency of ZYBO Z7-20 is 125MHz and using PLA block in SoC it can generate up to 250MHz. At each clock cycle, a one-pixel value is out from RAM and as an input to the left or right data window which is inside data controller block. In an implemented design does not require to load the same pixel again and again. Therefore, calculation of disparity value takes less clock cycle than other cases.

Procedure methodology

The proposed method required a stereo image for a disparity image. In ZYBO Z7-20 development board having 4MB BRAM memory to store two images known as stereo images. For this project, we used 256 x 256 size images as a stereo image. The image which is in the format of (.jpg or .png) are converted in text format (.txt) by MATLAB tools. The image must be resized in the size of 256 x 256 with each pixel size is 8 bits.

The address controller blocks in Fig. 3, addressing the left image and right image stored in BRAM. The pixel data of both images is out serially with respective addressing position. Left data window and right data window in data controller block are works as a shift register. Its input serially and output parallelly and synch with the clock. Window matching block is scanning 1 x 3 window of the left data window parallelly to right data window. The final block is disparity map block compare minimum SAD value and generates disparity pixel value with corresponding pixel address. HDMI controller block helps to display disparity image stored in disparity image RAM on a monitor.

IV. RESULT

The FPGA used for implementation was Xilinx ZYBO Z7-20 development board, which is useful for our project. The proposed system designed to generate a disparity image from stereo images, each having 256 x 256 in size. Table no.1 shows that the maximum frame rate of a various platform. From the table no.1, we can conclude that to work on FPGA platform is better than other two platforms.

TABLE I. COMPARISON OF FRAME RATE OF VARIOUS PLATFORM USE FOR DISPARITY MAP

System used for disparity generation	Maximum frame rate
MATLAB tools	20 frames per sec
Embedded system	30 frames per sec
FPGA	50-60 frames per sec

Simulation Result

A simulation waveform of a disparity pixel as shown in Fig.5 and Fig.6. Each disparity clock generates new disparity pixel value. On positive edge of a disparity clock, it generates disparity address and a disparity pixel value.

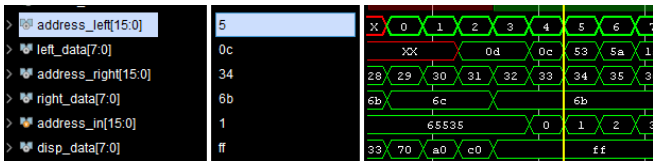


Fig. 5. Simulation waveform (first pixel from first horizontal line)

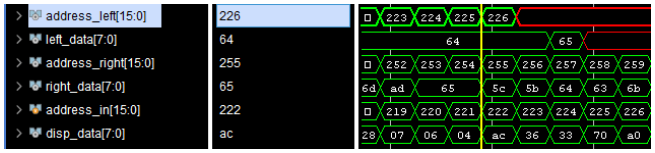


Fig. 6. Simulation waveform (last pixel from first horizontal line)

Due to window based matching algorithm size of the disparity image is reduced from 256x256 to 256x227 in resolution. On each disparity clock, it calculates disparity pixel value and address of that disparity pixel. All disparity pixel stored in respective disparity address in BRAM. For this proposed system to generate one disparity frame using N X N size required (N*N) + 3 clock cycles.

V. CONCLUTION

The window-based stereo matching algorithm implemented in the ZYBO Z7-20 FPGA with two images namely a left image and a right image (i.e. stereo images) stored in BRAM. The disparity image gets from the stereo image using FPGA platform in real-time with 50-60 frame per seconds. Using FPGA platform to generate the disparity map overcome the issue like low frame rate and execution delay.

VI. REFERENCES

[1] Qian, Zheng, and Kunio Takaya. "Dense stereo disparity maps by dynamic time warp with sparse features and FPGA acceleration." *Electrical and Computer Engineering (CCECE), 2011 24th Canadian Conference on*. IEEE, 2011.

[2] Aguilar-González, Abiel, et al. "An FPGA correlation-edge distance approach for disparity map." Electronics, Communications and Computers (CONIELECOMP), 2015 International Conference on. IEEE, 2015.

[3] Sledevič, Tomyslav. "The evaluation of configurable core for disparity map computation based on FPGA." Advances in Information, Electronic and Electrical Engineering (AIEEE), 2016 IEEE 4th Workshop on. IEEE, 2016.

[4] Wang, Yingqian, et al. "Disparity Estimation for Camera Arrays Using Reliability Guided Disparity Propagation." IEEE Access6 (2018): 21840-21849.

[5] Takaya, Kunio, and Zheng Qian. "FPGA based stereo vision system to display disparity map in realtime." Information science and applications (ICISA), 2012 International Conference on. IEEE, 2012.

[6] Michailidis, Georgios-Tsampikos, Renato Pajarola, and Ioannis Andreadis. "High performance stereo system for dense 3-D reconstruction." IEEE Transactions on Circuits and Systems for Video Technology 24.6 (2014): 929-941.

[7] Zitnick, C. Lawrence, and Takeo Kanade. "A cooperative algorithm for stereo matching and occlusion detection." IEEE Transactions on pattern analysis and machine intelligence22.7 (2000): 675-684.

Hirschmuller, Heiko, and Daniel Scharstein. "Evaluation of stereo matching costs on images with radiometric differences." IEEE transactions on pattern analysis and machine intelligence 31.9 (2009): 1582-1599.

Solar Based Net Metering Using Arduino Uno

Vijay M. Birari, Jayashree L. Gadhave, Kaveri B. Handore, Devashree M. Jadhav

Electronics & Telecommunication NDMVPS'S KBTCOE, NASHIK
hod.etc@kbtcoe.org, jgadhave0@gmail.com, kaverihandore@gmail.com, devashreej61@gmail.com

***Abstract*—Solar energy is emerged as a most important source of renewable energy and are now widely used for industrial and domestic applications. Solar energy generated through solar panel which is based on a solar collector, they are designed to collect the sun's energy and convert it into electrical power. The proposed system utilizes solar energy at home. It will generate excess power than that required for home utilities and the power not required at home is send to the service provider. It also provides credit which will lead to zero electricity bill. The proposed system ensures the energy generated at home through solar energy does not go to waste. The proposed system uses solar panel to collect solar energy which is stored in battery in DC form of electric energy which further is connected to inverter to convert DC signal into AC signal. The AC signal is given to bidirectional transformer which controls supply switching either to home or the supplier according to the utility. Arduino UNO is controller used and it displays utilized and unutilized power on LCD.**

Keywords- Solar panel, Battery, Inverter, Bidirectional transformer, Arduino UNO, LCD.

I. INTRODUCTION

A. Necessity

As electricity demand increases, with supply depending largely on fossil fuels plus some hydropower and then nuclear energy, concerns arose about carbon-dioxide emissions leading to global warming.

The fuels are on verge of extinction. Thermal engines requires huge amount of lubricating oil that is very expensive. Nuclear thermal power plant demands excessive amount of water for cooling purposes. This type of power station ultimately responsible for raise in sea water level. Attention again turned to the huge sources of energy surging around us in nature-sun.

B. Need

Today we are well advanced in meeting the challenge of generating electricity through solar energy. Net metering allows customer who generate their own electricity from solar power to feed electricity and the unutilized power back to supplier. The large magnitude of solar energy available makes it a highly appealing sources of electricity. No greenhouse gas emissions are released into the atmosphere when solar panels are used to create electricity. Solar net metering also provides revenue generation as in this process credits are gained at the supplier's end.

C. Motivation

India along with the world is facing 'Energy Crisis'. Fossil fuels are falling short as there is increase in demand for electricity. There is a significant gap in the demand and supply for electricity. Day by day as our country progresses towards development, this gap is increasing and tackling this situation is very important to continue our country's ascending path. The best option is utilizing the solar energy which is available in abundance. Government of India announced a policy to provide incentives on power generated from solar energy and fed to the power supplier.

To support this background, it is essential that we acquire capability to design, develop and install solar plants.

D. Objectives

- To use solar energy at home.
- To generate revenue by returning the excess unutilized power.

II. IMPLEMENTATION

A. Block Diagram:

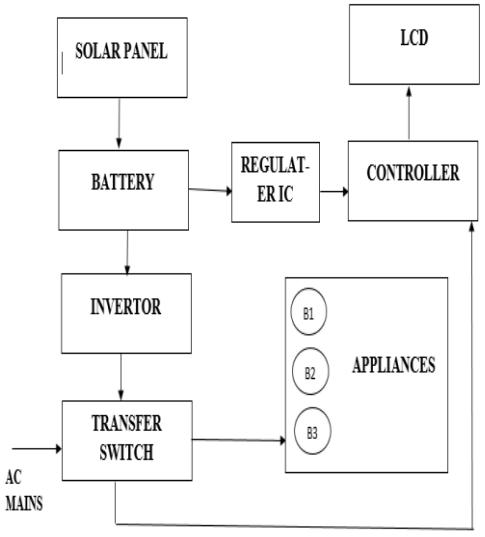


Fig 1: Block diagram of System

Proposed System is the block diagram of Solar Based Net Metering using Arduino. In this solar panels works to collect solar energy. Net meter work as bidirectional meter which takes data of utilized and unutilized energy. Battery is use to save energy and use energy as our requirement. The generated energy is first fed to load that is the home appliances. The unutilized energy will be fed to line.

Solar panel: Solar panel is used to convert the light rays of sun into DC or electric form of energy. This system absorbs sunlight and converts it into electricity that we can use. Every second, the sun emits approximately 42 trillion kilocalories of energy to the Earth.



Fig 2: Solar Panel of Proposed System

In general, most homes will produce excess electricity in the summer months and will use more electricity from the grid in the winter. Because these variations in production are fairly predictable, your utility won't send you a monthly check when you produce more than you need.

Instead, you will build up extra credits during the summer months so that you can draw from them at night and during the winter months when you need them. With the right design, your system can generate enough power to match your total electricity use for a year.

Battery: Battery gets input from solar in DC form and charge according to rating and gives the output to regulated IC and inverter which is in AC form. Battery is for storage and also used simultaneously for domestic use. This stored energy is used when solar is off or there is poor sunlight.

Inverter: Inverter is used to convert the DC voltage into AC signal. In proposed system it

will not generate the frequency it will only convert the signal. This is main part of proposed system. This is a power inverter which is used to convert DC electricity into AC electricity with an ability to synchronize to interface with a utility line. It is used in the converting DC sources such as solar panels into AC for tying with the grid.

For Proposed system the controller is Arduinouno microcontroller which gives the information and takes control over the entire input output system with the reference input provided to it and also display it on the LCD for the ease of understanding.

Transfer switch: This switch is main part of our system It connect the power switch either solar or power supplier as per the requirement.

Transfer switch: This switch is main part of our system I connect the power switch either solar or MSEB supply as per it required. It will not give power back to MSEB when it will shut off, because some work is ongoing on MSEB cables this will harmful. As fig. 3 shows transfer switch which has AC1 and AC2, in this AC1 is for main supply from MSCB and AC2 is used for solar panel energy utilization.

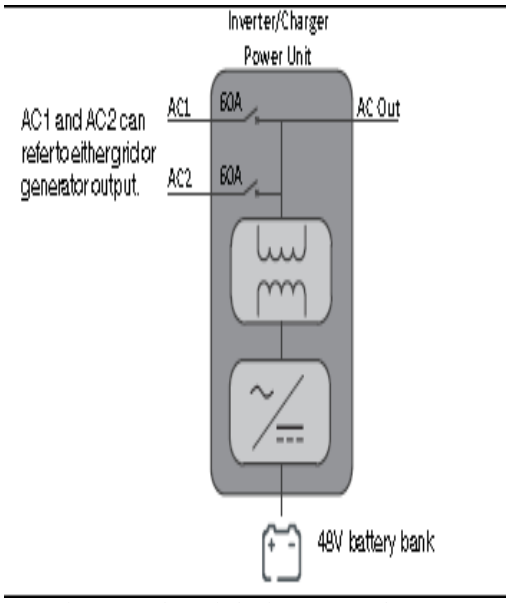


Fig 3: Transfer Switch of our Proposed System

Controller: Our controller is Arduinouno microcontroller which gives the information in display format to LCD. In this we display the home appliances connected to solar panel or power supplier and also display what amount of power is used or fed back to the grid.

Regulator IC: Regulator IC which converts the DC source into low dc power to connect the microcontroller. In this we can use voltage regulator or zener diode. Net metering mechanism provides usage of a bidirectional transformer which has facility to record both import and export energy values. In net metering, prime focus is on utilizing self produced electricity by renewable energy sources and excess or surplus is sold to utilities. Solar photovoltaic system is used significantly in net metering. When your solar power system generates more electricity than you use over the course of a month, your utility bill will receive a credit based on the net number of kilowatt-hours you gave back to the grid. If you produce less electricity than that you use in a given month, you must buy electricity from your utility to make up the difference. In these instances, you would pay for the electricity you use, minus any excess electricity your solar panels generated.

LCD: LCD (Liquid Crystal Display) screen is an electronic display module and find a wide range of applications. A 16x2 LCD display is very basic module and is very commonly used in various devices and circuits. These modules are preferred over seven segments and other multi segment LEDs. The reasons being: LCDs are economical; easily programmable; have no limitation of displaying special and even custom characters (unlike in seven segments), animations and so on. A 16x2 LCD means it can display 16 characters per line and there are 2 such lines. In this LCD each character is displayed in 5x7 pixel matrix. This LCD has two registers, namely, Command and Data. The command register stores the command instructions given to the LCD. A command is an instruction given to LCD to do a predefined task like initializing it, clearing its screen, setting

the cursor position, controlling display etc. The data register stores the data to be displayed on the LCD. The data is the ASCII value of the character to be displayed on the LCD.

III. CONCLUSION

In this project, we try to show the glimpse of the immense potential that Renewable energy resources hold. Some advancement which can significantly boost the practicality and effectiveness of the solution provided by our project are: Research can be done to increase the power output so as to make it more useful and dependable. Bulk production of the Arduino based solar power Inverter will further lower the cost. If implemented on a bigger scale, it can become quite economical. By this project we can establish effective way of generating power. All this will be done with low man power. There will be collection of money after using the power because people they themselves will produce the power and the excess power will be sold.

REFERENCES

[1] RuchikaThukral, Gulshan Kumar, Ankit Gupta.,2015., Microcontroller Based Solar Power Inverters, IJEE Volume 7, Issue 5 2015

[2] Andreas Poullikkas, George Kourtis, Ioannis Hadjipaschalis., 2013, A review of net metering mechanism for electricity renewable energy sources, International Journal of Energy and Environment., Volume 4, 2013 975-1002

[3] Kapil Narula, B. SudhakaraReddy.,2015. Will Net Metering Model for Residential Rooftop Solar PV Projects Work in Delhi? A Financial Analysis, International Journal of Renewable Energy Research, Vol.5, No.2 ,2015

[4] Fares S. El-Faouri, Munther Sharaiha, Daoud Bargouth., 2016 , A smart street lighting system using solar energy., PES Innovative Smart Grid Technologies

[5] Akshay Kumar Sahoo, Satya Prasad ,Sibashish purohit ,Lalaji gyansu ray, 2016, Exploring the impact of Net metering with Hybrid Solar wind-Grid based distributed Generation.

[6] Pragya Sharma, Tirumalachetty Harinarayana, 2012, Enhancement of energy generation from two layer solar panels

[7] Santhosh Kumar, M. L. Sudheer,2016, PIC16F72 Micro-controller Based Solar Powered UPS.

[8] Protik Kumar Das, Mir Ahasan Habib, Mohammed Mynuddin, 2015, Micro-controller Based Automatic Solar Tracking System with Mirror Booster.

Vehicle Theft Detection System with Automatic Indication

Akshada D. Avhad, Priti R. Gojare, Santosh S. Nagpure, Sarika M. Jagtap

E&TC Department SPPU NDMVP's KBTCOE, Nashik,
akshuavhad110@gmail.com, p.gojare98@gmail.com, santoshnagpure0510@gmail.com, jagtap.sarika@kbtcoe.org

***Abstract*—This paper presents the design of an automobile anti-theft system based on GSM and GPS module, with automatic directional indicator system. The system is developed based on the high-speed mixed type single-chip LPC2148. Few of the important features supported by our system are alerting owner by SMS about the theft attempt, allowing user to control the system remotely by SMS, fuel locking system. Automobile location can be obtained with the GPS module integrated in anti-theft system. The system can keep in touch with automobile owner through the GSM module, to monitor the safety and reliability of automobile. Automatic indication system is based on the direction of handsle or tilt angle, to prevent the accidents.**

Key words — GSM and GPS module, LPC2148uc, Left and Right indicator, Angle measurement sensor, Power supply, key.

I. INTRODUCTION

The increment in thefts of two wheelers and increasing number of road accidents which are prone only due to negligence in following few traffic rules, such as laziness shown by the drivers in order to indicate the signal lights while shifting of lanes or turning on roads etc. As a result of such accidents, calls for the need of a system that could both prevent the theft of the vehicle as well as ensure the safety of the rider by giving automatic indication system. In this project, ARM7TDMI-S microcontroller based vehicle tracking system with GPS and GSM is attached to the moving vehicle to enable the owner/user to track the location of that vehicle and automatic indication system to give automatic indication. The GPS receiver will continuously give the data indicating the position of the vehicle in terms of latitude and longitude in real time. The GSM module will send the position (Latitude and Longitude) of the vehicle to cell phone from a remote place. For automatic indication according to the angle measured by the angle measurement sensor automatic direction indicator is activated/deactivated.

An efficient security system is implemented for anti-theft using an embedded system integrated with Global Positioning System (GPS) and Global System for Mobile Communication (GSM). This proposed work is an attempt to design and develop a smart anti-theft system that uses GPS and GSM system to prevent theft and to determine the exact location of vehicle. The system contains GPS module, GSM module, LPC2148 microcontroller. GPS system track the current location of vehicle and GSM system is also installed in the vehicle for sending the information to the owner of the vehicle because GPS system can only receive the vehicle location information from satellites. In case of theft attempt this system automatically sends the message to owner. The owner can lock his/her vehicle with the help of SMS. This complete system is designed taking in consideration the low range vehicles to provide them extreme security. This proposed system will continuously monitor a moving vehicle and report the status of the vehicle. [1]-[7]

A. MOTIVATION

Automobile theft has increased in the world. According to National Crime Record Bureau (NCRB) data, more than four vehicles stolen on an average every day in Delhi only. And recovery rate is very low. This is serious problem in automobile field. The traditional security system available in market for two wheelers do not provide enough security to their owners. They can only track and monitor the vehicle such locks are well known to the thieves and they can be easily broken by them. These issues are motivation to design the system.

B. OBJECTIVES

1) To develop efficient automotive security system with GPS for real time tracking of stolen vehicle and GSM for mobile communication.

2) To implement a system which allow user to stop the vehicle remotely by SMS.

3) To design an Automatic directional indicator system for two wheelers to prevent the road accidents.

II. LITERATURE REVIEW

Ref no.	Year	Name of Researcher	Name of Paper	Technique Used
1	2005	Zhixiong Liu,Guiming He	Research on vehicle Anti-Theft System Using Facing Recognition	The system consists of IR illuminator to capture the driver's video. If the current driver is found to be unauthorized alarm is activated and details are send to the owner using GPRS or CDMA.
2	201 1	G.S.Prasanth, Ganesh B.Balaji, T.A.Srinivasa Varadhan	Anti-Theft Tracking System for Automoibles(AutoGSM)	It consists of a GSM module and several other components. The system can be switched ON by an SMS from the owner using GSM, to turn OFF the vehicle, and receive information

				about the vehicles current location & then forward the details to the owner
3	2012	Hu Jianming, LiJie, Li Guang-Hui	Automobile Anti-Theft system Based on GSM and GPS Module	An automobile anti-theft system based on GSM and GPS module. Automobile location can be obtained with the GPS. The system can keep in touch with automobile owner through the GSM module.
4	2015	K. A. Mamun1, Z. Ashraf	Anti-Theft vehicle security system with preventive action	An Anti-Theft Vehicle Security System(ATV2 S)has been design and implemented utilizing sensor-network system which employ Global System for mobile communicatio n (GSM) technology to track the vehicle
5	2016	Md. Maru Rahman, Jannatul Robaiat Mou, Kusum Tara, Md. Arif Mahmud	Real Time Google Map and Arduino based Vehicle Tracking System	Real time Google Map and Arduino based vehicle tracking system is implemented with Global Positioning System (GPS) to provides geographic coordinates at regular time intervals and Global system for mobile communicatio n (GSM) technology GPS module.
7	2017	MrinmoyDey Md. Akteruz zaman Arif and Md.Asif Mahmud	Anti-Theft Protection of Vehicle by GSM	GPS with fingerprint verification & A system has been developed based microcontrolle r that consists of a GPS and GSM for two-way communicatio n. And fingerprint verification of the driver of the vehicle is used to protect the vehicle from anti-theft.

III. ANALYSIS AND DESIGN APPROACH

There are major components in system, including GPS, GSM, LPC2148 microcontroller, Angle measurement sensor and directional indicator system.

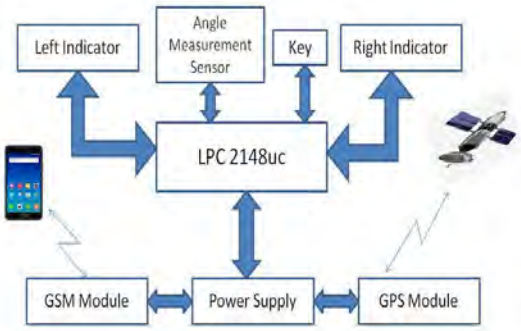


Fig 3.1: Block Diagram of System

A. KEY INSERT UNIT

This unit is used to insert the key. It detects the insertion of the key to the Key hole.

GSM (Global system mobile)

GSM technology is used for intimating the owner. An alert message is sent to the owner if anyone tries to start the vehicle without key. And information about location of vehicle is sent to the owner through message. The module sends SMS to the owner of the vehicle using a predefined time interval which is neither too periodic nor too overdue. Besides, the GSM module also responds to any SMS sent by the owner of the vehicle such as stop the vehicle, change the mobile number registered into the security system and reset the system. The owner has also been given the control to turn the vehicles off just by sending an SMS.

SIM900A GSM/GPRS MODULE:

This is an ultra-compact and reliable wireless module. The SIM900A is a complete Dual-band GSM/GPRS solution in a SMT module which can be embedded in the customer applications allowing you to benefit from small dimensions and cost-effective solutions. Featuring an industry standard interface, the SIM900A delivers GSM/GPRS 900/1800MHz performance for voice, SMS, Data, and Fax in a small form factor and with low power consumption. With a tiny configuration of 24mm x 24mm x 3 mm, SIM900A can fit in almost all the space requirements in your applications, especially for slim and compact. [2-4-7]

B. BUZZER

A buzzer is activated to alert the nearby people if the anyone tries to start the vehicle without key.

C. ANGLE MEASUREMENT SENSOR

It is used to measure the tilt angle of the handle. According to the angle automatic direction indicator is activated/deactivated.

ADXL335:

The ADXL335 is a complete 3-axis acceleration measurement system. The ADXL335 has a measurement range of ±3 g minimum. It contains a polysilicon surface-micro machined sensor and signal conditioning circuitry to implement an open-loop acceleration measurement architecture. The output signals are analog voltages that are proportional to acceleration. The accelerometer can measure the static acceleration of gravity in tilt-sensing applications as well as dynamic acceleration resulting from motion, shock, or vibration. The sensor is a polysilicon surface-micro machined structure built on top of a silicon wafer. Polysilicon springs suspend the structure over the surface of the wafer and provide a resistance against acceleration forces. Deflection of the structure is measured using a differential capacitor that consists of independent fixed plates and plates attached to the moving mass.

The fixed plates are driven by 180° out-of-phase square waves. Acceleration deflects the moving mass and unbalances the differential capacitor resulting in a sensor output whose amplitude is proportional to acceleration. Phase-sensitive demodulation techniques are then used to determine the magnitude and direction of the acceleration. The demodulator output is amplified and brought off-chip through a 32kΩ resistor. The user then sets the signal bandwidth of the device by adding a capacitor. This filtering improves measurement resolution and helps prevent aliasing. [10]

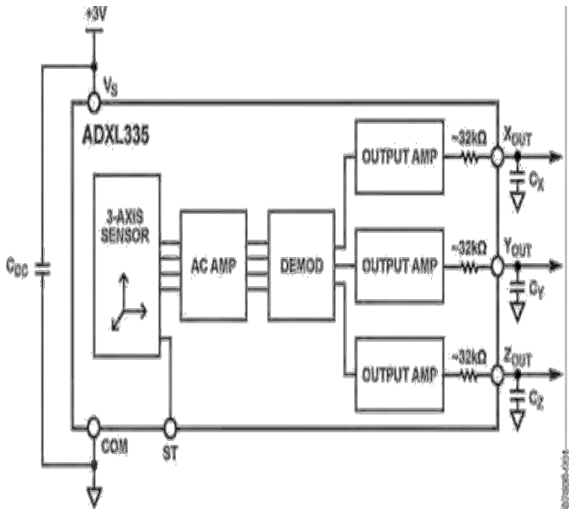


Fig 3.2: Internal circuitary of ADXL33

Features of ADXL335

1) 3-axis sensing.

2) Small, low profile package.

3) 4 mm × 4 mm × 1.45 mm LFCSP.

4) Low power: 350 μA (typical).

5) Single-supply operation: 1.8 V to 3.6 V.

6) 10,000 g shock survival.

7) Excellent temperature stability.

8) BW adjustment with a single capacitor per axis.

D. GPS MODULE

GPS technology is used for tracking vehicle. The Global Positioning System (GPS) is a space-based navigation system that provides location information of moving target can be determined and be transmitted with messaging platform of GSM network.

GPS has become an efficient tool in the field of scientific use, surveillance of objects and tracking. GPS uses Trilateration process to compute the position of an object. In our system, GPS is the key component in getting the vehicles coordinates for tracking. [3-5-6-7]

E. U-BLOX NEO-6M GPS Module with EPROM :

The NEO-6 module series is a family of stand-alone GPS receivers featuring the high performance u- blox 6positioning engine. These flexible and cost effective receivers offer numerous connectivity options in a miniature 16 x 12.2 x 2.4mm package. Their compact architecture and power and memory options make NEO-6 modules Ideal for battery operated mobile devices with very strict cost and space constraints.

Features:

1) EEPROM to save configuration settings.

2) Rechargeable battery for Backup.

3) The cold start time of 38 s and Hot start time of 1s.

4) Configurable from 4800 Baud to 115200 Baud rates. (Default 9600)

5) SuperSense ® Indoor GPS: -162 dBm tracking sensitivity.

6) Support SBAS (WAAS, EGNOS, MSAS, GAGAN)

7) Separated 18X18mm GPS antenna.

F. MICROCONTROLLER

A microcontroller is a computing device which usually consists of built-in peripherals. It included CPU, USART (Universal synchronous asynchronous Receiver Transmitter), timers, counters and other components. It is a programmable device which takes decision. It decodes commands received from GPS module and receives angle information from angle measurement sensor. And provides input to the GSM module buzzer, left and right indicator.

Microcontroller LPC2148 :

The LPC2141/42/44/46/48 microcontrollers are based on a 16-bit/32-bit ARM7TDMI-S CPU with real-time emulation and embedded trace support, that combine microcontroller with embedded high speed flash memory ranging from 32 kB to 512 kB. A 128-bit wide memory interface and a unique accelerator architecture enable 32-bit code execution at the maximum clock rate. For critical code size applications, the alternative 16-bit Thumb mode reduces code by more than 30 % with minimal performance penalty. [9]

SOLENOID VALVE

A solenoid valve is an electro mechanical valve for use with liquid or gas controlled by running or stopping an electrical current through a solenoid, which is a coil of wire, thus changing the state of the valve. The operation of a solenoid valve is similar to that of a light switch, but typically controls the flow of air or water, whereas a light switch typically controls the flow of electricity. Solenoid valves may have two or more ports: in the case of a two port valve the flow is switched on or off; in the case of a three port valve, the outflow is switched between the two outlet ports.

LEFT AND RIGHT INDICATOR

Used to indicate the direction of turn while turning on road. Indicator lights are amber in colour and can be front, the rear and sometimes at the side of the two wheeler on both the left and right hand sides. We use our indicators to show and intended change of direction, whether turning left or right or moving.

POWER SUPPLY

The system is powered by battery present in two wheelers which is generally of 12 Volts. IC7805 is used to provide supply regulated voltage of 5 Volts to MICROCONTROLLER, and other units in the system.

ADDITIONAL COMPONENTS

Along with the major modules and components, a number of other components are also used in our project.
Some of them are listed below:

1) 12V Battery.

2) LEDs, Buzzers and Indicators.

3) MAX232 IC, IC55 and Other Electronic Components.

IV. ADVANTAGES

1) Used to alert and detect automobile theft instantly by monitoring GPS location and allowing user to control the system remotely by SMS.

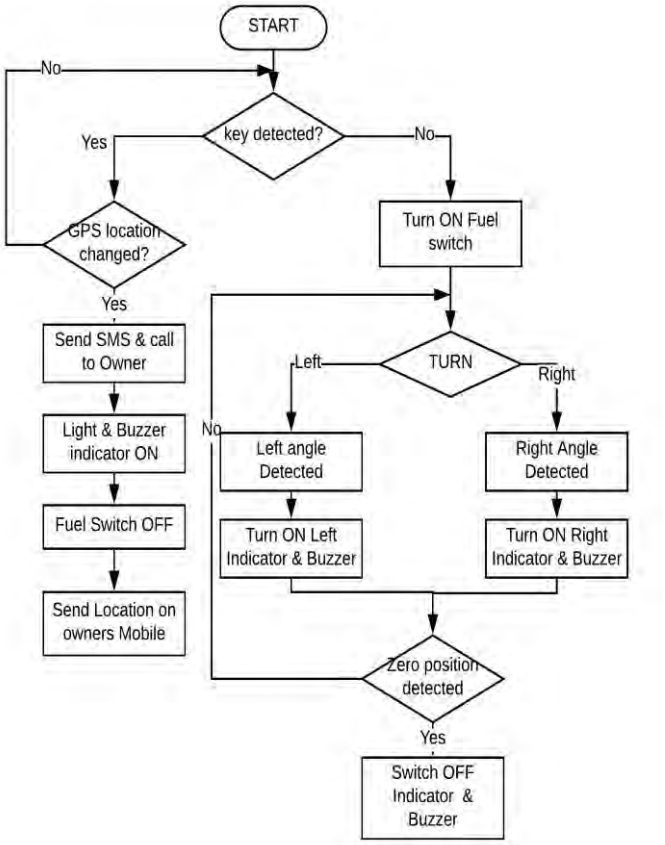
2) Antitheft detection system empowers you to switch off the flow of fuel during theft attempt.

3) This system allows the owner to add or delete the mobile numbers registered in the security system very easily and quickly.

4) Provides self-directional indicator system.

5) This system can be installed in all types of two wheelers irrespective of model and make.

V. FLOWCHART



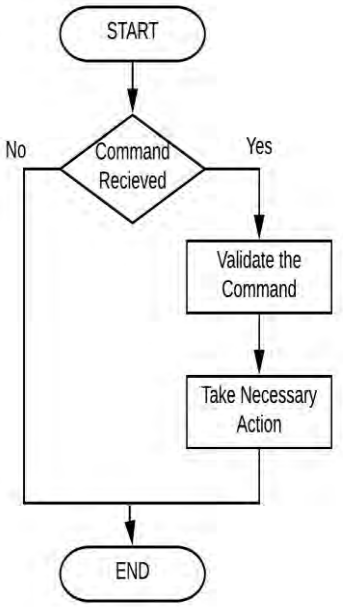


Fig 5.1: Flowchart

VI. APPLICATION

1) Used in two wheelers for Vehicle Theft Detection and Tracking Based on GSM and GPS to ensure security.

2) Theft detection system for two wheelers can be used for bikes used by:

a) Sales man
b) Delivery boys
c) Collection team in banking sectors.
d) Test drive bikes.
e) Rental bikes for tourist.

3) Used to provide automatic indication of signal lights of a two wheeler.

VII. CONCLUSION

We develop the system based on GSM and GPS module. GPS can precise locate out of doors. User can set the alarm directly with the code sent from mobile phone. When buzzer will be triggered then the information regarding the location of vehicle every once in a while to the owner via. SMS. Automobile Anti-Theft System Based on GSM and GPS Module has the functions of remote monitoring, high sensitivity responding and observation location of automobile online. This system has good properties of security integrate with traditional warning system of automobile. We have developed system that can control fuel tank flow via. valve using SMS. Buzzer will also be triggered if theft is detected.
We develop automatic side indication system for two wheelers which can give automatic indication while turning left or right using ADXL335 which is a small, thin, low power, complete 3- axis accelerometer with signal conditioned voltage outputs.

ACKNOWLEDGMENT

We thankfully acknowledge Sivananda Electronics, Deolali camp, Nashik for providing various resources such as laboratory with all needed software platforms for our project.

REFERENCES

[1] Zhixiong Liu, Guiming He, “Research on Vehicle Anti-theft and Alarm System Using Facing Recognition” 2005 International Conference on Neural Networks and Brain.

[2] G.S.Prasanth, Ganesh B.Balaji, T.A.Srinivasa Varadhan “Anti-Theft Tracking System for Automobiles (AutoGSM)”. 2011 IEEE International Conference on Anti-Counterfeiting, Security and Identification.

[3] Hu Jian-ming ,Li Jie,LiGuang-hui “Automobile Anti-theft System Based on GSM and GPS Module” Fifth International Conference on Intelligent Networks and Intelligent Systems 2012.

[4] K. A. Mamun1, Z. Ashraf “Anti-theft vehicle security system with preventive action” in 2012.

[5] Shruthi.K, Ramaprasad.P, Ruschil Ray, Manjunath A. Naik, Shubham Pansari “Anti-theft vehicle Tracking System with a Smartphone Application” 2015 International Conference on Information Processing (ICIP) 2015.

[6] Md. Marufi Rahman, Jannatul Robaiat Mou, Kusum Tara, Md. Ismail Sarkar “Real Time Google Map and Arduino Based Vehicle Tracking System”. 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE) 8-10 December 2016.

[7] Mrinmoy Dey*, Md. Akter uzzaman Arif and Md. Asif Mahmud “Anti-theft Protection of Vehicle by GSM & GPS with Fingerprint Verification”. International Conference on Electrical, Computer and Communication Engineering (ECCE), February 16-18, 2017.

[8] Muhammad Ali Mazidi, Shujen Chen, and Eshar Ghaemi’s “Atmel ARM programming for embedded system Volume 5” (Mazidi and Naimi ARM series.

[9] S.McNamaraY.B.Gianchandani’s“MEMS-Based Sensors”.

[10] ADXL335,smallpower,3axis,3gravityAccelerometer datasheet

[11] https://www.sparkfun.com/datasheets/Components/SMD/adxl335.pdf.

Autotracker Solar Based Multilevel Inverter For Medium Power Application

Lalit Nitin Chopade,Suwarna Prasad Rao,Rani Sham Shinde,Prof.Balasaheb Jagannath Pawar

Electronics & Telecommunication NDMVPS's KBTCOE, NASHIK.
lalit.chopade193 @gmail.com,suwarna612 @gmail.com, bjpawar10777@gmail.com

Abstract—In this paper inverters are nowadays very much in demand due to their serviceability. Based on the area of application there are various types of inverters available. Among them modular multilevel inverter is very useful in every industrial applications, power system and home appliance. There are numerous limitations of conventional two level inverter like presence of harmonics and high rate of change of voltage that is caused in output. The elimination of presence of harmonics in the inverter output involves various techniques. Modular multilevel inverter is one of the very efficient ways to eliminate the presence of harmonics and high change in the output voltage. It provides very less or negligible change in voltage between two levels that is ultimately solution for two level inverter. In this project we are using modular multilevel inverter where due to presence of multiple DC source and large number of switches change in output voltage from one level to other will be less. In this system we have used renewable energy as a source which is solar energy. The design and implementation of a system based on solar energy is a challenge because the angle of inclination of sun-rays with the solar panels continuously changes due to the movement of the sun from east to west because of earth's rotation. Therefore we have used dual axis solar tracking system for getting maximum efficiency of solar panel.

Keywords — Modular Multilevel Inverter, Harmonics, Renewable Energy, Solar Tracking, Dual Axis.

I. INTRODUCTION

From last few decades the requirement of low power electronic devices has been increased. People are selecting devices which have low power consumption with good efficiency. And also industrial applications have begun to required higher power instrument in recent year. Some medium voltage motor drivers and utility application required medium voltage and megawatt power level. The concept of multilevel converters has been introduced since 1975. A multilevel converter not only achieves high power ratings, but also enables the use of renewable energy sources. Renewable energy sources such as photovoltaic, wind, and fuel cells can be easily interfaced to a multilevel converter system for a medium and high power application. Now-a-days solar energy is becoming increasingly lucrative with the increasing cost and continuous depletion of the non-renewable energy resources. However, in spite of the multiple benefits of solar energy, solar panels which capture sunlight are stationary (solar array has a fixed orientation to the sky). These stationary as well as expensive solar panels are unable to extract the maximum solar energy as there is no stability of weather conditions. The power output of solar panels is maximum when it is oriented perpendicularly to the direction of sun rays as both the area of illumination of sunlight on solar panels and intensity of sun-rays is maximum in this case. It has been found out that the efficiency of solar panels improve by 30-60 percent when we use a solar tracking system instead of a stationary array of solar panels. The design and implementation of a power efficient solar tracker is therefore a challenge owing to the immobility of the solar panels. When the level of inverter is increased the output voltage waveform will be ripple free and harmonics will be reduced. The harmonics is mainly caused due to low switching losses. This results in high output voltage with the harmonics, there are many type of techniques can be done to analysis the harmonics and to get a pure sinusoidal waveform. In our project we are studying the 5-level single phase Modular Multilevel Inverter with various sinusoidal pulse width modulation techniques. The harmonics has been decreased up to a certain level and the output waveform has been measured in a digital oscilloscope. In our project, we are using an Arduino UNO microcontroller for generating the pulses.

II. MOTIVATION

The need for introducing multilevel inverter becomes an essential to overcome the shortcoming of conventional two level inverter and efficiently handle high power loads. Also multilevel inverter gives good power quality, low switching losses, and high voltage capability. A multilevel inverter not only achieves high power ratings, but also enables the use of solar energy which is a pollution free and environment friendly renewable energy resources. This is the motivation behind this project.

III. OBJECTIVES

[1]Study of the basic operating principle of the MMC.
[2]Development of a modulation technology, aiming to reduce the distortion occurs in conventional two level inverter like EMI and RFI.
[3]To provide sinusoidal waveforms with low-level harmonic content to reduce distortion.
[4]To use renewable energy resource which is solar energy.
[5]To increase efficiency of solar panel by using automatic solar tracking system because it has been found out that the efficiency of solar panels improve by 30-60 percent by using solar tracking system.

IV. LITERATURE REVIEW

Ref no.	Year	Name of Researcher	Name of Paper	Technique Used
1	2014	B.Rajashekar, T.Praveen Kumar, R. Ramesh	Cascaded H-Bridge Multilevel Inverter	

			with a New Selective Harmonic Mitigation Technique to Meet Grid Codes Under Non-Equal Dc Link Voltages with Power Quality Enhancement	
2	2013	Estíbaliz Solas, Gonzalo Abad, Jon Andoni Barrena,, Sergio Aurtenetxea,, Ainhoa Cárcar, and Ludwik Zaja¸c	Modular Multilevel Converter With Different Submodule Concepts-Part I:	Submodule concept
3	2015	Suman Debnath, Jiangchao Qin, Behrooz Bahrani	Operation, Control, and Applications of the Modular Multilevel Converter	MMC Topology
4	2009	S.Albert Alexander, T.Manigandan, N.Senthilnathan	Digital Switching Scheme For Cascaded Multilevel Inverters.	A new digital switching technique for cascaded multilevel inverter (CMLI) for the purpose of power quality improvement is proposed in order to ensure an efficient voltage utilization and better harmonic spectrum.
5	2016	Nader Barsoum, Rinoe Nizam , Evin Gerard	New Approach on Development a Dual Axis Solar Tracking Prototype	To track the sun in two directions that is elevation and azimuth, a dual-axis tracking prototype is developed to capture the maximum sun rays by tracking the movement of the sun in four different directions.

V. ANALYSIS AND DESIGN APPROACH

There are major components in system, including Solar panel, Batter, sensor, inverter circuit and ATmega 328P processor.

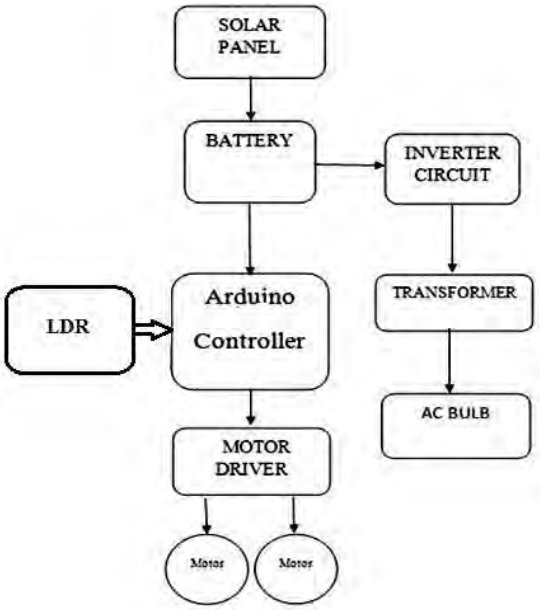


Figure 1: Block Diagram of System

1. SOLAR TRACKING SYSTEM:

Solar tracking system consists of LDR as a sensor, Solar Panel, Battery, Motor Driver and DC Motor. There are total four LDRs and Four 10K resistors are connected in a voltage divider fashion and the output is given to four Analog input pins of Arduino. LDRs are used as the main light sensors. Two DC motors are fixed to the structure that holds the solar panel. Four LDRs are fixed at four corners of solar panel. The motor driver circuit is used to drive the 12V DC motor and IC L239D is used here as motor driver which can drive two DC motors.

The motor requires 12V DC supply. When light falls on LDR its resistance gets decreased which is sensed by Arduino Controller. Hence DC motor is turned ON and Solar panel is then turned towards the direction of sun. Here we are using two axis solar tracking system. Hence efficiency of solar panel gets increased.

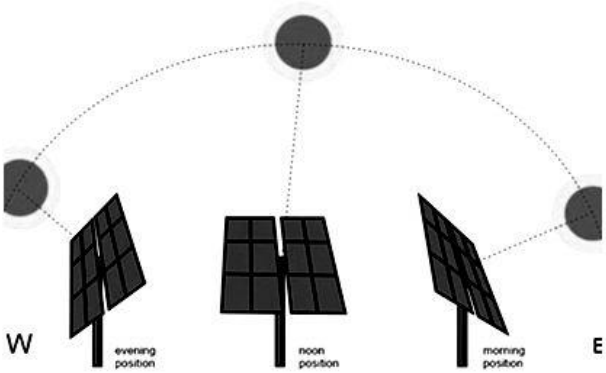


Figure 2: Solar Tracking System

When light falls on LDR its resistance gets decreased which is sensed by Arduino Controller. Hence DC motor is turned ON and Solar panel is then turned towards the direction of sun. Here we are using two axis solar tracking system. Hence efficiency of solar panel gets increased. For east west tracking, the analog values from two top LDRs and two bottom LDRs are compared and if the top set of LDRs receive more light, the vertical DC motor will move in that direction. If the bottom LDRs receive more light, the DC motor moves in that direction. For angular deflection of the solar panel, the analog values from two left LDRs(upper left and lower left) and two right LDRs(upper right and lower right) are compared. If the left set of LDRs receives more light than the right set, the horizontal DC motor will move in that direction. If the right set of LDRs receive more light, the DC motor moves in that direction. The battery gets charged from solar panel which further supplies power to the whole system. The battery used is having rating of 12V, 7A.

2. INVERTER CIRCUIT:

It consists of several MOSFETs. Inverter circuit takes the input as PWM signal from controller and generate multilevel voltage waveform at its output which is very much similar to sinusoidal waveform.

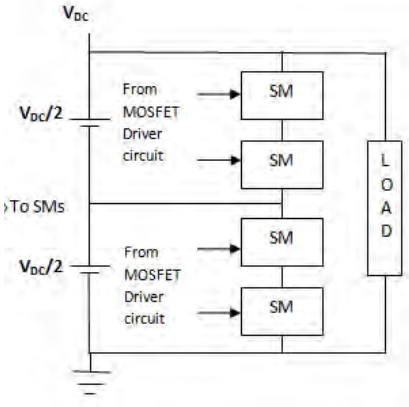


Figure 3: Modular Inverter Cicuit

3. CONTROLLER:

In this system we are using Arduino Uno Controller. It is used for auto tracking of solar panel. As the sun moves from east to west, the microcontroller checks the output of the sensor and once it detects a high output, the stepper motor starts rotating towards the direction of maximum solar radiation. Accordingly electricity is generated and stored into the battery. Also it is used for generating PWM signal which is given to the input of inverter circuit that is gate of MOSFET to controller their switching. Microcontroller requires dc supply which is directly provided from battery.

Technical specifications:

- Microcontroller: Microchip ATmega328P
- Operating Voltage: 5 Volt
- Input Voltage: 7 to 20 Volts
- Digital I/O Pins: 14 (of which 6 provide PWM output)
- Analog Input Pins: 6
- AC Current per I/O Pin: 20 mA
- DC Current for 3.3V Pin: 50 mA
- Flash Memory: 32 KB of which 0.5 KB used by bootloader
- SRAM: 2 KB
- EEPROM: 1 KB
- Clock Speed: 16 MHz
- Length: 68.6 mm
- Width: 53.4 mm
- Weight: 25 g

4. Step up Transformer:

The output from inverter circuit is not sufficient to supply the required current to the load hence a step up transformer of 12-0-12V is used after inverter stage.

5. Load(AC Bulb):

An AC bulb of 7W is connected as load at the output of inverter

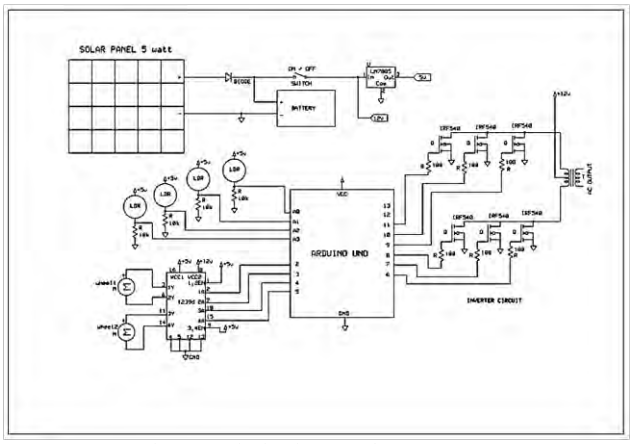


Figure 4: Cicuit Diagram of System

sV. ADVANTAGES

1) Better waveform quality of output voltage (more sinusoidal).
2) Reduced dV/dt,
3) Reduction in EMI.
4) Lower voltage rating devices can be used.
5) Solar trackers generate more electricity than their stationary counterparts due to an increased direct exposure to solar rays.
6) Tracking system enhances clean and emission free power production.

VI. FLOWCHART

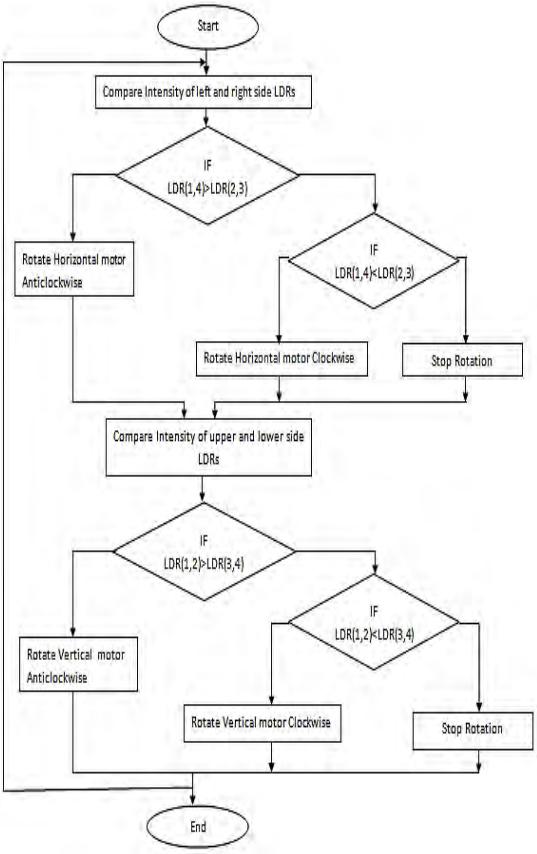


Figure 5 :Flowchart

VII. APPLICATIONS

1. Motor Drivers
2. Active Filters
3. Electric vehicle drivers
4. DC Power source utilization
5. Power factor compensators
6. Back to back frequency link systems
7. Interfacing with renewable energy resources
8. For home appliances.
9. Power generation at remote places.

VIII. CONCLUSION

The idea of Autotracker Solar Based Multilevel Inverter For Medium Power Application serves as an alternative to traditional two level inverter. The drawback of two level inverter is creating harmonic distortions in the output voltage and also has a high dv/dt as compared to that of a multilevel inverter. This drawback has been overcome by using multilevel inverter which works on the concept of submodules. When the level of inverter is increased the output voltage waveform will be ripple free and harmonics will be reduced. The harmonics is mainly caused due to low switching losses. Also it includes the design and implementation of a dual axis solar tracking system which gives maximum power efficiency.

ACKNOWLEDGMENT

We thankfully acknowledge Raj Electronics, Nashik for providing various resources such as laboratory with all needed software platforms for our project and special to Prof. B.J. Pawar giving us all the help and guidance we needed.
Balasaheb Jagannath Pawar M.E (Electronics) from T.K.L.E.T. Warana Nagar, Kolhapur, Presently working as a Assistant professor at N.D.M.V.P.S's, College Of Engineering, Nashik in the Department of Electronics and Telecommunication Engineering.
Lalit Nitin Chopade B.E (Electronics and Telecommunication) from N.D.M.V.P.S's, College Of Engineering, Nashik.
Rani Sham Shinde B.E (Electronics and Telecommunication) from at N.D.M.V.P.S's, College Of Engineering, Nashik.
Suwarna Prasad Rao B.E (Electronics and Telecommunication) from at N.D.M.V.P.S's, College Of Engineering, Nashik

REFERENCES

1. Cascaded H-Bridge Multilevel Inverter with a New Selective Harmonic Mitigation Technique to Meet Grid Codes Under Non-Equal Dc Link Voltages with Power Quality Enhancement by B.Rajashekar1, T.Praveen Kumar2, R.Ramesh3, Vol. 3, Issue 9, September 2014
2. Modular Multilevel Converter: by R. Marquardt, The 2010 International Power Electronics Conference
3. Operation, Control, and Applications of the Modular Multilevel Converter: A Review by Suman Debnath, Student Member, IEEE, Jiangchao Qin, Student Member, IEEE, Behrooz Bahrani, Member, IEEE, Maryam Saeedifard, Senior Member, IEEE, and Peter Barbosa, Senior Member, IEEE, IEEE TRANSACTIONS ON POWER ELECTRONICS, VOL. 30, NO. 1, JANUARY 2015
4. The Age of Multilevel Converters Arrives by LEOPOLDO G. FRANQUELO, JOSE RODRÍGUEZ, JOSE I. LEON, SAMIR KOURO, RAMON PORTILLO, and MARIA A.M. PRATS
5. Modular Multilevel Converter With Different Submodule Concepts—Part I: Capacitor Voltage Balancing Method by Estíbaliz Solas, Gonzalo Abad, Member, IEEE, Jon Andoni Barrena, Member, IEEE,Sergio Aurtenetxea, Member, IEEE, Ainhoa Cárcar, and Ludwik Zaja¸c, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, VOL. 60, NO. 10, OCTOBER 2013

A Wearable And Accessible Smartphone Application For Blind People

Aishwarya Ramnath Shelar, Ashlesha Bhagwat Sable,Minal Gangaram Lokhande,Viraj R. Sonawane

Electronics & Telecommunication,NDMVPS'S KBTCOE,NASHIK

aishwaryashelar33@gmail.com, ashlesha10.me@gmail.com, minuu181212.ml@gmail.com, sonawane.viraj@kbtcoe.org

Abstract— In recent years, several approaches have been made to create systems that allow seamless tracking and navigation both in indoor and outdoor environments. This paper proposes a prototype which uses RFID technology to provide location-based services and navigation to the blind, or visually impaired. An overview of previous work made in the context of systems developed to assist the mobility of blind users. The reader reads the tags ID by RFID communication. It then sends this information through a Bluetooth channel to the PDA device. The ID string and data is processed and then an output is sent through another Bluetooth channel to the headsets which finally produce the audio output that is used by the visually impaired person. We propose a RFID based assistive material reading frame work to help blind persons and product packaging from hand-held object in their daily lives. RFID based material reading through unique bar code system. The proof of concept prototype is also evaluated on a dataset collected using ten blind persons to evaluated the effectiveness of the systems hardware. The recognized text codes are output to blind users in speech. Reading is obviously essential in todays society. Printed text is everywhere in the from of reports, receipts, bank statements, restaurant menus, classrooms handouts, product packages, instructions on medicine bottles, etc., This project is developed to make the life of blind people easy. This is a smartphone application based system to scan the RFID tag behind the image, product in museum or shops and read the Id stored in the tag. This is very beneficial in case of finding out the description of packaged information to the blind people and thus helping them in deciding to purchase a product or not especially which are packaged. This is because it becomes very difficult for the blind people to description the packaged information. In order to use this system, all the user needs to do is scan the RFID tag on the product or near by the product in the RFID reader which then resolves the RFID cards which means it scans the tag to find out the Id stored. Thus this application really benefits blind and visually impaired people and thus making their work of identifying products easy. This is very easy to use and affordable as it requires a RFID reader to scan the tag. The RFID cards to identify the Id stored in it And a smart phone to take the signals form the ESP8266 processing the signal and read out the product description. This illustrates a physical system which incorporates an application of RFID and wireless data base record entries. Our proposed system not only eliminates the time consumed in manual information gathering but also maintains the record of entries which can be used for statistical purposes.

Key words — Rfid Reader, Arduino Uno, Esp8266, Blindness.

I. INTRODUCTION

Design and implement an Android application for the visually impaired. The main focus is to understand the user's needs and design the application in an accessible manner. The function the application should offer the user is to provide information about public places like museum, shops, etc. This project each person has the RFID reader before enter into the museum or the shops. After enter the museum or shops the people should know the which object or product begin us using RFID reader and Arduino intefacing with ESP8266. The RFID can read the product or object name and information of blind people. The main study of the project is to the blind people will going to the purchasing or visit to the museum, they can easily identify the what is the product or any other things are begin them. The reader attached to processor, the tag information will process at Arduino and send that information to Android Appllication via ESP8266. The ESP8266 send this information to android application over the internet. Android application convert text to voice signal and send to the respective blind person using headphones.

II. MOTIVATION

About 253 million people are visually impaired worldwide: 39 million are blind and 217 million have low vision (severe or moderate visual impairment) About 90 per cent of the world's visually impaired people live in developing countries. Globally, uncorrected refractive errors are the main cause of visual impairment Cataracts are the leading cause of blindness 65 per cent of visually impaired, and 82 per cent of blind people are over 50 years of age, although this age group comprises only 20 per cent of the world population. Top causes of visual impairment: refractive errors, cataracts and glaucoma. Top causes of blindness: cataracts, glaucoma and age-related macular degeneration. While there is still a long way to go when it comes to eye health, the number of people visually impaired from infectious diseases, however, has greatly reduced in the last 20 years.

ANALYSIS & DESIGN APPROACH

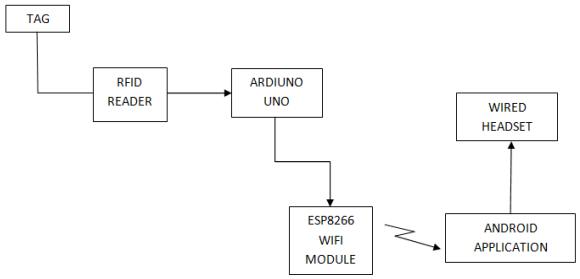


Figure 1. Block Diagram of System

In this paper we are using RFID that is Radio Frequency Identification Tag which will store the unique id in each tag which will placed on different products. The RFID reader will read the unique code from the tag and give that code to the Arduino which is controlling all components. Tag information will process at Arduino and sends that information to Android application via Esp8266 wi-fi module. The esp8266 send this information of respective product to android application over internet. The Android application receive this information and convert text format information to voice using text to speech synthesizer and send that information to in the form of voice to respective person using headphones.

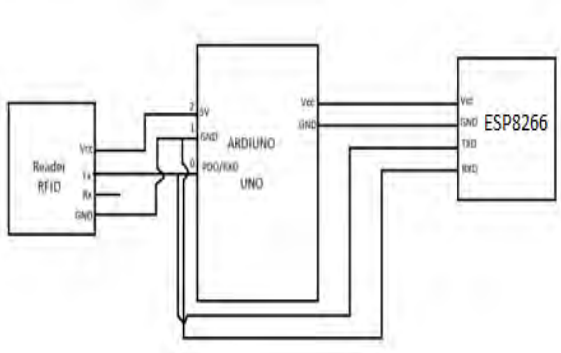


Figure 2. Detailed Circuit Diagram of System

In the given circuit Arduino Uno is used as a controller and RFID reader and ESP8266 module is also used. RFID has 4 output terminals which are VCC, GND, TX and RX. From which VCC and GND are connected to VCC and GND pins of the Arduino Uno respectively and TX pin of RFID is connected to the RX pin of the Arduino Uno. Similarly as of RFID ESP8266 wifi module is also connected to the Arduino Uno. This module has 8 pins which are VCC, GND, GPIO1, GPIO2, CHPD,RST, TX and RX. They are connected as VCC and GND of ESP8266 module are connected to the VCC and GND pins of the Arduino Uno respectively. CH pin of the ESP8266 module is also connected to the VCC pin of Arduino Uno and RST pin of the ESP8266 module is connected to any communication pin of the Arduino Uno.

CIRCUIT DIAGRAM RFID AND ARDUINO UNO

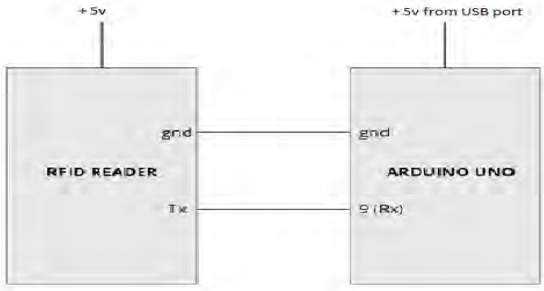


Figure 3. Circuit Diagram of Rfid and Arduino Uno

Make connections as shown. Make sure you connect GND Pin of RFID reader to GND Pin of Arduino. I am using the Software Serial Library of Arduino which enables digital pins to be used in serial communication. I have used pin 9 as the RX of Arduino.

3.1 ARDUINO UNO:

The Arduino UNO is an open-source microcontroller board based on the Microchip ATmega328P microcontroller and developed by Arduino.cc. The board is equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards (shields) and other circuits. The board has 14 Digital pins, 6 Analog pins, and programmable with the Arduino IDE (Integrated Development Environment) via a type B USB cable. It can be powered by a USB cable or by an external 9 volt battery, though it accepts voltages between 7 and 20 volts. It is also similar to the Arduino Nano and Leonardo. The hardware reference design is distributed under a Creative Commons Attribution Share-Alike 2.5 license and is available on the Arduino website. Layout and production files for some versions of the hardware are also available. "Uno" means one in Italian and was chosen to mark the release of Arduino Software (IDE) 1.0. The Uno board and version 1.0 of Arduino Software (IDE) were the reference versions of Arduino, now evolved to newer releases. The Uno board is the first in a series of USB Arduino boards, and the reference model for the Arduino platform. The ATmega328 on the Arduino Uno comes preprogrammed with a bootloader that allows uploading new code to it without the use of an external hardware programmer. It communicates using the original STK500 protocol. The Uno also differs from all preceding boards in that it does not use the FTDI USB-to-serial driver chip. Instead, it uses the Atmega16U2 (Atmega8U2 up to version R2) programmed as a USB-to-serial converter.

3.1.1 ARDUINO UNO FEATURES:

1) Operating Voltage: 5 Volt.

2) Input Voltage: 7 to 20 Volts.
3) DC Current per I/O Pin: 20 mA.
4) DC Current for 3.3V Pin: 50 mA.
5) Clock Speed: 16 MHz.
6) EEPROM: 1 KB.

3.2 ESP8266 WI-FI MODULE:

The ESP8266 series, or family, of Wi-Fi chips is produced by Espress if Systems, a fabless semiconductor company operating out of Shanghai, China. The ESP8266 series presently includes the ESP8266EX and ESP8285 chips. It is A WI-FI module. ESP8266 interface to pic controller. It provide the internet hotspot to Android application. Esp8266 attached to object. ESP8266EX (simply referred to as ESP8266) is a system-on-chip (SoC) which integrates a 32 bit Tensilica microcontroller, standard digital peripheral interfaces, antenna switches, RF balun, power amplifer, low noise receive amplifer, flters and power management modules into a small package. It provides capabilities for 2.4 GHz Wi-Fi (802.11 b/g/n, supporting WPA/WPA2), general-purpose input/output (16 GPIO), Inter-Integrated Circuit (IC), analog-to-digital conversion (10-bit ADC), Serial Peripheral Interface (SPI), IS interfaces with DMA (sharing pins with GPIO), UART (on dedicated pins, plus a transmit-only UART can be enabled on GPIO2), and pulse width modulation (PWM). The processor core, called "L106" by Espressif, is based on Tensilica's Diamond Standard 106Micro 32-bit processor controller core and runs at 80 MHz (or overclocked to 160MHz). It has a 64 KB boot ROM, 32 KB instruction RAM, and 80 KB user data RAM.(Also, 32 KB instruction cache RAM and 16 KB ETS system data RAM.) External flash memory can be accessed through SPI. The silicon chip itself is housed within a 5 mm 5mm Quad Flat No-Leads package with 33 connection pads 8 pads along each side and one large thermal/ground pad in the center.

3.2.1 ESP8266 FEATURES:

1) 16 GPIO pins.
2) 10-bit ADC (successive approximation ADC).
3) Integrated TCP/IP Protocol Stack.
4) Support Antenna Diversity.
5) IEEE 802.11 b/g/n Wi-Fi

3.3 RFID:

Radio-frequency identifcation (RFID) uses electromagnetic fields to automatically identify and track tags attached to objects. The tags contain electronically-stored information.

Passive tags collect energy from a nearby RFID reader's interrogating radio waves. Active tags have a local power source (such as a battery) and may operate hundreds of meters from the RFID reader. Unlike a barcode, the tag need not be within the line of sight of the reader, so it may be embedded in the tracked object. RFID is one method for

Automatic Identification and Data Capture (AIDC). RFID tags are used in many industries, for example, an RFID tag attached to an automobile during production can be used to track its progress through the assembly line; RFID-tagged pharmaceuticals can be tracked through warehouses; and implanting RFID microchips in livestock and pets allows for positive identification of animals. Since RFID tags can be attached to cash, clothing, and possessions, or implanted in animals and people, the possibility of reading personally-linked information without consent has raised serious privacy concerns. These concerns resulted in standard specifications development addressing privacy and security issues. ISO/IEC 18000 and ISO/IEC 29167 use on-chip cryptography methods for untraceability, tag and reader authentication, and over-the-air privacy. ISO/IEC 20248 specifies a digital signature data structure for RFID and barcodes providing data, source and read method authenticity. This work is done within ISO/IEC JTC 1/SC 31 Automatic identification and data capture techniques. Tags can also be used in shops to expedite checkout, and to prevent theft by customers and employees. In recent years, radio frequency identification technology has moved from obscurity into mainstream applications that help speed the handling of manufactured goods and materials. RFID (Radio Frequency Identification) enables identification from a distance, and unlike earlier bar-code technology, it does so without requiring a line of sight. Furthermore, RFID systems can discern many different tags located in the same general area without human assistance.

3.3.1 RFID FEATURES:

1) Operating voltage of EM-18: +4.5V to +5.5V.
2) Current consumption: 50mA Can operate on LOW power.
3) Operating temperature: 0C to +80C.
4) Operating frequency: 125 KHz.
5) Reading distance: 10cm, depending on TAG Integrated Antenna.

3.4 ANDROID APPLICATION:

Thing Speak has APIs for collecting data produced by sensors and APIs for reading that data from applications. Think of an IOT project as two parts. One part of the project is where you need to program a thing to send data. And, the second part is where you want to see that data. ThingSpeak sits in the middle and makes it handy to do both. Android is a Linux based operating system it is designed primarily for touch screen mobile devices such as smart phones and tablet computers. The operating system have developed a lot in last 15 years starting from black and white phones to recent smart phones or mini computers. One of the most widely used mobile OS these days is Android. The Android is software that was founded in Palo Alto of California in 2003. The android is a powerful operating system and it supports large number of applications in Smartphones. These applications are more comfortable and advanced for the users. The hardware that supports Android software is based on ARM architecture platform. The Android is an open source operating system means that its free and any one can use it. The android has got millions of apps available that can help you managing your life one or other way and it is available low cost in market at that reasons android is very popular. The android development supports with the full java programming language. Even other packages that are API and JSE are not supported. The first version 1.0 of android development kit (SDK) was released in 2008 and latest updated version is jelly bean. We will use ThingSpeak in the middle to collect data from sensors and then display the sensor readings on a custom Android app running on a mobile phone. We use the MIT App Inventor to create a custom Android app to see the sensor data and status of the system. This project uses easily accessible hardware to build a proof-of-concept IOT system to

send the signal to the wired headset. First of All you need to make Sure Bluetooth is Enabled then Search for Unpaired Devices, Then Using Device Address, You pair the Device. After Successful Pairing you will need to connect to device and Also HSP or HFP profiles. Note without HSP (Headset Profile) or HFP (Hands-Free Profile) you won't be able to connect and stream calls to your Headset or Speaker.

3.5 WIRED HEADSET:

Headphones (or head-phones in the early days of telephony and radio) are a pair of small loudspeaker drivers worn on or around the head over a user's ears. They are electro acoustic transducers, which convert an electrical signal to a corresponding sound. Headphones are also known as ear-speakers, earphones.

FLOW CHART

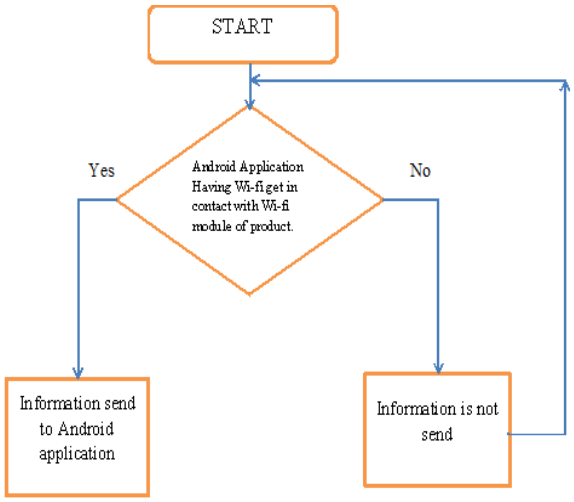


Fig.3.1 Flow chart

4.1 ADVANTAGES

1) Low design time.
2) Low production cost.
3) This system is applicable for both the indoor and outdoor environment.
4) Less space.
5) Low power consumption.
6) It is a dynamic system.

4.2 APPLICATION

1) Blind people without any help get information about object like museum, mall, etc.
2) This system can be used in the home, hospitals and colleges.
3) This system can be used in both the known and unknown environments like airports, malls and public parks, etc.
4) Shops.
5) Banks.

CONCLUSION

In conclusion, the system is a low cost system which is designed to withstand any terrain and surrounding, providing tactical and surveillance and better comfort. Moreover, implementation of wireless & GSM connection in control board allows the system install in more simple way. RFID technology positively promises an increased effectiveness and improved efficiency for business and administrative processes. All the future work is expected without spend extra cost, even one cent from the current system

ACKNOWLEDGMENT

We thankfully acknowledge Corpokin, Nashik for providing various resources such as laboratories with all needed software platforms for our project and special to Prof. V.R.Sonawane giving us all the help and guidance we needed. Viraj R. Sonawane M.Tech(Electronics and Telecommunication) Presently working as a Assistant professor at N.D.M.V.P.S's, College Of Engineering, Nashik in the Department of Electronics and Telecommunication Engineering. Aishwarya R. Shelar B.E(Electronics and Telecommunication) from N.D.M.V.P.S's, College Of Engineering, Nashik. Ashlesha B. Sable B.E(Electronics and Telecommunication) from N.D.M.V.P.S's, College Of Engineering, Nashik. Minal G. Lokhande B.E(Electronics and Telecommunication) from N.D.M.V.P.S's, College Of Engineering, Nashik.

REFERENCES

[1]. A. Adelmann, M. Langheinrich, and G. Floerkemeier. A toolkit for bar-code- recognition and resolving on camera phones - jump starting the internet of things. In Workshop on Mobile and Embedded Interactive Systems (MEIS06) at Informatik 2006, 2006.

[2]. D. Chai and F. Hock. Locating and decoding EAN-13 barcodes from images captured by digital cameras. In Information, Communications and Signal Pro-cessing, 2005 Fifth International Conference on, pages 15951599, 2005.

[3]. O. Gallo and R. Manduchi. Reading challenging barcodes with cameras. In IEEE Workshop on Applications of Computer Vision (WACV) 2009, Snowbird, Utah, December 2009.

[4]. E. Joseph and T. Pavlidis. Deblurring of bilevel waveforms. IEEE Transactions on Image Processing, 2, 1993.

[5]. Benjamin J. M., Ali N. A., Schepis A. F., A Laser Cane for the Blind Proceedings of the San Diego Biomedical Symposium, Vol. 12, 53-57.

[6]. Osama Bader AL-Barrm International Journal of Latest Trends in Engineering and Technology (IJLTET) .

[7]. Smart Cane Assisted Mobility for the Visually Impaired, World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:6, No:10, 2012.

[8]. Lise A. Johnson and Charles M. Higgins, A Navigation Aid for the Blind Using Tactile-Visual Sensory Substitution.

Energy Management for Hybrid Generation System and Automation

Tejaswini C. Janjale,Shubham K. Khodade,Vishal A. Murkute,Atul R. Chaudhari

E & TC Department,NDMVP'S KBTCOE, NASHIK
SPPUPuneUniversity

ABSTRACT— To generate maximum energy from these resources, hybrid mechanism is being implemented. In this paper, an attempt is made to achieve maximum utilization of the available resources. This is done by adding features such as solar mechanism and wind are used for hybrid energy generation. The hybrid power generation systems will in turn be used for charging the batteries. The Hybrid (Wind / hydro / solar) system is more economical, environmental friendly. The power generation capacity of hybrid system is more than the power generation capacity of the individual system.

Keywords — Hydro power, Solar energy, Wind energy

1. INTRODUCTION

We all know that the world is facing a major threat of fast depletion of the fossil fuel reserves. Most of the present energy demand is met by fossil and nuclear power plants. A small part is met by renewable energy technologies such as the wind, solar, biomass, geothermal etc. There will soon be a time when we will face a severe fuel shortage. As per the law of conservation of energy, "Energy can neither be created, nor be destroyed, but it can only be converted from one form to another". Most of the research now is about how to conserve the energy and how to utilize the energy in a better way. Research has also been into the development of reliable and robust systems to harness energy from nonconventional energy resources. Among them, the wind and solar power sources have experienced a remarkably rapid growth in the past 10 years. Both are pollution free sources of abundant power. With high economic growth rates and over 17 percent of the world's population, India is a significant consumer of energy resources. Despite the global financial crisis, India's energy demand continues to rise. India consumes its maximum energy in Residential, commercial and agricultural purposes in comparison to China, Japan, and Russia. Solar energy is energy from the Sun. It is renewable, inexhaustible and environmental pollution free. Solar charged battery systems provide power supply for complete 24 hours a day irrespective of bad weather. By adopting the appropriate technology for the concerned geographical location, we can extract a large amount of power from solar radiations. More over solar energy is expected to be the most promising alternate source of energy. The global search and the rise in the cost of conventional fossil fuel is making supply-demand of electricity product almost impossible especially in some remote areas. Generators which are often used as an alternative to conventional power supply systems are known to be run only during certain hours of the day, and the cost of fueling them is increasingly becoming difficult if they are to be used for commercial purposes. Wind energy is the kinetic energy associated with the movement of atmospheric air. It has been used for hundreds of years for sailing, grinding grain and for irrigation. Wind energy systems convert this kinetic energy to more useful forms of power. Wind energy systems for irrigation and milling have been in use since ancient times and at the beginning of the 20th century it is being used to generate electric power. Windmills for water pumping have been installed in many countries particularly in the rural areas. Wind turbines transform the energy in the wind into mechanical power, which can then be used directly for grinding etc. or further converting to electric power to generate electricity. Wind turbines can be used singly or in clusters called wind farms.

There is a growing awareness that renewable energy such as photovoltaic system and Wind power have an important role to play in order to save the situation. Hybrid power system consist of a combination of renewable energy source such as wind generators, solar etc. of charge batteries and provide power to meet the energy demand, considering the local geography and other details of the place of installation. These types of systems are not connected to the main utility grid. They are also used in stand-alone applications and operate independently and reliably. The best application for these type of systems are in remote places, such as rural villages, in telecommunications etc. The importance of hybrid systems has grown as they appear to be the right solution for a clean and distributed energy production. This paper presents the Solar-Wind hybrid Power system that harnesses the renewable energies in Sun and Wind to generate and supply electricity to a private house, farm house, a small company, an educational institution or an apartment house depending on the need at the site where used.[1-2-3]

1.1. MOTIVATION

In this present modern world, everyone is habituated to use electrical power for all the applications. As the world population is increasing, electrical power demand is also increasing. The conventional sources such as coal, diesel, and gas are going to depleting. The need of renewable energy

sources becomes important in this context. Many renewable energy technologies today are well developed, reliable and cost competitive with the conventional fuel generators. Cost of renewable energy is on falling trend as the demand and production increases. Renewable Power generation gives the best solution to load demand and environment issues. SWHES is proposed as Distributed generation to meet the increased load demand.

1.2. OBJECTIVES

[1]To provide uninterruptable power supply to the different applications.

[2]To provide best alternative for non-renewable energy resources.

2. LITERATURE REVIEW

1. Power Generation by Wind Energy

Author: Saskatechewan

Published on: 2 May 2004

Wind power requires realistic cost reliability evaluation techniques that respond to the inherent random variables. This paper presents an evaluation model and applies it to a test system to illustrate the influence of different factors on generation planning.

2. Power Generation By Fuel Cell System

Author: Hans Rainer

Published on: 22 Sept 2005

Fuel cell plants can be installed close to the energy consumers that means directly at the clients site.The paper provides a brief insight into the fuel cell strategy in general and into the most actual fuel cell projects in particular of E.ON Energie - the largest subgroup within E.ON.

3. Fuel Cell Based Hybrid Power Gneration Strategies For Microgrid Applications

Author:U. R.Prasanna, kaushik rajashekara.

Published on: 22 Oct 2015

In this paper, hybrid power generation systems with different combinations of fuel cells and renewable energy sources for microgrid applications are discussed. The hybrid systems discussed in this paper are: combined cycle operation of a solid oxide fuel cell (SOFC) and a micro turbine; Proton exchange membrane (PEM) fuel cell and wind turbine; combination of SOFC and PEM fuel cell; and SOFC and solar thermal power generation system.

4. Power Generation Automation System Designing For Highway

Authors: Anuragh Pandey, J.P.Sharma

Published on: 27 Aug 2015

This paper tries to show how energy can produced using the road and transport with different -2 ways and their storage. The produced power can be use for the lamps near the road and that will be a great boon for the rural villages too. Four ways of electricity generation (wind, solar, breaker pressure) and automation technology is discussed.[1-2-3-4]

3. SOLAR WIND HYBRID GENERATION SYSTEM

The solar wind hybrid energy system is used where the load is relatively small in rural areas. It can be used in schools especially in residential apartments. Also this usage is extended to commercial and industrial sectors.

Below Fig 1 describes the block diagram of Solar Wind Hybrid Energy System.

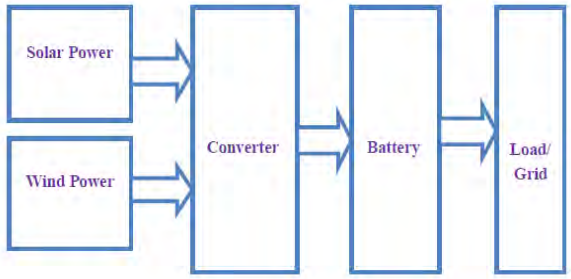


Figure 1: Hybrid Energy System

Solar and wind hybrid power systems are designed using solar panels and wind turbine for generating electricity. According to many renewable energy expert, a “hybrid” electric system that combines home wind electric and home solar electric (photovoltaic) technologies offers several advantages over either single system. In much of the United States, wind speeds are low in the summer when the sun shines brightest and longest and the wind is strong in the winter when less sunlight is available. Because the peak operating times for wind and solar system occur at different times of the day and year, hybrid systems are more likely to produce power when you need it. For the times when neither the wind nor the solar systems are producing, most hybrid systems provide power through the batteries. In the winter when less sunlight is available. Because the peak operating times for wind and solar system occur at different times of the day and year, hybrid systems are more likely to produce power when you need it. For the times when neither the wind nor the solar systems are producing, most hybrid systems provide power through the batteries. In hybrid solar wind generator, the electrical energy generated by solar panel is in AC form which can be converted to DC using invertors and used effectively..These solar panels are available at the output ratings like 5 watts, 10 watts, 20 watts, 100 watts etc. Hence we can select the solar panel as per our need. The huge wind turbines are rotated and thus kinetic

energy is generated by these rotations which can be converted to electrical energy. Minimum wind speed required for connection of the generator to the power grid is known as cut in speed while, maximum wind speed required for the generator for disconnecting the generator from the power grid known as cut off speed. Generally, wind turbines are accessible to the range of speed between cut in and cut off speeds. Wind turbine is a device consist three blades which on rotation produces the electricity in such a way that that the axis of rotation must be aligned with the direction of blowing wind. A gear box is termed as a high-precision mechanical system because it converts energy from one device to another device. Horizontal axis turbines and vertical axis turbines are the most frequently used turbines. An electrical generator is followed wind turbine; hence it is known as wind turbine generator. It generate electricity into battery bank with the help of solar charge controller and wind controller and after that DC load which is stored into battery. Proposed system has the concept of load sharing, depends on the amount of load requirement sources will come in to operation. These are more suitable for distributed and remote power applications.

4. ANALYSIS AND DESIGN APPROACH

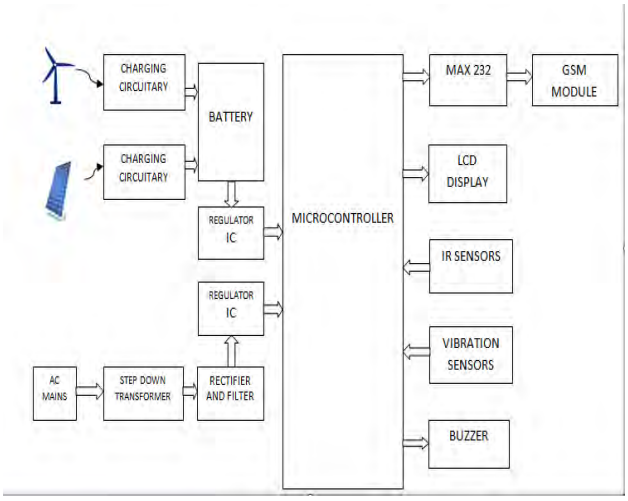


Figure 2: Block Diagram of System

Fig. 2 shows the block diagram of hybrid energy generation system. Intermittent energy resources and energy resources unbalance are the most important reason to install a hybrid energy supply system. The Solar PV wind hybrid system suits to conditions where sunlight and wind has seasonal shifts. As the wind does not blow throughout the day and the sun does not shine for the entire day, using a single source will not be a suitable choice. A hybrid arrangement of combining the power harnessed from both the wind and the sun and stored in a battery can be a much more reliable and realistic power source. The load can still be powered using the stored energy in the batteries even when there is no sun or wind. Hybrid systems are usually built for design of systems with lowest possible cost and also with maximum reliability. The high cost of solar PV cells makes it less competent for larger capacity designs. This is where the wind turbine comes into the picture, the main feature being its cheap cost as compared to the PV cells. Battery system is needed to store solar and wind energy produced during the day time. During night time, the presence of wind is an added advantage, which increases the reliability of the system. In the monsoon seasons, the effect of sun is less at the site and thus it is apt to use a hybrid wind solar system. The system components are as follows. [1-3]

4.1. PHOTOVOLTIC SOLAR POWER

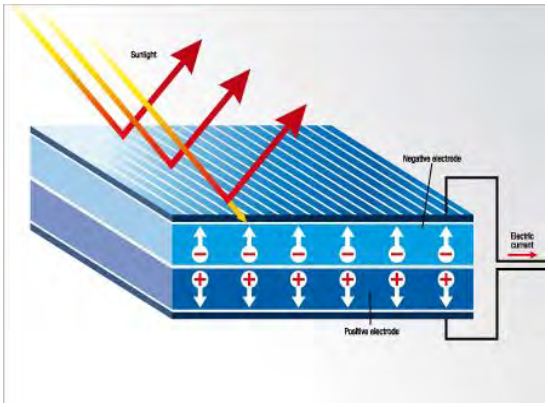


Figure 3: Internal Structure of Solar Panel

Solar panels are the medium to convert solar energy into the electrical energy. Solar panels can convert the energy directly or heat the water with the induced energy. PV (Photo-voltaic) cells are made up from semiconductor structures as in the computer technologies. Sun rays are absorbed with this material and electrons are emitted from the atoms. This release activates a current. Photovoltaic is known as the process between radiation absorbed and the electricity induced. Solar power is converted into the electric power by a common principle called photo electric effect. The solar cell array or panel consists of an appropriate number of solar cell modules connected in series or parallel based on the required current and voltage. [4]

4.2. WIND POWER

The wind energy is a renewable source of energy. Wind turbines are used to convert the wind power into electric power. Electric generator inside the turbine converts the mechanical power into the electric power. Wind turbine systems are available ranging from 5MW. The energy production by wind turbines depends on the wind velocity acting on the turbine. Wind power is able to feed both energy production and demand in the rural areas. It is used to run a windmill which in turn drives a wind generator or wind turbine to produce electricity.[1]

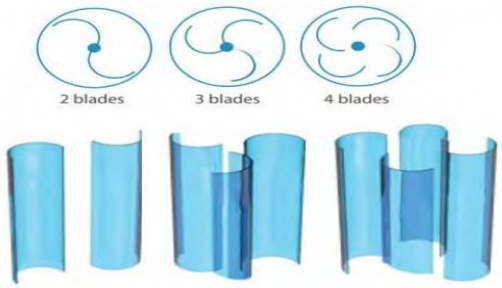


Figure 4: Structure of Wind Turbine

4.3. BATTERIES

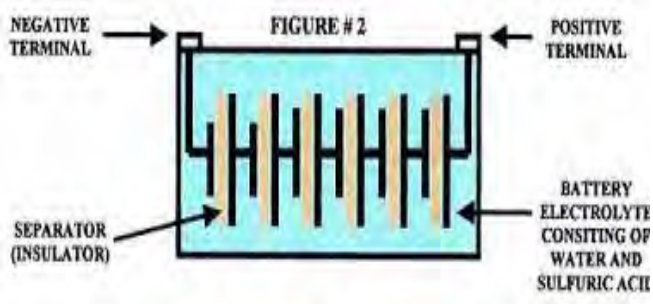


Figure 5: Battery Structure

The batteries in the system provide to store the electricity that is generated from the wind or the solar power. Any required capacity can be obtained by serial or parallel connections of the batteries. The battery that provides the most advantageous operation in the solar and wind power systems are maintenance free dry type and utilizes the special electrolytes. These batteries provide a perfect performance for long discharges.

4.4. REGULATOR IC

The regulator IC 7805 is used to supply constant 5V to the different applications like microcontroller, GSM module IR sensors,vibration sensor etc.

4.5. MICROCONTROLLER

The microcontroller P89V51RD2 is an 80C51 microcontroller which is used to give the control signals to the different applications. [7]

4.6. GSM MODULE

The GSM interface board supports the communication between CPU board and external devices. Whenever the earthquake is detected or over speed is detected the message will be delivered to the user through the GSM module.

4.7. LCD DISPLAY

The LCD (Liquid Crystal Display) display 16*2 matrix is used to display the vehicle count and the speed of each vehicle.

4.8. IR SENSORS

The two pairs of IR sensors are used to count the vehicles. Whenever the vehicle cuts the first and second IR pair, the count will be incremented by one. Also, we can fixed the distance between the two IR pairs and the speed can be calculated using the formula,

Speed= Distance/Time

4.9. VIBRATION SENSORS

The vibration sensor SW420 is used for earthquake detection. When the earthquake is detected the message will be displayed through the GSM module.

4.10. BUZZER

Whenever the over speed or earthquake is detected the buzzer will be turned ON.

5. ADVANTAGES

[1]When one source isn't available the system will still be able to provide energy from the alternate energy source. The clean, emission-free power acts as a reliable backup or supplemental power source.
[2]When both sun and wind is present, the batteries will charge even faster.
[3]Solar panel have a higher initial cost than wind turbines. With this hybrid system, you can spend more on wind power to offset the number of required solar panel and still achieve the same output.
[4]Solar panels require very little maintenance while wind power need only periodic service.

6. APPLICATION

[1] Grid connected: SWHES can be connected to the grid, no matter how much power generating from this source.
[2] Stand alone: Almost all SWHES are the examples of stand – alone systems.
[3] Street lighting: The best running application of SWHES is the solar street lighting. It reduces the load from conventional power plants.
[4] Household: SWHES are integrated with the buildings, equipped with the house for good appearance also.
[5] Remote Applications: like military services where it is impossible to provide conventional power supply

5. CONCLUSION

Hybrid power generation system is good and effective solution for power generation than conventional energy resources. It may provide greater efficiency. It can be provided to remote

places where government is unable to reach. So the power can be utilize where it generated so that it will reduce the transmission losses and cost. Cost reduction can be done by increasing the production of the equipment. People should motivate to use the non-conventional energy resources. It is highly safe for the environment as it doesn't produce any emission and harmful waste product like conventional energy resources. It is cost effective solution for generation. It only need initial investment. It has also long life span. Overall it is good, reliable and affordable solution for electricity generation.

ACKNOWLEDGMENT

We thankfully acknowledge Sivananda Electronics, deolali camp, Nashik for providing various resources such as laboratory with all needed software platforms for our project.

REFERENCES

[1] Saskatchewan, "Power generation by wind energy", IEEE, 2004,International conference on wind energy system.

[2] Anuragh Pandey, "Power generation automation system designing for highway", IEEE, 2015,IEEE international conference on electrical computer and communication technology.

[3] Kavita Sharma, Prateek Haksar "Designing of Hybrid Power Generation System using Wind Energy- Photovoltaic Solar Energy- Solar Energy with Nanoantenna" Internationa Journal of Engineering Research And Applications (IJERA) Vol. 2, Issue 1,Jan-Feb 2012, pp.812-815 .

[4] [4] Sandeep Kumar, Vijay Kumar Garg, "A Hybrid model of Solar-Wind Power Generation System", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE), Vol. 2, Issue 8, August 2013, pp. 4107-4016.

[5] [5]Arjun A. K., Athul S., Mohamed Ayub, Neethu Ramesh, and Anith Krishnan," Micro-Hybrid Power Systems – A Feasibility Study",

[6] Journal of Clean Energy Technologies, Vol. 1, No. 1, January 2013, pp27-32

[7] [6] J.B.V.SUBRAHMANYAM, P.K.Sahoo and Madhukarreddy," Local PV-Wind hybrid systems development for supplying electricity to industry"Acta Electrotechnica,Vol.53,No.1,2012,pp10-15

[8] [7] Muhammad Ali Mazidi, Shujen Chen, And EsharGhaemi's, "Atmel ARM programming for embedded system Volume5(Mazidi and Naimi ARM series).

.

Kinect: From Gaming To Measuring Improvements In Physical Therapy

Akshay Pakhle, Ashvin Khairnar, Animesh Barmukh, Namita Pulgam, Mohammed Waqar Ali

Department of Computer Engineering Ramrao Adik Institute Of Technology, Navi Mumbai, India
akshaypakhle10@gmail.com, yashkhrnr@gmail.com, animeshbarmukh7@gmail.com, ashwinigudewar@gmail.com, mwaqar29@gmail.com

***Abstract*—The use of the commercial video games as rehabilitation tools has recently gained much interest in the physical therapy arena. We present the development of a virtual trainer for use by physiotherapists and patients in exercise based physiotherapy programs. It allows a therapist to tailor exercise requirements to the specific needs and challenges of individual patients. The Microsoft Kinect sensor gives three-dimensional information about the user's body, recognizing skeleton and joint positions allowing developers to innovate in the healthcare field by developing games that cater to the physiotherapeutic needs of the patients. This enables an immersive and natural interaction between the user and the system. Most importantly, the recorded skeletal joint data facilitates quantitative analysis and feedback of patient's body-movements.**

***Index Terms*—Kinect Sensor, Digital Goniometer, Elbow Re-habilitation, Physiotherapy, Unity 3D.**

I. INTRODUCTION

The process of physical therapy and recuperation are effective when patient engagement, timely monitoring and constant checking is done. Traditionally. These exercises are boring and debilitating rendering the patient exhausted and frustrated. Technology and its numerous methods are constantly assessed for their use and implementation in the physical therapy arena and medical technology. The Microsoft Kinect sensor involves a lot of advanced sensing hardware components. Most importantly, it has a depth sensing unit, a color camera component, and a four-microphone array that provides full-body 3D motion detection and capture. Microsoft Kinect will allow the patients to interact with the system and engage with various movements, where they perform prescribed movement combinations barring the need of any external monitoring system. A Kinect-based system can provide proper performance of physiotherapeutically rendered services, increase patient engagement, allow the clinician to monitor progress remotely.

II. LITERATURE REVIEW

In clinical settings and experimental studies of muscular disorders, participants emerged in interactive experience re-duced level of pain, general distress/unpleasantness and report a desire to use Interactive approaches for medical procedures of these kind. Literature survey includes study of various research papers about Physiotherapy using Kinect. Following research papers were studied based on physiotherapy and similar existing projects were compared and analyzed.

Carlo Camporesi, et, al [1], The authors proposed a new solution based totally on Virtual Reality technologies for improving the transport of physical therapy and rehabilitation. Three fundamental factors are addressed: 1) the ability to allow therapists to create new exercises and remedy packages intuitively with the aid of direct demonstration, 2) computerized remedy shipping and monitoring with the use of an independent digital tutor that can screen and quantitatively assess the motions carried out by the patient, and 3) networked collaborative faraway remedy classes with the aid of linked purposes displaying the motions of each the therapist and the patient.

Professor. Wanda Stryla et al [2], the main goal of the article is to attempt to preliminarily analyze the impact of physiotherapy in virtual reality on improvement in motor performance of the paretic upper limb in patients after ischemic stroke. An additional aim of this work is to determine the effect of exercises in a virtual environment on nonpareil's upper limb in the same group of patients.

K. Sooda, A. Muralidharan, A. Sultana, A. Farheen, and K. Majumder, [7]. The first novelty of the proposed technique is that the hand gesture cognizance is performed by way of the first detecting the hand area thru heritage subtraction. Some previous works want the users to put on information glove to collect hand gesture data. However, the different sensors of information glove are pricey and avoid its broad utility in real life.

A.D. Cualin, H.F. Pop, and R.F. Boian [8]. This study incorporate machine learning methods to analyze Kinect body gestures involved in the user interaction with exergaming systems designed for physical rehabilitation. They propose a method to improve gesture recognition accuracy and motion analysis, by extracting from the full body motion data recorded by the Kinect sensor three important features which are relevant to physical therapy exercises: body posture, movement trajectory and range of motion. Thus, this approach has the potential to improve gesture recognition accuracy and provide user feedback on how to improve the movement performed.

Grigore Burdea, et, al [4]. Interactive digital rehabilitation represents the provision of therapeutic interventions locally or at some distance, the usage of Virtual Reality hardware, software and simulations. Such therapy practices have been utilized to various affected person populations, such as muscular-skeletal, post-stroke, muscle debilitation and cognitively impaired. This paper reviews the benefits introduced through VR-enhanced and VR-based rehabilitation to the above patient groups. Also mentioned are the many challenges in integrating this new technological know-how into the medical care system.

III. PROBLEM DEFINITION

Rehabilitation is by its nature repetitive and repetition tends to decouple the mind and reduce patient's motivation. Also, Traditional physiotherapy involves different processes and these processes are usually intensive, time consuming, dependent on the expertise of the therapist and implies the

Collaboration of the patient who is usually asked to perform the therapy multiple times. The passage of time and technological progress made computer technology become ever-present in the lives of majority of people. We are aiming to develop Kinect based games that could transform tedious exercise and injury rehabilitation programs into an engaging enjoyable experience that patients actually look forward to doing. The objectives of this project are:

1. To observe the change in range of motion in injured joint by conventional physiotherapy rehabilitation using digital goniometer.
2. To observe the change in range of motion of injured joint by Kinect-Based-Gaming along with the conventional physiotherapy process.
3. To compare the degree of change in these observations of range of motion and analyze them for further improvement.

IV. PROPOSED SYSTEM

The implementation of this system basically involves the interaction between 4 main entities – Patient, Kinect, Game Software and Web Interface. The interaction between these entities in coherence will allow the effective functioning of this system to cater the ability to carry out physiotherapeutic movements whilst playing a computer game or a visually engaging mechanism, thus hopefully enabling the patient to shift his focus from the pain and more playing the game.

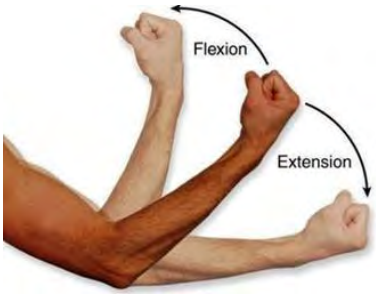


Fig 1: Elbow Flexion and Extension

The following methodology has been proposed in implementing this system:

1. Web Module: The web module will have a registration form which consists of three parts: Admin's module, Patient's module and Doctor's module. The Admin's role will be to monitor the doctor's and the patient's activities, record values and generate reports. The doctor's module will be used to retrieve patient's information, track their improvements and suggest appropriate exercises for them. The patient will use their module to create their profile, update their information and view the assigned doctors and exercises. Database manipulation will be done using PostgreSQL. Interactive charts and graphs will be displayed on the web which will show the patient's exercise data in a graphical form so that it is easy for the doctor to track patient's improvement.
2. Game Module: Physiotherapy exercise based interactive games will be developed on Unity 3D. Unity is a game development tool with a set of resources for rapid development of interactive 3D or 2D applications using C#. In our initial version of system, two types of games will be displayed: one for elbow rehabilitation and one for wrist rehabilitation. The patient can choose the one which he has been assigned by the doctor, then play the game for the allotted time, and exit the game. After this, the scores will be available in the doctor's section with date and time. The recent scores will be updated in the chart also.

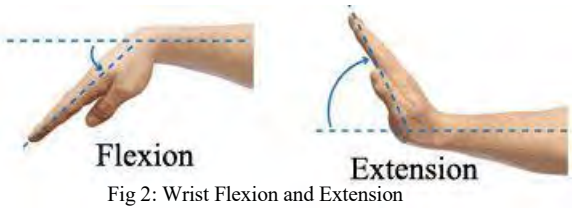


Fig 2: Wrist Flexion and Extension

This methodology of taking the measurements and data from the Kinect instead of traditional goniometer is better than the previous as it reduces the inaccuracies that used to occur in the previous traditional methods. For example, in this system we calculate the distance between the elbow joint and wrist joint with Kinect using the formula and geometric calculations given in Fig 3.0.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

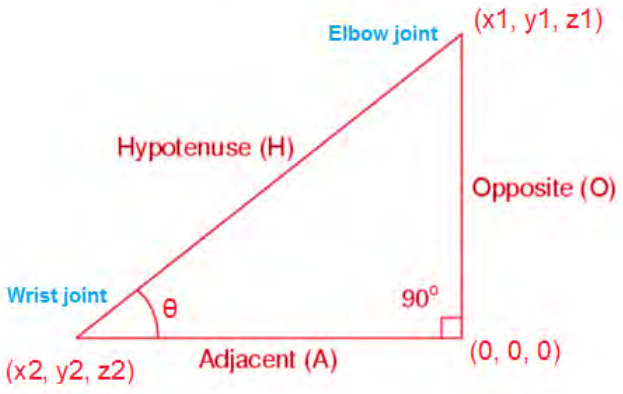


Fig. 3: Geometrical Representation of Elbow Joint

The Microsoft Kinect provides the ability to detect and sense joint movements of a given body. This data can then be used to track the movements and compare with standard values to determine its ability of movements and free Range of Motion. For any given muscle or joint movement, its overall improvement towards normalcy can be determined by calculating the Range of Motion during several repetitions of the prescribed exercise and then calculating an overall average to compare with standard values.

This same concept can be applied to Elbow Flexion and Extension and Wrist Supination exercises. A simple example demonstration of the calculation of the Elbow Range of Motion angle can be seen in figure 4.0.

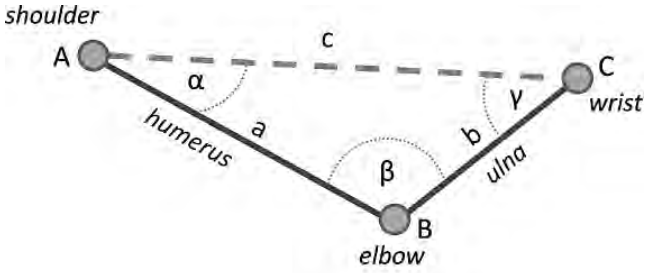


Fig 4: Calculating the Elbow Angle

V. CONCLUSION

In traditional physiotherapy patients tend to get tired and tend to lose interest due to repetitive and tedious exercises, and it also involves use of manual devices like goniometer which becomes quite tedious for both, the patients as well as the doctors. So, an interactive motion detection system using Kinect can be a boon to this problem. Kinect based games for rehabilitation can increase motivation by offering a rich and distinctive environment for the patient. It can also allow networked and remote sessions for rehabilitation where clinicians will still be able to monitor patient progress. The Kinect is continuing to advance, update and develop into a very feasible option for rehabilitation of joint injuries. So, Kinect has solved this problem in this way.

REFERENCES

[1] Carlo Camporesi, Marcelo Kallmann, Jay J. Han.VR Solutions for Improving Physical Therapy, In Proceedings of IEEE Virtual Reality, Orlando, Florida, 2013.

[2] Professor. Wanda Stryla, Journal of Neurology and Neuroscience, Phys-ical Education Academy named Eugeniusz Piasecki in Poznan, Poland, 2015.

[3] JUNGPIL SHIN, (Member, IEEE), AND CHEOL MIN KIM School of Computer Science and Engineering, The University of Aizu, Fukushima, 965-8580, Japan 2 Department of Computer Education, Jeju National University, Jeju, 63243, Korea, Non-Touch Character Input System Based on Hand Tapping Gestures Using Kinect.

[4] Grigore Burdea, Virtual Rehabilitation Benefits and changes, CAIP Center, Rutgers University, USA, 2003.

[5] Juraj MIHAL'OV, Kinect: From Entertainment to Scientific Research on Virtual Movement, Technical University of Kosice,˘ 2016.

[6] Mateus Trombetta, Patricia Paula Bazzanello Henrique, Manoela Rogofski Brum, Eliane Lucia Colussi, Ana Carolina Bertoletti De Marchi, Rafael Rieder, Motion Rehab AVE 3D: a VR-based exergame for post-stroke rehabilitation, Computer Methods and Programs in Biomedicine (2017).

[7] K. Sooda, A. Muralidharan, A. Sultana, A. Farheen, and K. Majumder, "Design and development of hand gesture recognition system," Interna-tional Journal of Advanced Research in Computer Science, vol. 7, no. 7, 2016.

[8] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with Microsoft Kinect sensor: A review," IEEE transactions on cybernetics, vol. 43, no. 5, pp. 1318–1334, 2013.

[9] B. Lange, C.-Y. Chang, E. Suma, B. Newman, A. S. Rizzo, and M. Bolas, "Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor," in Engineering in medicine and biology society, EMBC, 2011 annual international conference of the IEEE, pp. 1831–1834, IEEE.

Earthquake Perdiction Using Support Vector Machine.

Supriya Mandlik, Sakshi Mudrale, Ravita Mishra, Apurva Rawal, Prachi Mirkute

Department of Information Technology Ramrao Adik Institute of Technology, Nerul,NaviMumbai,India.
supriyamodak7@gmail.com, mudralesakshi@gmail.com, ravita.mishra@rait.ac.in, apurvar1508@gmail.com, prachimirkute27@gmail.com

***Abstract*— A natural calamity occurs due to natural hazards. Earthquakes are one of those natural calamity which leads to environmental, financial as well as human loss. In today's scenario, predicting earthquakes is a necessity. The current system explores data mining methods to predict the earthquake. In tweet analysis, tweets were monitored continuously and accordingly predictions were given based on the keywords in the tweet, words in the tweet and the context. But these tweets may contain false information and may give wrong predictions. The proposed system uses Support Vector Machine to predict earthquakes which uses classification and regression techniques, in order to provide enhanced results. The proposed system can be used in construction sites where the builder can use the software in order to predict earthquake in advance so that he/she can decide whether one should construct buildings, offices, hospitals, schools, etc in a particular region.**

Keywords—machine learning, support vector machine (SVM), prediction, data mining, genetic algorithm (GA).

I. INTRODUCTION

The aim of this paper is to study the perception of time reliant seismicity which implies that present seismicity should be estimated on the basis of previous data. This has become an important topic in estimating seismic threats. Generally there are two different methods for predicting earthquake so far used :short term memory cells which can predict exact time, date, location of occurring earthquake whereas using long term memory cells where we will be focusing is predicting earthquake even before a couple of days or months before. In earlier days, earthquake prediction was done using neural networks, tweet analysis and burst detection techniques. In neural network system, raw sensors were used to fetch seismic waves, but there human intervention. Also not always seismic activities can lead to earthquake. In tweet analysis, tweets were monitored continuously and accordingly predictions were given, but those tweets may contain false information and may give wrong predictions. A search input is important for tweets that might be relevant, for this it is necessary to obtain a good set of input queries. Continuous monitoring was done on social media and then predictions were made. This paper presents the first attempt to apply SVM algorithm as well as Genetic Algorithm for long-term earthquake prediction. We have defined huge set of seismic indicators than the existing work in long term earthquake prediction and conducted encyclopedic assessment of various classification techniques in this task. Instead of training individual classification model on small data sets we have mingled the data of various regions in to one data table to bear a common classification model. This paper mainly focuses on the areas which are situated on tectonic plates and are vulnerable to earthquake. It can be used in: construction sites where the builder can use our system in order to predict earthquake in advance so that he/she can decide where one should build offices, hospitals, schools, etc.

II. LITERATURE SURVEY

Predicting Dangerous Seismic Activity with Recurrent Neural Networks. This paper was a solution presented to forecast multiple time data of elevated seismic activities taking place in side coal mines, which can lead to tremendous human loss, presented by Karol Kurach (university of Warsaw) et.al. [2]. As coalmine workers life is in danger while working in it, till date no such system has been deployed which will exactly predict occurrence of seismic events. Predicting Dangerous Seismic Activity with Recurrent Neural Networks paper can directly predict such dangerous activities. The model is generic and works on memory cells. Long term memory cells include data of couple of months and years short term memory cells includes data of hours and days. Robust Detection of Extreme Events Using Twitter: Worldwide Earthquake Monitoring. This paper was published by using Twitter. Twitter is a social media platform which is more prominent for achieving real time data. Huge amount of messages from twitter stream along with false buzz were input source, presented by Barbara Poblete et.al. [1]. Semi supervised initiation was scoped to track various signals in vigorous manner. This system compares the suggested system to bring out the method of predicting earthquake using social media platform. Its implementation is easy and has worldwide scalability. It is good enough for moderate intensity earthquake and has quite acceptable accuracy. Robust detection was an improved system of tweet analysis as it removes the noise from the data and actually searches for the location. This removes the fault of irrelevant information in the existing system but its accuracy was not up to the mark term (months and years) memory cells were used in recurrent neural networks. Robust detection was an improved system of tweet analysis as it removes the noise from the data and actually searches for the location. This removes the fault of irrelevant information in the existing system but its accuracy was not up to the mark.

Table 2.1 Summary of literature survey

Author name and year	Technique used	Advantages	Drawbacks

Karol Kurach, *et.al.* Year: 2016. [2]	Recurrent Neural Network with Long Short-term memory cells.	Applied in Various time variant problems	No human crafted features.
Barbara Poblet, *et.al.* Year: 2018. [1]	Burst detection method has been evaluated.	Volume variation in twitter stream is detected and change parameters to ensure robustness to stream fluctuation	The system verifies the tweet and then gives analysis but it is time consuming.
G..V, Otari, *et.al.* Year: 2012. [3]	Data Mining techniques are used.	Model can be used for any prediction result.	A review of data mining technique in predicting natural disasters.
Amir Abolfazl-Suratgar *et.al.* Year: 2008. [4]	Simulation is used to predict earthquake 2 days before.	Simulation results are promising.	Not all geomagnetic attributes leads to earthquake

Here are some of the gaps presented in existing systems which we will be overcoming in this system. The existing system used only raw sensors which were used to predict seismic activity, i.e. no human intervention in system. As not all seismic activities can lead to earthquake, our system will overcome it by using real time data which is true. Also in tweet monitoring technique, continuous monitoring was required and tweets are needed to be verified as well. This was quite time consuming; our system produces results based on previous records which reduces the time for verification of data. Thus using SVM and GA methodologies, we will be increasing the efficiency of our system.

III. PROPOSED FRAMEWORK.

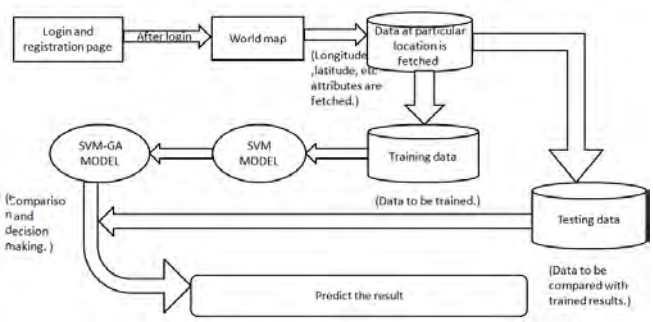


Figure 3.1: System Architecture

A. *System design:*

- *SampleData*: The data collected is the accumulation of all the seismic zone data namely (zone, city, country, previous date of earthquake, Richter magnitude scale, and distance covered)
- *Normalization*: In svm, the data normalization is often referred to whitening as svm refer data in range of 0-1.Soft Normalization is a process which subtracts the mean of the values and divides by twice the standard deviation (again, by dimension).If our input has d dimensions, then we will have d means and d standard deviations, no matter how many training examples we will be having. These normalized vectors are sent to your svm library for training.
- *Training Model*: In machine learning, the set of data used to teach/develop a model is called training model. In Training the model, precise features are taken from training set.. If the training set is labeled correctly, then the model will be able to learn something from these features.
- *Testing Model*: The test set is a dataset which is used to assess how well the model executes Earthquake prediction using support vector machine at making predictions on that test set. In order to make any prediction, we need some original dataset that our model can learn from, and we need a test set to see how well the model actually does at making predictions. Without a test set, there is no way to know whether or not our model over fits or under fits the training data.
- *SVM Model*: SVM is a machine learning algorithm for pattern classification and non-linear regression. The methodology involved in SVM modeling is : learning with combined input and target outcoming values.
- *Genetic Algorithm*: Genetic Algorithm is a biometric algorithm which emulates the natural selection and natural process of genetics. It can abolish the astigmatism of human made judgements and raises the anticipative performance of SVM model.

B. *Methodology used:*

1. Support Vector Machine

Support Vector Machine (SVM) is a machine learning algorithm for pattern classification and non-linear regression. The methodology involved in SVM modeling is: learning with combined input and target out coming values

SVM algorithm:

Input: A (Training data) and B (Testing data)
Output: Predicted result.

1. Start.
2. Select input of (A, B) for SVM.
3. Randomly divide the dataset into equal size subsets.
4. Form subset of them for the SVM model.
5. Verify the model using one subset.

6. Repeat 3 and 4 for all the subsets.
7. Calculate the overall accuracy prediction.
8. End.

2. Genetic Algorithm

Genetic Algorithm is a biometric algorithm which is a natural process of genetics. It can abolish the astigmatism of human made judgments and raises the anticipative performance of SVM model.

GA Algorithm:

1. Start
2. Initiating the data from SVM.
3. Selection of best fitting data.
4. Crossover: It means bringing quality data towards local minima/maxima.
5. Mating: It means throwing out data out of local minima/maxima.
6. Mutation: It is meant for comparison of this data with testing data.
7. Predicting result.
8. End.

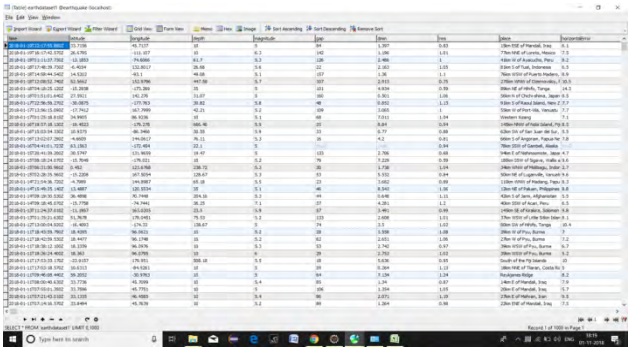


Figure 4.1 : Dataset

Above figure contains training data which is normalized using mention any technique

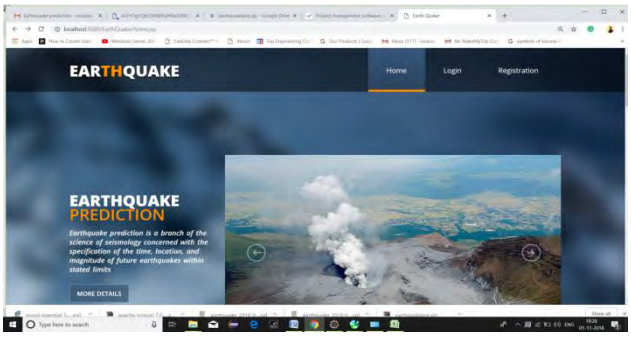


Figure 4.2: Home Page

Above figure shows the GUI of our system which includes home, login and registration.

IV. RESULTS AND DISCUSSION

Following results are based on the dataset used. It includes various attributes time, latitude, longitude, depth, magnitude, etc.But the main focus is on attributes latitude, longitude, magnitude, time and temperature.

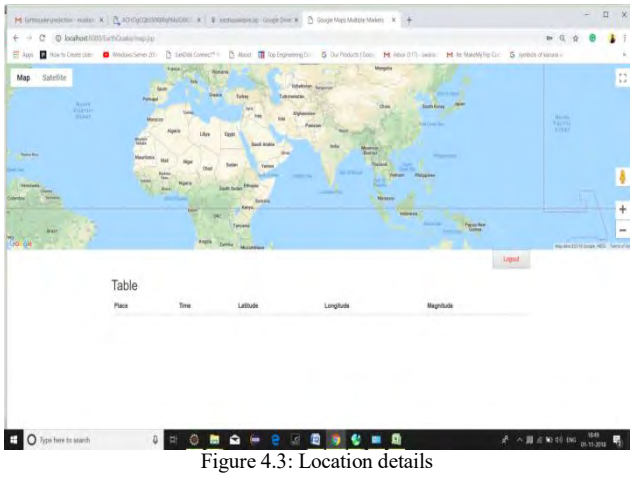


Figure 4.3: Location details

Above figure illustrates, after successful registration and login, clicking on a particular location the system will provide an alert for latitude, longitude, city, location of that region.

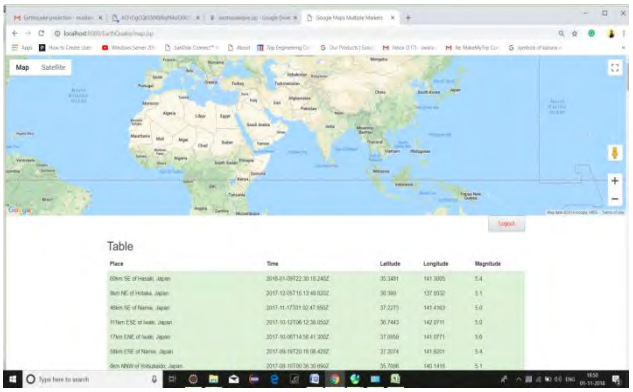


Figure 4.4: Region based output

Above figure displays the records for the selected location.

V. CONCLUSION

Existing system would predict earthquake using recurrent neural networks, tweet analysis etc, but the efficiency of their system would deal only with earthquake. The proposed system can also deal with other natural calamities (tsunamis, hurricanes, cyclones, etc) along with earthquake. Our model will be testing 2 parameters using SVM: one is past data and other is real-time data, compare them and resultant sample will be again compared by GA (Genetic Algorithm). The paper mainly focuses on the areas which are situated on tectonic plates and are vulnerable to earthquake. It can be used in construction sites where the builder can use this system in order to predict earthquake in advance so that he/she can decide where one should construct buildings, offices, hospitals, schools, etc.

VI. REFERENCES

[1] Barbara Poblete, Jheser Guzman, Jazmine Maldonado, Felipe Tobar Robust Detection of Extreme Events Using Twitter: Worldwide Earthquake Monitoring.

[2] Karol Kurach, Krzysztof Pawlowski Predicting Dangerous Seismic Activity with Recurrent Neural Network.

[3] G.V.Otari, Dr.R.V.Kulkarni A Review of Application of Data Mining in Earthquake Prediction.

[4] Amir Abolfazl Suratgar, Farbod Setoudeh, Amir Hossein Salemi, Ali Negarestani Magnitude of Eartquake Prediction Using Neural Network.

[5] W.Dzwinel et al Non multidimensional scaling and visualization of earth quake cluster over space and feature space, nonlinear processes in geophysics 12[2005] pp1-12.

[6] C.Lomnitz. Fundamentals of Earthquake prediction [1994] .

[7] B.Gutenberg and C.H. Richtro, Earthquake magnitude, intensity, energy and acceleration bulseism soc. Am 36,105-145 [1996].

[8] C.Brunk, J.Kelly and Rkohai Mineset An integrate system for data access, Visual Data Mining and Analytical Data Mining, proceeding of the 3rd conference on KDD 1997.

[9] Predicting the Earthquake using Bagging Method in Data MiningS.Sathiyabama, K.Thyagarajah, D. Ayyamuthukumar.

[10] A Bagging Method using Decision Trees in the Role of Base Classifiers, Kristna Machov, Frantiek Bark, Peter Bednr.

[11] Cluster Analysis, Data-Mining, Multi-dimensional Visualization of Earthquakes over Space, Time and Feature Space, Witold Dzwinel, David A. Yuen, Krzysztor Boryczko, Yehuda Ben-Zion, Shoichi Yoshioka.

[12] P. Varotsos, M. Lazaridou, ”Latest Aspects of Earthquake Prediction in Greece based on Seismic Electric Signals”, Techtonophysics, vol. 188, 1991.

[13] https://ieeexplore.ieee.org/document .

[14] A. Ma, Z. Qiang, ”Remote Sensing Information Models for Earthquake Forecasting [J]”, Proceeding of the Twelfth International Conference Applied Geologic Remote Sensing, pp. 11-10, 17 19 Nov. 1997.

[15] A.F. AL-Anazi, I.D. Gates, ”Support vector regression to predict porosity and permeability”, Computer and Geoscience, vol. 39, pp. 64-76, 2014.

[16] B. Mehrkian, A. Bahar, A. Chaibakhsh., ”Geneticoptimized neuro-fuzzy inference system (GONFIS) in nonlinear system identification”, IEEE, International Conference on Control System, Computing and Engineering, 2011. RYCHETSKY,Matthias,Algorithms and Architecture for machine learning based on Regularized Neural networks and Support Vector Approaches.

[17] https://www.springer.com/in/book/9780387772417.

[18] https://journals.plos.org/plosone/article?id=10.1371/journal.pone.

[19] https://en.wikipedia.org/wiki/Supportvectormachine.

[20] The Headquarters for Earthquake Research PromotionEarthquake Research Committee, ”Long-term evaluations of Earthquake on Nankai Trough (in Japanese),” 8th Conference of Subcommittee for Instituting Results n Society, 2001.

Smart E-Glasses For Industrial Applications

Suraj R. Dokhale, Nayan V. Nikam, Akshaykumar G. Ushir, Arvind P. Meshram

Department of E & TC, NDMVP's KBTCOE, Nashik

***ABSTRACT*— This paper presents the design and applications of smart e-glasses.The system is developed based on the wireless bluetooth technology. Important feature supported by our system is displaying real time readings on glasses which are present on measuring device .Wireless communication can be established using bluetooth module HC-05. The system uses arduino pro micro controller to monitor readings on measuring device and display readings on glasses. So this smart e-glasses provide safe and secured environment that will create feeling of safety.**

Key words — Wireless technology, Bluetooth module HC-05, Arduino pro micro, OLED Display, LiPo battery.

I . INTRODUCTION

Nowadays any wireless devices can be controlled from any Part of the world. In this digital world we already use smart phones, smart watches and many smart gadgets. Smart glasses are going to one of them. In industry while working on electronic circuits it is very important to have undisturbed focus.Working in high voltage environment is dangerous. Suppose people are working on fuse boxes or on complicated circuit boards or PLC panels, every time they check the readings on multimeter; they get distracted. So it is most likely that they may misplace their hands and end up touching live voltage point. This may cause serious injury to the respective person. So it is important to provide safe and secured environment or working condition.

This proposed work is an attempt to design and develop a smart e-glasses system that uses bluetooth for wireless communication and OLED display along with lens and acrylic glass structure for displaying the readings within eye sight. System involves arduino microcontroller, bluetooth module, OLED display, acrylic glass, bifocal convex lens, LiPo battery.

System is divided into two parts. One is measuring side and another is displaying side. Measuring side contains voltage measurement circuit. It contains center-tap transformer, half wave rectifier and bluetooth module. Voltmeter circuit works as a transmitting circuit. Arduino Bluetooth module is connected to voltmeter, so output of voltmeter is transmitted through Bluetooth module. Center-tap transformer is used to convert high AC voltage to low AC voltage. A half-wave rectifier is used to convert the ACvolt to DC volt, because Arduino can't handle the AC voltage. This unregulated DC voltage is filtered out by using filter capacitor and zener diode. Displaying side contains bluetooth module, OLED display, acrylic glass, reflector, lens and housing structure. In the receiving circuit Arduino Bluetooth module is connected to OLED display, so the value of Voltmeter is displayed on OLED display. The value from the OLED display goes over the mirror through lens to the reflector (acrylic sheet) right into user's eyes and project a virtual image at a distance of 30cm right into user's view field which is at distance of distinct vision.An efficient security system is implemented for anti-theft using an embedded system integrated with Global Positioning System (GPS) and Global System for Mobile Communication (GSM). The system contains GPS module, GSM module,LPC 2148 microcontroller.GPS system track the current location of vehicle and GSM system is also installed in the vehicle for sending the information to the owner of the vehicle because GPS system can only receive the vehicle location information from satellites. In case of theft attempt this system automatically sends the message to owner.The owner can lock his/her vehicle with the help of SMS. This complete system is designed taking in consideration the low range vehicles to provide them extreme security.This proposed system will continuously monitor amoving vehicle and report the status of thevehicle.[1-2-5-12]

II . MOTIVATION

8% deaths that occur in factories are due to electricity. 56% disasters that occur in factories is due to electrical faults. Almost 12 people die due to electrocution every day. 42% of total fires occur due to electrical sources. 400 deaths from job related electrical accidents per year. Electrocution is the fifth leading cause of death. So it is very important to provide solutions for problems in industries which involves work related to electricity to ensure safe working environment. [5-8-11]

III . OBJECTIVES

- To provide safe and secure environment.
- To make process of taking multimeter reading efficient.
- To avoid deaths from electrocution.
- To reduce time of working while using multimeter.
- To create and maintain feeling of safety. [3-6-9]

IV . ANALYSIS AND DESIGN APPROACH

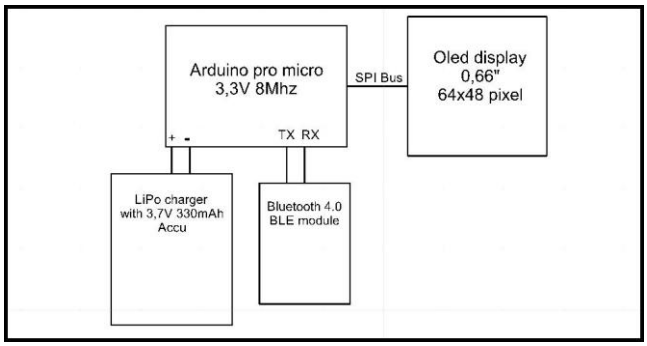


Figure 1: Block Diagram of System

ARDUINO

Arduino board designs use a variety of microprocessors and controllers. The boards are equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards (shields) and other circuits. The boards feature serial communications interfaces, including Universal Serial Bus (USB) on some models, which are also used for loading programs from personal computers. The microcontrollers are typically programmed using a dialect of features from the programming languages C and C++. In addition to using traditional compiler tool chains, the Arduino project provides an integrated development environment (IDE) based on the Processing language project.[3-4-9-11-13]

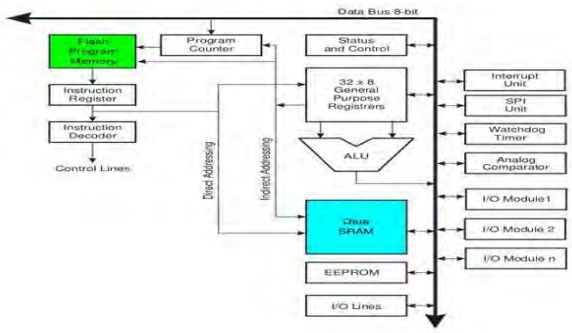


Figure 2: Arduino internal architecture

- Inexpensive - Arduino boards are relatively inexpensive compared to other microcontroller platforms. The least expensive version of the Arduino module can be assembled by hand, and even the pre-assembled Arduino modules cost less than $50
- Cross-platform - The Arduino Software (IDE) runs on Windows, Macintosh OSX, and Linux operating systems. Most microcontroller systems are limited to Windows.
- Simple and clear programming environment - The Arduino Software (IDE) is easy-to-use for beginners, yet flexible enough for advanced users to take advantage of as well. For teachers, it's conveniently based on the Processing programming environment, so students learning to program in that environment will be familiar with how the Arduino IDE works.
- Open source and extensible hardware - The plans of the Arduino boards are published under a Creative Commons license, so experienced circuit designers can make their own version of the module, extending it and improving it.

Bluetooth Module

HC-05 module is an easy to use Bluetooth SPP (Serial Port Protocol) module, designed for trans-parent wireless serial connection setup. The HC-05 Bluetooth Module can be used in a Master or Slave configuration, making it a great solution for wireless communication. This serial port Bluetooth module is fully qualied Bluetooth V2.0+EDR (Enhanced Data Rate) 3Mbps Modulation with complete 2.4GHz radio transceiver and baseband. It uses CSR Blue core 04 External single chip Bluetooth system with CMOS technology and with AFH (Adaptive Frequency Hopping Feature).

HC05 model can be configured either for a Master or for a Slave mode, thus most preferred for a Microcontroller project. The HC-05 module can build a connection to other modules, while HC04/HC-06 modules can be a slave only. Two Slave modules cannot communicate with each other themselves. Bluetooth is a wireless technology standard for exchanging data over short distances (using short-wavelength UHF radio waves in the ISM band from 2.4 to 2.485 GHz) from fixed and mobile devices, and building personal area networks (PANs). Range is approximately 10 Meters (30 feet).[8-11]

Features:

1. Typical 80dBm sensitivity.
2. Up to +4dBm RF transmit power.
3. 3.3 to 5 V I/O.
4. PIO (Programmable Input/output) control.
5. UART interface with programmable baud rate.
6. With integrated antenna and edge connector.

Micro OLED Display

OLED displays are not just thin and efficient they can also be made flexible (even rollable) and transparent. OLED (Organic Light Emitting Diodes) is a light emitting technology, made by placing a series of organic thin lms between two conductors. When electrical current is applied, a bright light is emitted. OLEDs can be used to make displays and lighting. Because OLEDs emit light they do not require a backlight and so are thinner and more efficient than LCD displays (which do require a white backlight).

OLEDs are used today in mobile phones, digital cameras, VR headsets, tablets, laptops and TVs. Samsung is the clear leader in OLED production for mobile devices, and the company uses OLEDs in all of their flagship devices such as the Galaxy S6, the Note 5 and the very latest Tablet Pro S. Other companies, including Apple, Motorola, Dell, Sony, Microsoft, LG and Lenovo are also using OLEDs in some of their mobile devices. OLEDs are used in mobile devices to-day because they are thin, efficient and bright. OLEDs carry a price premium over LCDs, but companies are using these displays more and more as performance increases and prices decrease. Here's our list of products and gadgets with an OLED display. Our own OLED Marketplace is the world's most comprehensive OLED display catalog.

The Micro OLED Breakout Board breaks out a small monochrome, blue-on-black OLED. Its micro, but it still packs a punch the OLED display is crisp, and you can't a deceivingly large amount of graphics on there. This breakout is perfect for adding graphics to your next Arduinoproject, displaying diagnostic information without resorting to serial output, and teaching a little game theory while creating a fun, Arduinobased video game. Most important of all, though, is the Micro OLED is easy to control over either an SPI or I2C interface.

Features:

1. Operating Voltage: 3.3V
2. Screen Size: 64x48 pixels (0.66" Across) Monochrome Blue-on-Black
3. SPI or I2C interface

Lithium Polymer Battery:

Lithium Polymer batteries (henceforth referred to as LiPo batteries), are a newer type of battery now used in many consumer electronics devices. They have been gaining in popularity in the radio control industry over the last few years, and are now the most popular choice for anyone looking for long run times and high power, much higher discharge rates, meaning they pack more LiPo batteries offer three main advantages over the common Nickel-Metal Hydride (NiMH) or Nickel Cadmium (NiCd) batteries:

- LiPo batteries are much lighter weight, and can be made in almost any size or shape.
- LiPo batteries offer much higher capacities, allowing them to hold much more power.

The capacity of a battery is basically a measure of how much power the battery can hold. The unit of measure here is milliamp hours (mAh). The C Rating is simply a measure of how fast the battery can be discharged safely and without harming the battery. One of the things that make it complicated is that it's not a stand-alone number; it requires knowing the capacity of the battery to ultimately configure out the safe amp draw (the \C" in C Rating actually stands for Capacity).

Features:

- Rechargeable 3.7v 280ma battery
- Nominal Voltage: 3.7V
- Discharge cut off voltage: 3.0V
- Charge cut off voltage: 4.2V
- Charging Current (Std.): 0.2CA
- Charge Temperature: can be -40c +80c
- Charging Time (Std.): 3 4 hours
- Charging Time (Max.): 2 3.0 hours

Circuit Diagram:

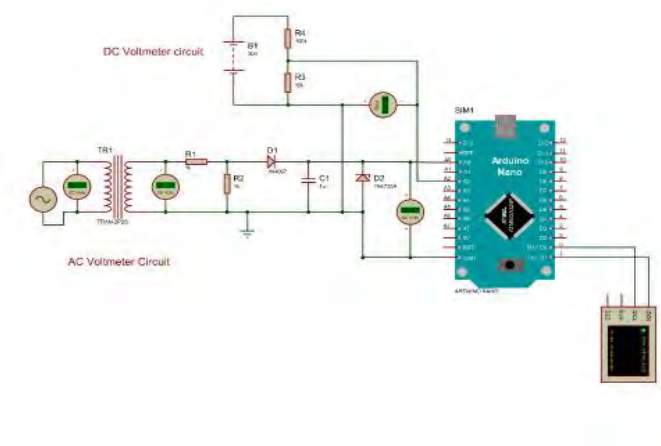


Figure 3: Transmitter side

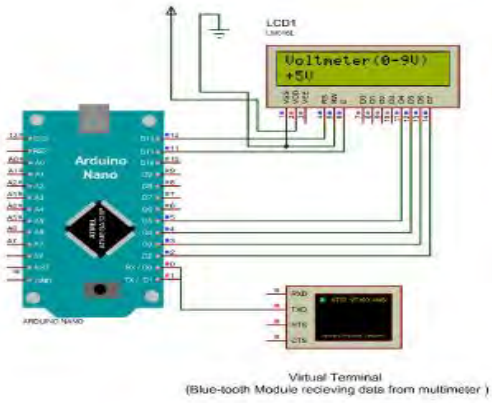
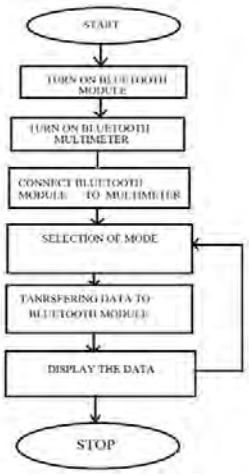


Figure 4: Receiver side

Flow chart:



Conclusion

Thus we have studied the smart glasses for measuring and displaying electrical parameters such as voltage. The simulation of same system are carry out using Arduino pro micro Proteus software. As it is difficult to observe readings on voltmeter of high voltage panels while measuring current or voltage at points. So smart glasses are preferred over voltmeter for observing readings at such high voltage area and also used where the user cannot perform multitasking such measuring the voltage points and observing readings on voltmeter screen. Due to use of Smart glasses" user can easily perform multitasking, this avoids user getting an electric shock. This reduces human efforts specially at high risk area.

References

[1] Anna Syberfeldt, Oscar Danielsson, PatrikGustavsson, Augmented Reality Smart Glasses in the Smart Factory: Product Evaluation Guidelines and Review of Available Products, IEEE Access Year: 2017, Volume: 5. https:ieeexplore.ieee.org

[2] Tomasz Kocejko ,Krzysztof Czuszynski, JacekRuminski, Adam Bujnowski, ArturPolinski, Jerzy Wtorek Extending touch-less interaction with smart glasses by implementing EMG module, 2017 10th International Conference on Human System Interactions (HSI).

[3] AliakseiAndrushevich, Martin Biallas, Rolf Kistler,JacekRumiski,AdamBujnowski, Jerzy Wtorek. Open smart glasses development platform for AAL applications, 2017 Global Internet of Things Summit (GIoTS).

[4] MgeGken; A. NuriBaolu; Marina Dabic, Exploring adoption of smart glasses: Applications in medical industry, 2016 Portland International Conference on Management of Engineering and Technology (PICMET).

[5] E. De Buyser; E. De Coninck; B. Dhoedt; P. Simoens, Exploring the potential of combining smart glasses and consumer-grade EEG/EMG headsets for controlling IoT appliances in the smart home,2nd IET International Conference on Technologies for Active and Assisted Living (TechAAL 2016).

[6] Paulo Alonso, Garcia Gaona. Virtual reality, UnivDistrital Francisco Jose de Caldas, Bogota, Colombia Published in: IEEE Latin America Transactions (Volume: 14, Is-sue: 6, June 2016) https:ieeexplore.ieee.org

[7] Anna Syberfeldt, Oscar Danielsson, PatrikGustavsson, Augmented Reality Smart Glasses in the Smart Factory: Product Evaluation Guidelines and Review of Available Products, IEEE Access Year: 2017, Volume: 5. https:ieeexplore.ieee.org

[8] Tomasz Kocejko ,Krzysztof Czuszynski, JacekRuminski, Adam Bujnowski, ArturPolinski, Jerzy Wtorek Extending touch-less interaction with smart glasses by implementing EMG module, 2017 10th International Conference on Human System Interactions (HSI).

[9] http://www.arduino.cc

[10] MgeGken; A. NuriBaolu; Marina Dabic, Exploring adoption of smart glasses: Applications in medical industry, 2016 Portland International Conference on Management of Engineering and Technology (PICMET).

[11] http://wiki.eprolabs.com

[12] Ali Emre Ok, Nuri A Basoglu. Zmir Institute of Technology, Dept of Engineering Man-agement, Urla, Turkey. Exploring the design factors of smart glasses, Published in: Management of Engineering and Technology (PICMET), 2015 Portland International Conference, publisher: IEEE. Publication year: 2015. https:ieeexplore.ieee.org

[13] Antoine Widmer , Roger Schaer, DimitriosMarkonis. HES-SO, Sierre, Switzerland. Facili-tating medical information search using Google Glass connected to a content-based medical im-age retrieval system, Published in: Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE. Publisher: IEEE, publication year: 2014. https:ieeexplore.ieee.org

Comparative Analysis of Different Predictive Models

Priyanka Bolinjkar, Niki Modi, Harshali Patil

Department of Computer Engineering Thakur College of Engineering and Technology, Mumbai,India
priyankabolinjkar33@gmail.com, nikimodi0102@gmail.com, harshali.patil9@gmail.com

***Abstract*—Predictive analytics encompasses a variety of statistical techniques from data mining, predictive modeling, and machine learning that anatomized current and historical facts to make predictions about future or otherwise unknown events. Predictive models exploit patterns found in historical and transactional data to identify risks and opportunities. There are various predictive models like regression, classification, clustering, decision tree etc which are used for predictive analysis based on their applications. In this paper the main aspiration is to perform the comparative analysis of different predictive models as in many areas such as marketing, performance, or fraud detection models. Predictive analytics provides a prognosticate score (probability) for each individual in order to determine, inform, or influence organizational processes that belong across large numbers of individuals, such as in marketing, credit risk imposition, fraud detection, manufacturing, healthcare, and government operations including law enforcement.**

Keywords—Predictive, Clustering, Classification, Regression, Decision Tree

I. INTRODUCTION

Predictive modeling is the process of creating, testing and validating a model to best predict the probability of an outcome. A number of modeling methods from machine learning, artificial intelligence and statistics are available in predictive analytics software solutions for this task. The models can be used to predict anything from sports outcomes and TV ratings to technological advances and corporate earnings. Predictive modeling is also often referred to as Predictive analytics, Predictive analysis, Machine learning.

These synonyms are often used interchangeably. However, predictive analytics most often refers to commercial applications of predictive modeling, while predictive modeling is used more generally or academically.

Machine learning is also distinct from predictive modeling and is defined as the use of statistical techniques to allow a computer to construct predictive models. Predictive modeling is useful because it gives accurate insight into any question and allows users to create forecasts. To maintain a competitive advantage, it is critical to have insight into future events and outcomes that challenge key assumptions.

II. PREDICTIVE MODELING PROCESS

This process involves running one or more algorithm on the data set where prediction is going to be carried out. This is an iterative processing shown below in figure [1] and often involves training the model, using multiple models on the same dataset and finally arriving on the best fit model based on business data understanding.

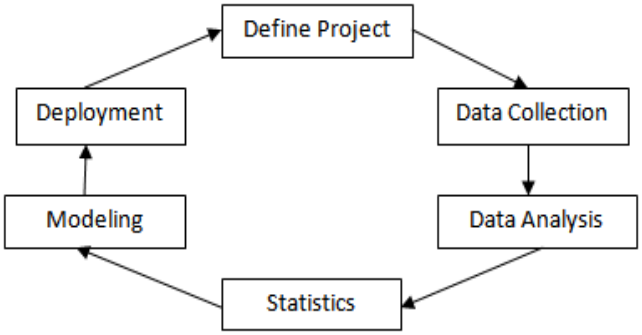


Fig 1: Predictive Modeling Process

1) **Define project:** Define the project outcomes, deliverable, scope of the effort, business objectives, identify the data sets that are going to be used.

2) **Data collection:** Data mining for predictive analytics prepares data from multiple sources for analysis. This provides a complete view of customer interactions.

3) **Data analysis:** Data Analysis is the process of inspecting, cleaning and modelling data with the objective of discovering useful information, arriving at conclusion

4) **Statistics:** Statistical Analysis enables to validate the assumptions, hypothesis and test them using standard statistical models.

5) **Modeling:** Predictive modeling provides the ability to automatically create accurate predictive models about future. There are also options to choose the best solution with multi-modal evaluation.

6) **Deployment:** Predictive model deployment provides the option to deploy the analytical results into everyday decision making process to get results, reports and output by automating the decisions based on the modeling.

III. TYPES OF PREDICTIVE MODELS

Predictive models come in all shapes and sizes. There are dozens, if not hundreds, of different methods that can be used to create a model, and more are being developed all the time. However, there are relatively few types of predictive models. The most common ones are:

A. Regression

- Multivariate
- Linear
- Logistic

- Polynomial

B. Classification

- Decision Tree
- Random Forest
- Naïve Bayes

C. Clustering

Regression:

Regression Algorithm which predicts continues variable based on other variable in the dataset. Regression models are the mainstay of predictive analytics. The focus lies on establishing a mathematical equation as a model to represent the interactions between the different variables in consideration. Depending on the situation, there are a wide variety of models discussed below that can be applied while performing predictive analytics.

1) **Multivariate:** Multivariate regression is a technique that estimates a single regression model with more than one outcome variable. When there is more than one predictor variable in a multivariate regression model, the model is a multivariate multiple regression. Multiple regression is an extension of simple linear regression. It is used when we want to predict the value of a variable based on the value of two or more other variables. The variable we want to predict is called the dependent variable (or sometimes, the outcome, target or criterion variable).

2) **Linear:** Regression is basically a statistical approach to find the relationship between variables. In machine learning, this is used to predict the outcome of an event based on the relationship between variables obtained from the data-set. Linear regression is one type regression used in Machine Learning.The linear regression model analyses the relationship between the response or dependent variable and a set of independent or predictor variables. This relationship is expressed as an equation that predicts the response variable as a linear function of the parameters. The goal of regression is to select the parameters of the model so as to minimize the sum of the squared residuals.

3) **Logistic:** Logistic model, it is also called as sigmoid function as shown below in figure [2] which is basically a method which transforms information about the binary dependent variable into an unbounded continuous variable and estimates a regular multivariate model. Logistic regression takes input data and *squishes* it, so that no matter what the range of the input is from 1 and 0.

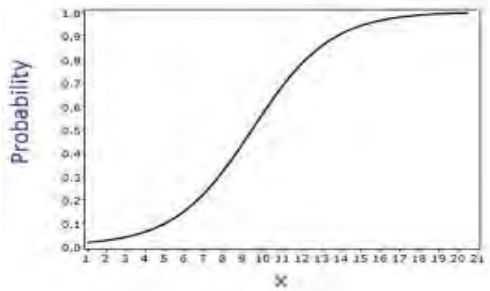


Fig 2: Logistic Regression (Sigmoid Function)

1) **Polynomial:** Polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted E(y |x), and has been used to describe nonlinear phenomena such as the growth rate of tissues.

Classification:

Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks. Classification models are tested by comparing the predicted values to known target values in a set of test data.

1) **Decision Tree:** Decision Tree algorithm classify and predict one or more discrete variables as shown below in figure [3] based on other variables in the dataset. Example: algorithm is C 4.5 and CNR type. Decision trees are formed by a collection of rules based on variables in the modeling data set: a) Rules based on variables' values are selected to get the best split to differentiate observations based on the dependent variable. b) Once a rule is selected and splits a node into two, the same process is applied to each "child" node (i.e. it is a recursive procedure)

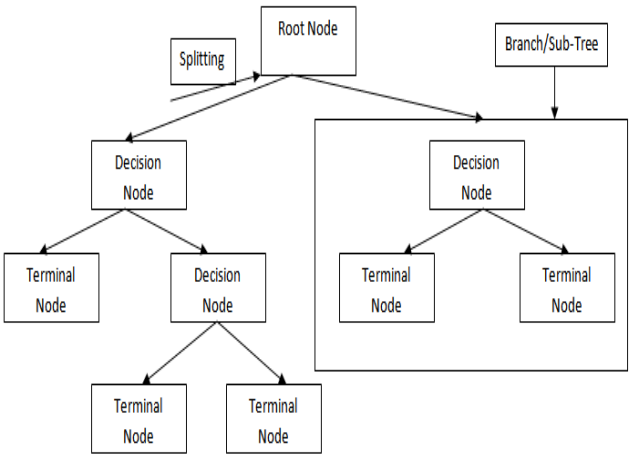


Fig 3: Decision Tree

2) **Random Forest:** Each branch of the tree ends in a terminal node. Each observation falls into one and exactly one terminal node, and each terminal node is uniquely defined by a set of rules. A very popular method for predictive analytics is Leo Bierman's random forests.

Random Forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because it's simplicity and the fact that it can be used for both classification and regression tasks. Random forests are made of many decision trees. They are ensembles of decision trees, each decision tree created by using a subset of the attributes used to classify a given population (they are sub-trees).

3) **Naïve Bayes:** Naive Bayes is a classification algorithm for binary (two-class) and multi-class classification problems. The technique is easiest to understand when described using binary or categorical input values. It is called naive Bayes or *idiot Bayes* because the calculation of the

probabilities for each hypothesis are simplified to make their calculation tractable. Rather than attempting to calculate the values of each attribute value.

Naïve Bayes are the probabilistic classifier based on applying Bayes' theorem with strong(naïve) independent assumptions. Naïve Bayes based on Bayes conditional probability rule is used for performing classification tasks. Naïve Bayes assumes the predictors are statistically independent which makes it an effective classification tool that is easy to interpret. It is best employed when faced with the "curse of dimensionality" problem, i.e. when the number of predictors is very high.

Clustering

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups as shown below in figure [4]. It is basically a collection of objects on the basis of similarity and dissimilarity between them. For Example- The data points in the graph below clustered together can be classified into one single group. We can distinguish the clusters, and we can identify that there are 3 clusters in the below picture.

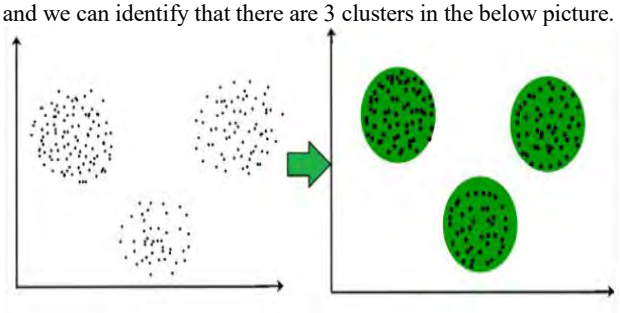


Fig 4: Clustering

IV. COMPARISON OF DIFFERENT PREDICTIVE MODELS

Features	Comparison of different predictive models		
	Regression	***Classification***	***Clustering***
Definition	It is set of statistical for estimating the relationship among varibales.	Is a process related to categorization the process in which ideas and objectives are recognized	Is a process partitioning a set of data into a set of meaningful subclasses.
Perforamnce	Slowly and use more memery	Performance on test data is process for performance on unseen data	Performance on external evaluation based on previous data set.
Complexity	Om^Wn	More Complex	Less complex than classification
Types	Multivaient ,Linear,Logistic,Polynoimal	Decision Tree, naïve Bayes, Random Forest	k-means
Benefits	Widely used,runs fast,highly interpretable	It is easy to understand and easy to implement, training is very fast	It is adaptable to change &helps single out useful features, It is important in data analysis
Leraning Algorithm	Supervised	Supervised	Unsupervised

Features	Comparison of different predictive models		
	Regression	***Classification***	***Clustering***
Data Output	Continuous data output	Predict discrete no of values	Partitioning data set into gap
Realibility	Less Relable	High reliable and accurate and reproducible, and highlight the potential use	Reliable for data grouping in hierarchical method
Accuracy	Logistic 75:25	Decision Tree: 86:79	-
Application/ Domain	Business are forecasting and Optimization	Credit approval, medical diagnosis, target marketing	Pattern recognition , image analysis
Prediction Evaluation	Ordered Data using Root mean square error	Unordered data measuring accuracy	clustering analysis doesn't have a solid evaluation metric

V. CONCLUSION

The main purpose of our paper is to compare the different predictive models based on different machine learning and data mining algorithm i.e. Regression, Decision tree, Classification and Clustering. So the comparative analysis on these models based on various features likes market trends, performance, risk imposition, quality, the main application in different areas and different domain.

ACKNOWLEDGMENT

We hereby take the privileged to present our paper on comparative analysis of different predictive models. We are very grateful to our paper guide Ms. Harshali Patil for contributing her valuable moments in the paper from her busy and hectic schedule. Thank you for being after us like a true mentor and great academic parents.

We are very thankful to Ms. Harshali Patil whose guidance and support was an immense motivation for us to carry on with our paper. Her suggestions have greatly contributed for the betterment of our paper.

REFERENCES

[1] J. Dem˘sar. Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research, 7:1–30, 2006

[2] Weiss, Sholom M. et al., Predictive Data-Mining: A Practical Guide. San Francisco, Morgan Kaufmann (1998)..

[3] Draper, N., and H. Smith, Applied Regression Analysis, 2nd edition. New York: John Wiley (1981), pp. 307-312..

[4] Dash, M., and H. Liu, "Feature Selection for Classification," Intelligent Data Analysis. 1:3 (1997) pp. 131-156.

[5] PrasadaBabu, S.HanumanthSastry." Big Data and Predictive Analytics in ERP Systems for Automating Decision Making Process". 2014. IEEE

[6] Hina Gulati. "Predictive Analytics Using Data Mining Technique". 2015. IEEE

[7] Friedman. N., Geiger, D., and Goldszmidt, M., "Bayesian classifier", Machine learning, Vol. 29, 1997, pp.131- 163

[8] J. R. (1986). Induction of decision trees. Machine Learning, 1, 81-106.

[9] M. L., Patell, J. M. & Wolfson, M. A. (1984). The experimental design of classification models: An application of recursive partitioning and bootstrapping to commercial bank loan classifications. Journal of Accounting Research, 22, 84- 113.

[10] C. F. (2009). Feature selection in bankruptcy prediction. Knowledge-Based Systems, 22(2), 120-127.

[11] Z. S. Hua et al., "Predicting corporate financial distress based on integration of support vector machine and logistic regression," Expert Systems with Applications, vol. 33, no. 2, pp. 434-440, 2007.

[12] M. Y. Chen, “Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches,” Computers & Mathematics with Applications, vol. 62, no. 12, pp. 4514-4524, December 2011.

[13] C. F. Tsai, Y. F. Hsu, and D. C. Yen, “A comparative study of classifier ensembles for bankruptcy prediction,” Applied Soft Computing, vol. 24, 977-984, 2014.

[14] M. Wagle, Z. Yang, and Y. Benslimane, “Bankruptcy prediction using data mining techniques,” in Proc. 8th International Conference on Information and Communication Technology for Embedded Systems, May 2017, pp. 1-4.

[15] K. (2008). Comparison of data mining techniques used to predict student retention. Ph.D. dissertation, Nova Southeastern University United States, Florida. Retrieved January 9th, 2010, from Dissertations & Theses @ Nova Southeastern University.

All Clean- Autonomous Cleaning Bot

Shrinidhi Gindi, Saad Naik, Safa Ansari, Sufiyan Shah
Information Technology MHSSCOE, Mumbai, India
shrinidhi.gindi@mhssce.ac.in, Saad.naik20@gmail.com, safa.ansari.a.q@gmail.com, sufiyan764@gmail.com

***Abstract*—AllClean is an autonomous cleaning robot equipped with a robotic arm in association with the cleaning unit. The cleaning unit comprises of suction mechanism and the floor cleaning mop. AllClean will provide an automated cleaning solution for today's technology driven world. The robotic arm empowers AllClean to pick and place garbage objects, differentiating the bot from its competitors. AllClean is empowered by two powerful programming languages which are Python and C and uses Raspberry pi for its main processing and algorithmic workflow. AllClean uses Opencv image processing library for performing operations on images captured by the cameras. AllClean is designed in such a way that it hides all the complexity of different cleaning mechanism from the user. Its design makes sure that the robot acquires minimum space and is light weighted; such factors make it highly compatible and user friendly.**

Keywords: Autonomous Cleaning Robot, Robotic Arm, Camera based garbage picking, Cost Efficient.

I. INTRODUCTION

This project aims at developing an all purpose autonomous cleaning robot AllClean which will be used to clean flat surface environments and places such as monuments, halls, malls from garbages ranging in size right from dust/dirt particles to thrown water bottles. The robot will be equipped with a robotic arm in association with the cleaning unit. The cleaning unit comprises of suction mechanism and the floor cleaning mop. The suction mechanism consists of a vacuum suction unit which cleans surfaces from small size dust particles. The floor polishing adds to the systems performance and gives it an expert cleaning touch, it comprises of a rotating mob attached beneath the robot following the suction mechanism such that the floor gets polished after it is cleaned from garbage particles and even objects. The bot has two wheels and a support wheel giving it optimum balance and load handling capabilities. AllClean turns either left or right about one particular axis of rotation allowing the bot to avoid loss of cleaning space and ensuring no floor space is left dirty. The robotic arm empowers AllClean to pick and place garbage objects, differentiating the bot from its competitors.



Fig. 1: AllClean

AllClean being autonomous would reduce the workload and overheads for authorities responsible for maintaining monuments and cut down cleaning expenses for companies who spend millions of dollars over the years in keeping their premises clean.

II. LITERATURE SURVEY

In this chapter the aim is to review the literature of cleaning robots. Moreover, concept and definition of the cleaning robots is reviewed and defined for the purpose of this study. From early human civilization human is increasingly dependent on the machines. If we look at past 30 years we will see robotics from large structure going to small and smaller in nano range. If we look at the cleaning robot we can see iRobot and some LG Hom-Bot are dominating the market.

LG Hom-Bot is robot vacuum cleaner which provides wide variety of cleaning tasks, because modern homes or any other place requires cleaning versatility. It has two cameras one on top and other on bottom. Top camera is used to mapping the room and bottom camera scans the floor for moving distance enhances precision and speed. iRobot Roomba vacuum cleaner it similar to LG Hom-Bot in design, but it does not have camera to map out the room or area to clean.

The locomotion of this robot is guided by sensors placed strategically in the left, right and front of the robot or digital signal processors which help the robot to make decision to avoid obstacles present in its current path. But both of them cannot clean the garbage which are bigger in size, these are only used to clean the dust from cleaning area. It will consider as it is obstacle.

All Clean Bot is the project we are currently working on, which can be used for both cleaning the dust and picking the garbage like bottle, paper. It has suction unit which is used to clean the dust or any particle on the floor. The opening of the suction unit is placed in front of the robot. It has robotic arm which is used to pick and place the garbage which comes under it range of picking. The arm is placed in top of the robot. It has two cameras, one of the camera is placed in front which is used to find the path for motion with obstacle. Other placed in the arm to send the coordinate of the object and to find if the object is in range of arm. This processing and signals is done in raspberry pi.

III. EXISTING SYSTEM

A. Roomba Robot

Roomba is a product of iRobots launched in September 2002 .It is an autonomous vacuum cleaner having few similar characteristics as AllClean as far as cleaning of small dust particles in concerned. Equipped with a set of features that help its navigation process while working in houses or similar small environments. Roomba includes actions right from detecting dirty spots on your floor cleaning them, making sure to avoid falling from steep drops and also obstacle avoidance. In order to swipe of the debris from the ground Roomba uses a rotating, 3-prong spinner brush .The two independent wheels placed on either side make sure that Roomba can rotate a complete 360° angle. It is powered by a removable NiMH battery which uses a wall power adapter to recharge itself. However recent models have also been equipped with self charging units/homebase. These units are separately placed and solely perform the function of recharging once the robot docks on it .Roomba has a recharge time of 3hrs.

B. LG Home-Bot

You can't fit a round robot in a square corner, so the Hom-Bot has contoured corners and longer side brushes to reach dust where it hides. LG's Home-Bot is set apart from its other competitors as it uses an advance mapping program known as the Dual Eye 2.0 mapping system which maps any home's layout the robot is currently operating in.

Collisions are avoided using a bumper which ensures Hom-Bot's safe motion also the Auto-Resume process allows the Hom- Bot to continue its process in case of any interruptions. Inclusion of seven different custom cleaning feature's, each designed to perform its own suited task and self-learning characteristic for identifying and avoiding furniture as well as other household equipments results in less bumping into undesired objects . LG Hom-Bot has a user friendly procedure for removal of dirt collected during the cleaning process.

IV. METHODOLOGY

The development of this project includes creation of an autonomous cleaning bot (AllClean) meant for all purpose cleaning of places having flat surfaces such as monuments, halls. It will involve units with shared cleaning responsibilities such as removing dust particles to polishing the surfaces to picking and placing of objects involving the use of a robotic arm. The robot will use Raspberry pi as its main processing unit along with Arduino for base movement configuration. Python and C will be the programming language used. The robot will be able to pick and place small size object with weight less than 200g, which will come in its path and is classified as pickable by AllClean's algorithm else the robot will move to next free path it perceives. The robot will be equipped with two web cameras having the role of capturing stream of images for different purposes which are garbage detection detection and path calibration; respectively. The robot will not include a self garbage disposal mechanism and will depend on the user to dispose the collected garbage.

A. Locomotion of the Robot

AllClean's base is supported by three tyres, one ahead and two at the back, the tyres at the back are connected to 30 rpm DC motors each. The tyre ahead is not connected to any motor; it has a free movement which supports the bot in turning left or right. The default motion of bot is forward. The bot will initiate its motion by taking a picture by the camera located ahead on the body. If the picture contains garbage and the garbage is in some specific perimeter range, then the robot will move towards it until the bot reaches a distance from where it can be easily picked by the arm. The perimeter algorithm is as follows:

1. Start
2. Front webcam captures images.
3. If (Object detected and perimeter range is between 500 and 1000) then
   Start Robotic Arm
   else
   Turn by angle θ and move forward.
4. Not done with cleaning? Go back to step 2
5. Stop

The mechanism used for turning left or right is based on single axis rotation. For turning left, the right tyre motor is accelerated and the left tyre motor is at halt, (as shown in the fig. 2.), with respect to Z- axis the body will move about left tyre.

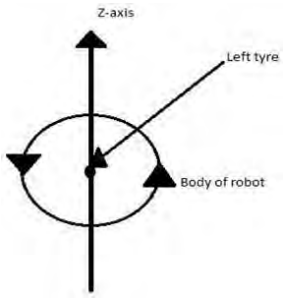


Fig 2: Mechanism for turning left

And vice versa for turning right the body will turn about right tyre (as shown in the fig.3).

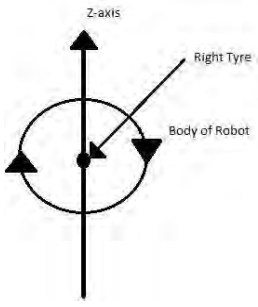


Fig 3: Mechanism for turning right

B. Circuitry

a) *Raspberry pi:* The core component of AllClean circuit is raspberry pi and what actions to be performed by arm and wheels are send using its gpio pins. We have also used noob os to have interface with electronic component and program them according to our need. We have used python and C language to code the program for motion of arm using servo driver PCA9685 and wheel motion using arduino and motor driver.

b) *Servo Driver:* Two servo drivers connected in series with each other to avoid them from overloading. Servo driver is connected to the raspberry pi via gpio pins and motor is connected to the servo driver.

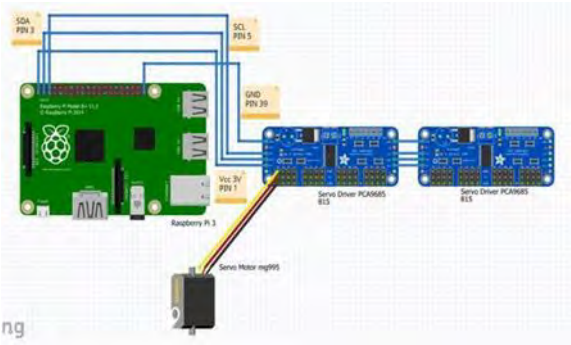


Fig 4: Connection between the raspberry pi and servo driver

c) *Arduino and Motor driver:* Arduino is connected to the raspberry pi and it sends the pwm signals to motor driver via pwm pins to control the motion of dc motor for wheel and mop rotation. We have used three motor drivers two for wheels and one for mopping.

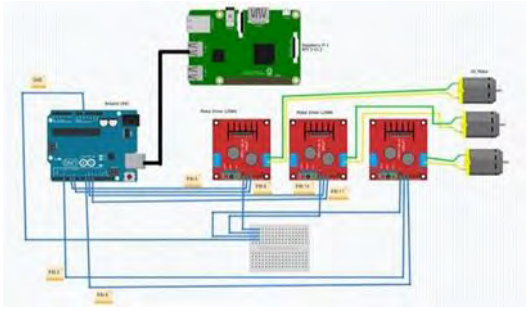


Fig 5: Circuit diagram for wheel and mop Rotation

C. Design of the Robot

a) *Robotic Arm:* The arm consists of 3 links which are attached to one another with 2 servo motors each container. A flexible plastic pipe is attached in front of the container and the motor is attached at the back side. There are two halves of the vacuum; one half is equipped with the motor fan while the other half is used for storing the dust particle.

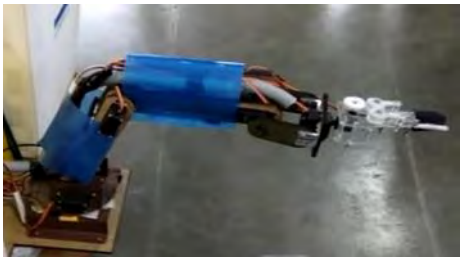


Fig 6: Robotic Arm

We have used MDF (Medium Density Fibreboard) sheets for making links because it's light weight, durable and cost efficient. All the links are laser cut precisely to maintain flexibility of the arm.

There are 9 motors used in the entire arm including gripper, 8 SERVO mg995 and 1 SERVO mg 90s.

Maximum and Minimum reach of the arm with gripper from the track is 40.5cm and 10.0cm and average weight the arm can lift in one pick is 100gm.Area coverage in one rotation of body is 4836.385sq.cm.Maximum opening of the gripper is 11cm.

b) *Camera Positioning:* We have used two cameras one is attached to the 2nd link of the robotic arm and another one is attached at the front of the robot so that it can detect the object and navigate easily.



Fig 7: Camera Positioning

c) *Gripper:* Static parts of the gripper are made entirely of acrylic while the dynamic parts such as gears are made with both acrylic and Teflon. The gripper can pick up any object of size ranging from 1cm to 10cm. we have used Teflon for making gears because it is highly durable and force absorbent. It protects the gears from breaking.

D. Vacuum

The vacuum is equipped with a motor fan which is used to suck small dust particles into the container. A flexible plastic pipe is attached in front of the container and the motor is attached at the back side. There are two halves of the vacuum; one half is equipped with the motor fan while the other half is used for storing the dust particle.

E. Mop

It will be used to clean the floor completely after the vacuum has sucked in the dust particles. It works with the help of a DC motor. The shaft of the motor is attached with a corrugated plastic sheet of circular shape and a soft wool material cloth is pasted on that circular sheet. The rotary motion is very effective and fast way to clean the floor. Motor resides inside the body of the robot with the shaft and the circular mop outside the robots body.

F. Block Diagram

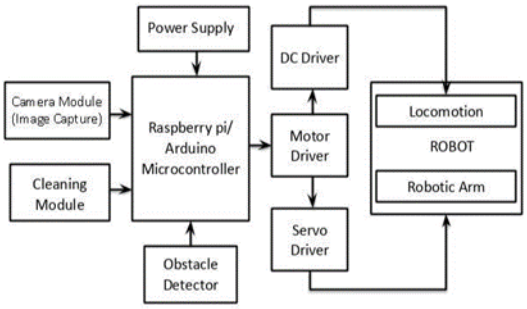


Fig 9: Block diagram of the proposed project

G. Flow Chart

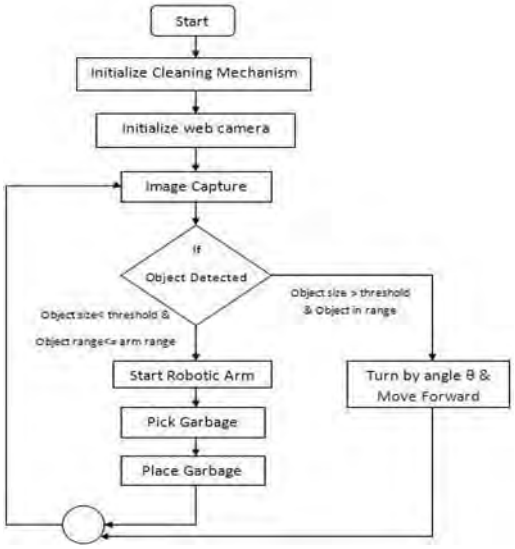


Fig 10: Flow chart of the proposed project

V. RESULT AND DISCUSSION

The purpose of the project was achieved to some extent. Cleaning process done efficiently. The garbage pickup range is minimum 10cm to maximum 40.5cm. The project is still in progress to achieve best results.

To hire househelp for picking up, mopping and suction of garbage is costly and time consuming work. AllClean cleans all surfaces at the push of a button. It is super convenient; at only inches tall it can easily work and absorb dirt in hard to reach places. The other capabilities that will be added to AllClean include the addition of a charging circuitry. Modification and further better advancements in the design and material that are currently being used for the robot. Work in the direction of enhancement of image processing logic of the robot for better object detection and classification. Improving the robots path detection strategies. Use of more efficient motors in the robotic arm which will make it possible for the arm to even pick heavier objects. Making the robot more user friendly by providing a mobile application through which any user can track and manage AllClean and also improving AllClean's power consumption.

REFERENCES

[1] International Journal of Internet of Things 2017; 6(2): 40-42 doi:10.5923/j.ijit.20170602.06

[2] International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 6 Issue 04, April-2017

Emergency Response System Using Smartphone

[1]Lallan Pandit, [2]Sankalp Mukadam, [3]Samant Sahani, [4]Rishabh Pathak, [5]Guide- Prof. Satish M. Ranbhise
Information technology Shree L. R. Tiwari College Of Engineering(Mumbai university) Mumbai,India
[1]lallanpandit3112@gmail.com, [2]sankalpmukadam.sm@gmail.com, [3]samantgsahani@gmail.com, [4]rishabhtesla@gmail.com, [5]ranbhise1976@gmail.com

***Abstract*—As nowadays the population in India is increasing drastically also with the advance features and opportunities in technology. Also, due to rise in various fields, it gives rise to number of problems and also provides us with the opportunity to overcome them. But, it is necessary to make proper use of the available resources. In today's world everyone has a super-minicomputer (smartphone) in his/her pocket which is easily available to everyone. Instead of such a great technology it is not used up to its fullest efficiency. This project introduces us with solution to this problem with the help of geolocation and web application. For example, if there is an accident in a particular area where the help is not available easily, with the help of this application you can get help with just one click and rest is done by the application. It grabs information about the nearest hospitals and other help and takes the required actions. Thus, this paper discusses the various issues and also gives an idea of the proposed system, its working details and its benefits in other application too.**

***Keyword*—**
Longitude,Latitude,Camera,Hospitals,Android,java, EMS (Emergency message service).

I. INTRODUCTION

When disaster strike, traditional communication and cellular access points may no longer be available, denying communication when it is most needed. Recent Disasters such as Kerala Flood, and many more in past such as Indian Ocean, Gujarat Earthquake have caused unprecedented damage and massive loss of life, and unveiled the necessity of immediate ad-hoc communication solutions for rescue forces and personnel. The research community has invested in the construction of management systems for communications at rescue areas, designing effective systems that provide timely access to comprehensive, relevant, and reliable information that is critical to humanitarian operations, as well as operations management. Various emergency handling services are present worldwide to cater to the emergencies faced by the common man. The USA has its own such system by the name 911.At present, India does not have a such similar well defined emergency-handling approach. EMS is developed keeping in mind the lack of any such system in India. It is loosely based on 911 system, but suited more to India. EMS is responsible for leading the department in providing a properly planned and organized emergency management force, capable of responding to public emergencies. The people in an emergency or anybody at the emergency site will Send or share EMS at avail the service. The common man can make use of this system in case of any emergencies. The system works is such a way, that victims will receive help in the form of police, fire, ambulances or fire-brigades depending on the situation. This system drives to and responds to emergencies promptly

II. RELATED WORK

Emergency Response Application is a messaging application for emergency services providing integrated medical, police and fire emergency services in Indian. Application has GPS System for location Tracking and Camera for taking the Images and shares the Messages to get the appropriate resources. Application also Shares the same information to the victim Family member and the nearest Helping hands. It also gives the alert message to the Victims to getting the resources.

III. FLAWS IN CURRENT SYSTEM

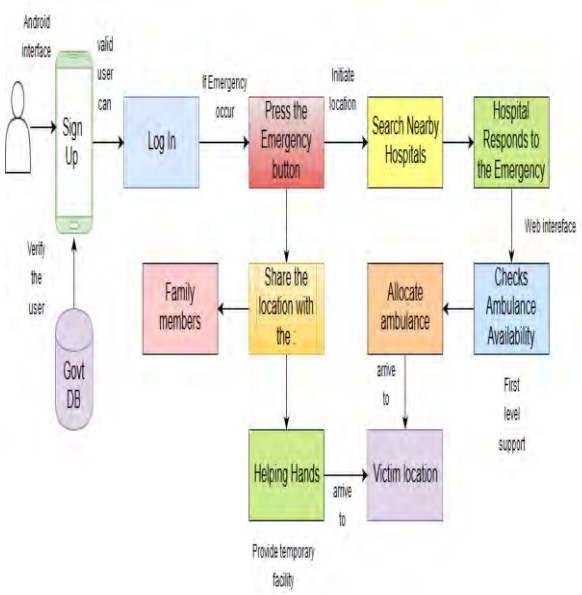
Application starts with the Validation of user by using his email id, mobile no, adhaar card. After Validation itself user Press the Help button then user has to select the appropriate emergency such as accident, fire, etc. When the user clicks on any emergency button at the moment victim location and his images sends to the nearest resources Such as hospital, fire bridge, etc. The information is also sends to the victim family member, and Nearest Helping Hands. This information goes to all the Hospitals nearest to the Victims Location. The moment the hospitals get the message it checks the resources availability. If resources are there it responds to the message that resources will be provided to the victim. This Messages get shared to all the hospital that resources is provided to the victim. This application provides the required resources in efficient time. Just by one click you will help any victims without any losses of time and effort.

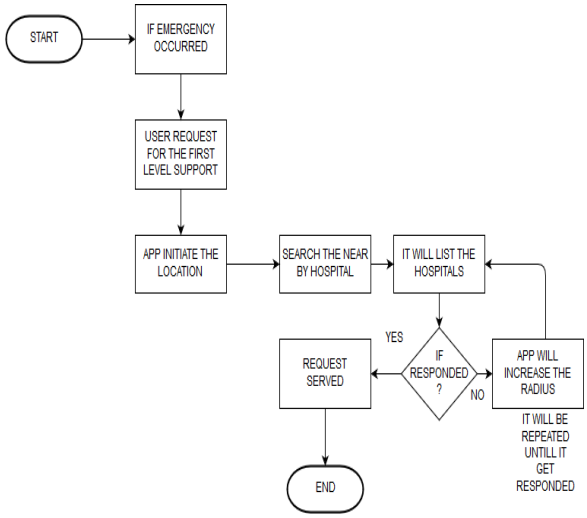
IV. FEATURES OF EMS

[1]. EMS system uses GPS for location tracking.
[2]. Authorized person will be responsible for rescuing the victim.
[3]. Report can be generated Area wise, Issues wise, Location wise, and Priority wise.
[4]. Location tracking will be done in Network mode also if the GPS is disable.
[5]. EMS system is useful in any type of emergency that would occurred in society.
[6]. EMS system is very cost effective as compared to other emergency system.
[7]. Admin will be able to send the emergency resources.
[8]. Victims Location is also shared to his family member and also to the Helping hands (First Aid).
[9]. At the moment the Location Shares it search the nearest Hospitals.
[10]. Increases the radius of Searching if nearest resources is not available (Scalability).

[11]. If appropriate hospitals responded to the message it gives the other hospitals message that Victims get Resources.

V. BLOCK DIAGRAM OF THE PROPOSED SYSTEM





VI. DESIGN AND IMPLEMENTATION

User Application: User Raise the issue in case of any emergency from his Android Application with description and nature of issue severity. User's location tracking is done with the help of GPS is OFF then location tracking is done in Network mode. All the details of user like Title, Description, Priority and Current location report is send to the server for further assistance. Server respond location based Emergency Contact Nos to the user application.

Admin Application: Admin gets notifications n his Android application with Emergency Issues. Admin can also view the person details and his/her location that has raised the issue. Admin can view the optimal path on Google map with marker to reach the victim who raised the issue in case of emergency. After resolving the issue, admin can mark the issue as resolved with certain comments.

Web Based Application for Reporting: Web Application is useful for generating the reports such as, Area wise issues report, Priority wise issues report, Location wise issues report.

VII. FUTURE WORK

We are implementing this application using GPS System network currently. If resources permit, we can implement this using cloud computing techniques as a future work. This would give a leverage to global access by expanding the scope of the application from an ad-hoc network to, say WAN.

Algorithm

The algorithm used in this application is based on Floyd Warshall algorithm.

Floyd Warshall Algorithm:

The Floyd Warshall Algorithm is for solving the All Pairs Shortest Path problem. The problem is to find shortest distances between every pair of vertices in a given edge weighted directed Graph.
We initialize the solution matrix same as the input graph matrix as a first step. Then we update the solution matrix by considering all vertices as an intermediate vertex. The idea is to one by one pick all vertices and updates all shortest paths which include the picked vertex as an intermediate vertex in the shortest path. When we pick vertex number k as an intermediate vertex, we already have considered vertices {0, 1, 2, .. k-1} as intermediate vertices. For every pair (i, j) of the source and destination vertices respectively, there are two possible cases.
1) k is not an intermediate vertex in shortest path from i to j. We keep the value of dist[i][j] as it is.
2) k is an intermediate vertex in shortest path from i to j. We update the value of dist[i][j] as
dist[i][k] + dist[k][j] if dist[i][j] > dist[i][k] + dist[k][j]

VIII. CONCLUSION

The mobile device has been positioned to become the world's leading solution for every problem. The main aim of the project is to minimize the sorrow and pain a family would go through due to the loss of the beloved once. GPS technology is used so that the hospital can take immediate action which might reduce the severity. This system is accurate and the main advantage is that it saves a lot of time. It also overcomes the drawbacks of the previous systems and inherits the advantages of them. Thus, it can be considered as an improvised version of the previous systems and the existing system.

REFERENCES

[1]. Osnat (Ossi) Mokryn Dror Karmi Akiva Elkayam Tomer Teller\Help Me: Opportunistic Smart Rescue Application and System"," 2012 The 11th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net).

[2]. D. Horowitz and S. Kamvar, The anatomy of a large-scale social search engine, in Proceedings of the 19th international conference on World wide web. ACM, 2010, pp. 431–440.

[3]. J. Fajardo and C. Oppus, A mobile disaster management system using the android technology,"WSEAS TRANSACTIONS on COMMUNICATIONS, vol. 9, no. 6, pp. 343–353, 2010.

[4]. A. Pietilainen, E. Oliver, J. LeBrun, G. Varghese, and C. Diot, MobiClique:middleware for mobile social networking, in Proceedings ofthe 2nd ACM workshop on Online social networks. ACM, 2009, pp.49–54.

[5]. B. Van de Walle, G. Van Den Eede, and W. Muhren, Humanitarian information management and systems, Mobile Response, pp. 12– 21,2009.

[6]. T. Catarci, M. de Leoni, A. Marrella, M. Mecella, B. Salvatore, G. Vetere,S. Dustdar, L. Juszczyk, A. Manzoor, and H. Truong, Pervasivesoftware environments for supporting disaster responses, IEEE InternetComputing, pp. 2637, 2008.

[7]. N. Tolia, M. Satyanarayanan, and A. Wolbach, Improving mobiledatabase access over wide-area networks without degrading consistency, in Proceedings of the 5th international conference on Mobile systems, applications and services. ACM, 2007, pp. 7184.

[8]. J. Su, J. Scott, P. Hui, J. Crowcroft, E. De Lara, C. Diot, A. Goel, M. Lim, and E. Upton, Haggle: Seamless networking for mobile applications, in Proceedings of the 9th international conference on Ubiquitous computing. Springer-Verlag, 2007, pp. 391408.

[9]. P. Currion, C. Silva, and B. Van de Walle \" Communications of the ACM, vol. 50, no. 3, pp.61-65, 2007.

[10]. J. Scott, P. Hui, J. Crowcroft, and C. Diot, Haggle: A networking architecture designed around mobile users, IFIP WONS, vol. 2006, 2006

Product Tracking System using RFID and Prediction Algorithms

Pooja Yadav, Kamini Nawale, Isha Raval, Lydia Suganya

Computer Department, Thakur College of Engineering & Technology, Kandivali (East), Mumbai-400101, India
poojayadav2124@gmail.com, kaaminiiknawale@gmail.com, iraval2772@gmail.com, lydia.jacob@gmail.com

Abstract-Effective inventory management plays an important role in keeping track of when a stock is going in and out that can help lower costs, speed up fulfillment, and fraud prevention. Sometimes it might happen that demand is more and supply is less or vice-versa. So, to maintain a proper balance between supply and demand we have come up with an idea of a product or stock tracking system.

The main aim of this project is to predict the next month's sales of a particular product by considering the past and current sales data by using various prediction algorithms, count the number of products sent by the company in the warehouse, keep a track of the product count dispatched from the warehouse. Use the RFID to get the data of the product for Quality assurance.

Keywords-Prediction algorithm, Inventory, stocks, Linear Regression, analysis, Quality assurance.

I. INTRODUCTION

Inventory management plays a very vital role in any company's growth. For many companies, inventory management is one of the most complicated processes. They need to decide how many products they need to stock and avoid excess inventory. Excess inventory can incur possible financial losses.

So to maintain a proper balance between supply and demand and to avoid excess inventory, we have come up with a solution wherein customers can place their bulk orders through a website. Once the company gets the order and its details from their database which signifies the total sale they will use that data to predict the next month's sales using machine learning algorithm in order to avoid overstocking and understocking.

II.THEORY

Django Framework

Django is a Python-based free and open-source web framework, which follows the model-view-template (MVT) architectural pattern. It is maintained by the Django Software Foundation (DSF), an independent organization established as a non-profit. Its primary goal is to develop a complex database driven website the front end of our project. We have developed a website using Django framework for making it easier to order products. As customers are added as well in the entire process that eliminated the use of RFID in the proposed model hence, it is a cost efficient product.

Linear Regression is one of the simplest algorithms for predicting future values based on past values. Hence a comparative study has been conducted by using different algorithms for predicting future sales. Company support, Warehouse involvement, and commitment are key elements required to gauge managerial feasibility in the proposed project. The success and the profitability of the project partly depend on managerial competence of the major ingredients of the proposed project which are Warehouses. The capability of the infrastructure of a process is to achieve and sustain the properties of isolation, atomicity, durability, consistency etc in the matter of the data stored of the products. Also, constant attention is required in the case of the number of log files being created.

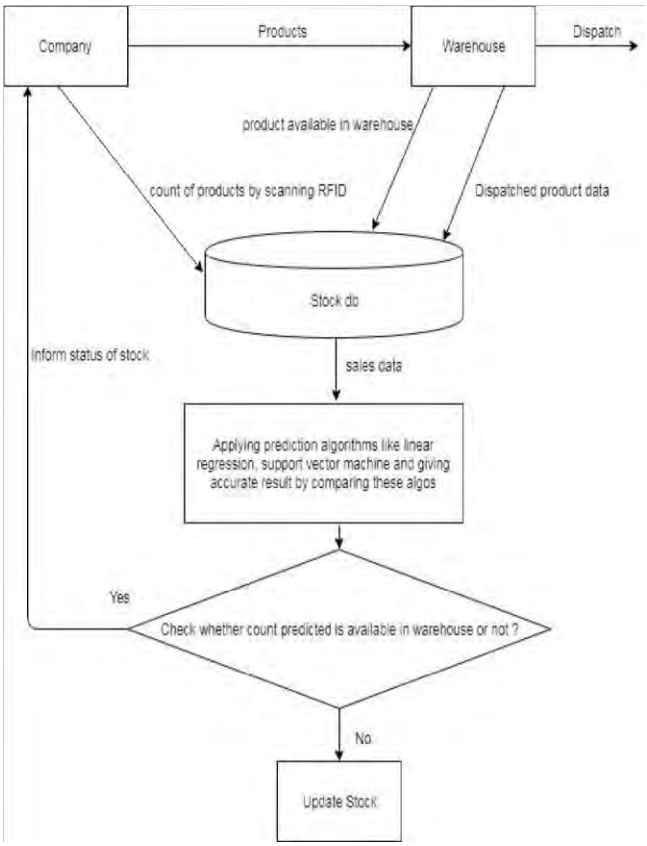
The proposed system is concerned with the sales prediction. This system will predict next month's sales on the basis of the current and previous month's delivery record of a particular product. our proposed system consists of various other features and modules such as it can predict sales of a particular product or overall sales prediction of next month for a particular company.

III.PROJECT WORK

Algorithm

- Order the number of product through the website (this task is to be performed by the customer).
- Order details will be stored in the database(backend).
- Using the data from the database predict future sales by applying various algorithms.

Flowchart



Literature survey

1. Packages Tracking using RFID:

System that transmits the identity of an object using wireless media while applying RFID. The main theme of the system is to distinguish between objects' categories. Throughout the paper we proposed an efficientackage tracking system that increases speed and accuracy and at the same time save money, as well.

2. Algorithm for the procurement and inventory management in the distributed

3. Supply chain:

The algorithm of procurement and inventory management in the distribution network and recommendations for improving the system of inventory management.

IV. RESULTS AND DISCUSSION

In product Tracking System customers will give order for products via a web portal. Total product order will be collected by the company and accordingly products will be sent to the warehouse after scanning them through RFID. Total count of the product will be kept with the company and the warehouse for future use. When the product will be dispatched for the warehouse they will be scanned again and company will be notified about this. All these count and dispatched data of current month and previous sales will be used in order to predict next month's sales using an algorithmic approach.

V. CONCLUSION

The project tries to maintain a balance between supply and demand and predicting sales in order to reduce stocking cost. it can be stationed at various organizations to maintain their inventory and track the stocks available in the warehouse. Different companies and organizations maintain a lot of products on a daily basis with this system their efforts will be reduced. The probability of error and time consumed will be reduced by a considerable proportion with the implementation of the system.

VI. FUTURE SCOPE

- RFID can be used for product Quality assurance by collecting product information.
- Live tracking of routes while it is being dispatched.
- More prediction algorithms can be used for getting more accurate results.

VII. ACKNOWLEDGEMENT

- We sincerely thank our guide Mrs. Lydia Suganya for her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work. We thank the HOD, Dr. Sheetal Rathi,
- Dean Academic Dr. R.R. Sedamkar, the Principal, Dr. B. K. Mishra, and the college management for their support.

REFERENCES

[1] http://opexanalytics.com/machine-learning-inv entory-optimization/
[2] https://www.unleashedsoftware.com/blog/usin g-machine-learning-inventory-management
[3] https://www.researchgate.net/publication/2895 05094_Packages_Tracking_Using_RFID_Tec
[4] hnology
[5] ttp://paper.ijcsns.org/07_book/200901/20090
[6] 105.pdf
[7] https://www.hindawi.com/journals/ijap/2013/5
[8] 31364/
[9] Distributed artificial intelligence , Agent technology and Collaborative Application (by
[10] vijayan sugumaran)

Antisocial Activity Detection Using Web Mining Intelligence

Ribah Shaikh, Sushmita Shinde, Rishabh Singh, Mrs. Purvi Sankhe
(Information Technology, Thakur College of Engineering and Technology, India)

Abstract: **There has been an increase in exorbitant use of internet in the past few years. This increase, has also driven to escalations in the number of illegal activities on the web with the help of internet. These antisocial activities have been causing menace and disturbance in the society and therefore there is a need to cease these antisocial activities for which the automatic detection of such activities is necessary. Antisocial activity detector will be a web portal where the user provides the link of the website, which he or she thinks is an illegitimate one. Initially the study of existing systems backed us in identifying their features and gaps which in turn helped in targeting the proposed system with an aim to provide a new system with high accuracy rates and lower costs. For the development of this project, python and some data mining algorithms for processing and classification of data will be used. The data in the link provided by the user, is scraped, pre-processed and classified. The scraping will be done with the help of web scraper. Once the scraped data is obtained, it will be preprocessed and classified with the help of pandas and scikit learn module of python. The result of the classification process generated is then provided to user. This result will then tell the user, whether the website is antisocial or not. The algorithm used in the project will provide accuracy on the amount of precise and correct data. This accuracy will range from 75- 85 %. The result provided by the system based on the percentage needs to be studied further by authorities and the necessary actions should be conducted.**

Keywords: Antisocial behavior, Data mining, URL detection, Python modules, Cyber security.

I. INTRODUCTION

Any unacceptable activity that has caused or was likely to cause harassment, distress or simply harm to an individual or a group, by a misconduct or by violating certain rules and regulation is an antisocial activity. It is broadly defined as an unconsidered action that may damage or cause evil to the society.Use of internet to carry out illegal activities like cyberbullying, influencing youngsters to join any of the terrorist groups, selling of faked products or offensive pornographic materials, etc. are some of the examples of an antisocial activity.In the past few years the percentage of people using the internet has excessively increased. This increase in the usage of internet has led to the growth of harmful content on the web resulting in an antisocial activity.In order to reduce the volume of harmful material on the web, there is a need for its automatic detection.Being able to automatically detect the negative content, will lead to reduction in the number of antisocial activities.

II. MATERIAL AND METHODS

Requirements

- Web browser
- Windows Operating System
- Windows XP, Windows 7,8
- MySQL Server
- Python 3.7 version

Methodology Used

For developing a project or software it is important to know the flow or which type of software development lifecycle model the system is going to use. Out of various SDLC models there is a need to choose the model which will be perfect for the project and therefore Iterative model has been selected.

As the name suggest Iterative model is basically an iterative process where the input is processed through various phases and if it doesn't satisfy it is repeated until it becomes satisfactory.

The model consists of five phases namely:

1. Planning.
   For developing any software or developing any project planning plays a crucial role as without planning a project will not react its maximum level. The planning phase deals with requirement of customers, understanding what the customer wants and making it available for the customer. It is also important to understand the purpose, scope and goal of the project. For developing the antisocial activity detection system, initially the system understands in what ways does the customer wants the result, and according to it the planning is done.
1. Analysis and Design.
   Once the requirements and purpose of the project is understood it is important to analyze and work on the design of the project.

Depending upon the requirement, the design of the project needs to be visualized. The design of the project should satisfy basic criteria like easy access tousers, low cost, etc. To decide the work flow of system various diagrams like DFD, use case,flow chart, etc. need to be created. In this phase the layout for the project has been designed along with DFD, UML and flow chart which help in providing better understanding of the proposed system.

2. Implementation.
   Once the design of system is ready the part of implementing it, i.e. coding begins. The implementation of system is to be carried out in a systematic manner. First the UI (UserInterface) is to be created with the help of web development languages namely HTML, CSS, BOOTSTRAP, etc. then the server-side processing is to be done. In this project the server-side processing is carried out with the help of python 3.7 version which is used for scraping and classification purpose and lastly the integration of both. After completing the implementation, the system is partially ready.
3. Testing.
   Testing plays an important role in any SDLC because without testing the product you cannot deploy it for user. Testing needs to be carried out in two ways black box and white box where theteam can understand both the scenarios and can solve the upcoming problems in order toimprove the system.
4. Deployment
   Once all the testing is done, the system is ready to be deployed and to be used by user. The web portal now can be hosted and the working of the system is to be monitored. Feedbacks from the users will be beneficial for improvement.

Feasibility Study

Prior developing the product or software it is necessary to know all the factors important for developing a good product. It is important to know the resource requirements, strengths and weakness of the project, etc. All these factors ultimately revolve around two major criteria cost and value, in order to know how much feasible, it is to create a project and how much profit will it return.

The feasibility of a particular software product can be determined from the following factors:-

1. Technical Feasibility.
2. Operational Feasibility.
3. Economic Feasibility.
4. Legal Feasibility.

Technical Feasibility:

For better availability and ease of use, it is important to create a product which is technically feasible i.e. which says that the product should be delivered with the available technologies, languages and skills. Technical feasibility focuses on the practicality of the project. While developing the system we have used python version 3.7 for the server-side function which is commonly available and can be installed easily.

Operational Feasibility:

Operational feasibility will check that how well the proposed system will solve a problem and how it satisfies the requirements identified in the requirements analysis phase of system development. With the help of this, the proposed system will be able to identify whether accurate results are delivered to the customer or not.

Economic Feasibility:

It is used to determine the benefit and saving that are expected from a proposed system and compare them with costs. It is an ongoing effort that will help in improvement in the accuracy at each phase of the system life cycle. This project is economically feasible as there is no use of any hardware device, thereby reducing cost.

Legal Feasibility:

Legal feasibility checks whether the proposed system conflicts with legal requirements or not. The project itself is working for a legal purpose, to detect the antisocial sites over internet and hence legal feasibility is assured.

III. RESULT

a) Output

Antisocial activity detection system will be able to detect any antisocial website on the internet. The user wants to know whether the website is antisocial or not. Therefore, he/she will enter the URL of the website which he thinks is illegitimate. The system will scrape the data present in the URL provided by the user which will then be preprocessed and classified. This classification will be provided as an output to the user which will tell that the website is legitimate or illegitimate.

b) Outcomes

The proposed system will strengthen protection against antisocial activities on the internet. With the help of this

system, the user has a platform to investigate about the website, whether it is legal or illegal. The system ensures that the alleged and proven fraud are dealt with, in timely manner.

c) Discussion of the results

Initially the literature survey was done where the research gaps were identified. Based on the research gaps the problem definition was derived which aims on identifying whether the website is antisocial or not. The system will take input in the form of URL from the user wherein the content of the URL will be scraped, preprocessed and classified. This classification will help the user clarify its doubt about the website that it is legal or illegal. The system is technically, economically and legally feasible.

IV. DISCUSSION

Internet today, is one of the most important part of our life. It is nothing less than a boon for person related to any field. Along with the benefits it also has certain negative points. Since anyone can use the internet, so the huge user base of this virtual world contains some naughty minds as well. A variety of dark net markets are currently available in the market. These markets allow for the anonymous illegal buying and selling of drugs and other illegal or control substances such as pharmaceuticals and weapons. Terrorist organizations use dark net for recruiting and sharing of information in order to plan terrorist attack.

There is a need to identify all such websites that are antisocial and cause harm to the society. So, in order to reduce the volume of harmful activities on the web, there is a need for an automatic detection. This project will help differentiating an antisocial website with a normal website with the automatic detection of illegitimate websites.

To detect the antisocial activities on the internet the system uses machine learning, data mining algorithms which classify the data provided to them into the classes which are defined.

The algorithm studies the data, learns it and based on it provides the output in which class the particular data lies. For the data analysis and processing the use of python and its libraries namely pandas, scikit-learn has been done. For getting data from the link, the system uses web scraper which is created with the help of beautifulsoup4 and requests modules.

The project ultimately gives a clear idea to the user that the link which is provided by him is antisocial or not. This will help us in reducing the probability of antisocial activity happening in future.

V. CONCLUSION

Antisocial activity detection system will be able to detect any antisocial website on the internet. The user wants to know whether the website is antisocial or not. Therefore, he/she will enter the URL of the website which he thinks is illegitimate. This system will scrape the data present in the URL provided by the user which will then be preprocessed and classified. This classification will be provided as an output to the user which will tell that the website is legitimate or illegitimate.

With the help of identification of antisocial activities system will help government security organizations to reduce online crimes to certain extent.

Just as terrorists seek to exploit technology, the government will harness developments in technology and use this system to protect the society. In future, the antisocial activity detection system will help in security of national assets leading to peace in the society.

REFERENCES

[1]. Hsinchun Chen “Exploring extremism and terrorism on the web: The dark web project” 2007

[2]. Amit Pratap Singh, Dr, R. C. Jain “A survey on different phases of web usage for Anomaly user Behavior Investigation” 2007

[3]. Floras Tsai, KapLuk Chan “Detecting cyber security threats in weblogs using probabilistic model” 2007

[4]. Yulei Zhang, Shuo Zeng, Li Fan, Yan Dang, Catherine A. Larson and Hsinchun Chen “Dark web forums portal: Searching and analyzing Jihadist forums” 2009

Implementation of Voice Controlled Virtual Assistant using STT, TTS and Logic Handling Engine

Nikunj Sartan, Varun Jasani, Akshit Singh, Mrs. Ashwini Patil

Computer Engineering, *TCET, Kandivali (East), Mumbai*

nikunjsartan123@gmail.com, jasanivarun014@gmail.com, akshitgiroud6@gmail.com, ashwini.patil@thakureducation.or

***Abstract*— The purpose of this paper is to illustrate the implementation of a Voice Command System as an Intelligent Personal Assistant (IPA) that can perform numerous tasks or services for an individual. These tasks or services are based on user input, location awareness, and the ability to access information from a variety of online sources (such as weather or traffic conditions, news, stock prices, user schedules, retail prices, telling time, local traffic, travel assistant, events, notification from social applications plus one can ask questions to the system, invoke its machine learning otherwise get it from Wikipedia… etc…..).**

Using voice as an input to convert into text using a speech to text engine. The text hence produced was used for query processing and fetching relevant information. When the information was fetched, it then be converted to speech using text to speech conversion and the relevant output to the user was given. Additionally, some extra modules were also implemented which worked on the concept of keyword matching.

It can help the visually impaired to connect with the world by giving them access to Wikipedia, any definiton, Twitter, and Music all through their voice Also it can be a source of entertainment and information for blind/visually impaired by playing their favourite music .

Keyword—STT, TTS, mic

I. INTRODUCTION

In today's world there are a lot of computerized tasks users have to perform manually. For example if user wants to search for data of a particular website, he uses multiple mouse clicks and keyboard button hits for searching that data he wants which consumes more time and energy.

Similarly there are many other computerized tasks which user has to perform manually in order to get the required outcomes. Therefore voice controlled virtual assistant came into existence to give required outcomes
without much effort.

Virtual assistants are useful for helping the users of a computer system automate tasks and accomplish tasks with minimum human interaction
with a machine. The interaction that takes place between a user and a virtual assistant seems natural; the user communicates using their voice,
and the software responds in the same way.

II. RELATED WORK

Voice recognition is being implemented nowadays in many smartphone devices and major other electronic devices too

Below given are some of the applications of speech recognition

Samsung uses bixby voice assistant which takes in voice commands to perform certain search operations on the phone

Similarly apple has its own voice assistant Siri and other devices have their own and so on

This feature is being implemented in IOT devices too like amazon's Alexa

With Alexa it is possible to completely automate your home nowadays

For instance, You just need to feed in voice command and you tunr the lights on or off in any room of your house

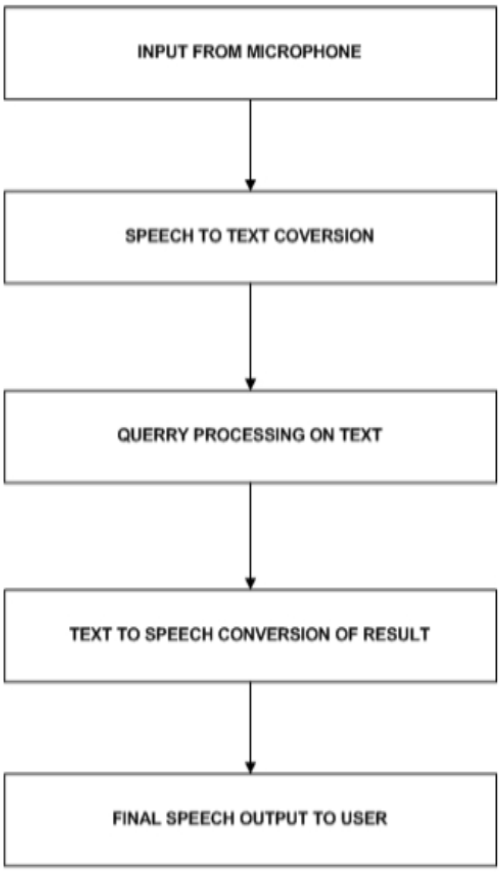
III. RESEARCH/TECHNICAL/PROJECT WORK

A. Flow of Events in Voice Command System

First, when the user starts the system, he uses a microphone to send in the input. Basically, what it does is that it takes sound input from the user and it is fed to the computer to process it further. Then, that sound input if fed to the speech to text converter, which converts audio input to text output which is recognizable by the computer and can also be processed by it.

Then that text is parsed and searched for keywords. Our voice command system is built around the system of keywords where it searches the text for key words to match. And once key words are matched then it gives the relevant output.

This output is in the form of text. This is then converted to speech output using a text to speech converter which involves using an optical character recognition system. OCR categorizes and identifies the text and then the text to speech engine converts it to the audio output.



B. *Algorithm:-*

Step 1:-Put the Speech

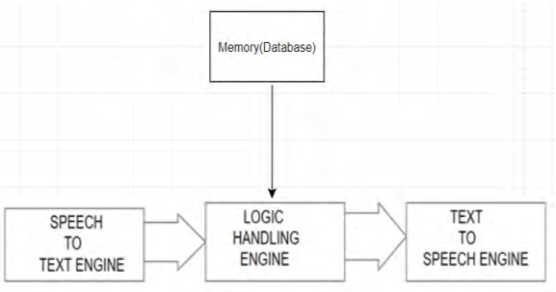
Step 2:-Convert the speech into series of texts as keywords

Step 3:-Search for the keywords in the database

Step 4:-If found in the database perform the required action

Step 5:-If not found then goto Step 1.

C. *Flowchart*



D. *Literature Survey*

The literature survey for research was done by referring to the research papers, articles, books and internet. Overall, this project describes a review of speech recognition task, speech recognition approaches, current speech recognition system, as well as different type of methods applied to speech recognition system. Based on the review of the advantages and disadvantages, this project discusses the most suitable techniques and methods to develop a speech recognition system

Some sources like books(The python playground book and automate boring stuff using python book),websites like pythonprogramminglanguage.com, Wikipedia.com, youtube.com and various other kinds of sources are being referred for literature survey of this project.

IV. RESULTS AND DISCUSSION

The advent of virtual assistants has been an important event in the history of computing. Virtual assistants are useful for helping the users of a computer system automate tasks and accomplish tasks with minimum human interaction with a machine. The interaction that takes place between a user
and a virtual assistant seems natural; the user communicates using their voice, and the software responds in the same way.
Such a virtual assistant can serve in the Internet of things as well as run a
voice-controlled coffee machine or a voice-controlled drone but for the sake
of this project we have limited it to its software usability.

V. CONCLUSION

In this paper we presented the working and the idea of our project. Our project is certainly aimed for physically disabled people who could use our software on the comfort of their chairs/desks with just their voice as an input. We sincerely hope our project will add benefits to some extent in the society.

REFERENCES

[1] Ass. Prof. Emad S. Othman, November-2017, Senior Member IEEE - Region 8, High Institute for Computers and Information Systems, AL-Shorouk Academy, Cairo – Egypt, PH- 0020-010-25830256. E-mail: emad91@hotmail.com, International Journal of Scientific & Engineering Research Volume 8, Issue 11

[2] Aditya K, Biswadeep G, Kedar S and S Sundar,2017, School of Electronics Engineering, VIT University, Vellore 632014, Tamil Nadu, India

[3] Lawrence Rabiner, Ronald W. Schafer 2013, Programs for supporting the teaching of digital speech processing IEEE Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE) pp. 290 – 295.

[4] Eric Matthes, Python Crash Course: A Hands-On, Project-Based Introduction to Programming Book,2015

[5] Al Sweigart ,Automate the Boring Stuff with Python: Practical Programming for Total Beginners,2015

[6] Recording audio using Google speech recognition referred from https://pythonspot.com/personal-assistant-jarvis-in-python

[7] Google speech recognition, API, Retrieved from

[8] https://pythonspot.com/en/speech-recognition-using-google-speech-api/

A web based course learning systems for needy students

Ravishankar S. Singh1, Pratik S. Singh2, Swaraj S. Kasliwal3, Mrs. Veena Kulkarni4
Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, India

Abstract- **The project is developed as a non – profitable web based system. The focus of this website is on equipping/educating the indigent and rural students, who find it difficult to pursue their studies in their own interested field. The project strives to educate the students by strengthening or developing their fundamentals with minimal cost. The proposed system includes various modes of learning. The students can have access to video lectures, conferences, pdf notes etc. The students can also solve their queries using virtual assistant. The system also consists of a notification and updating system. With regards to this, the one with appropriate access will be able to amend or modify certain, though limited, aspects which are not easily provided in other systems. From paper management to schedule adjustment and from that to portfolio assessment, the system provides an immense array of provision making the system more versatile and uniquely capable. The system is focused on removing the barrier caused due to the financial condition of the students and also the lack of modes of interaction. Being a web based application; it also gives the rural students some free courses. This is a very light weight system. The system would be very easy to use so any age group students can use it.**

I. INTRODUCTION

The project presented is an application based non-profitable system which is to be developed for certain category of students, especially indigent students and rural students. Since so many students manage education by engaging in rather external activities, system tries to provide a grandeur environment for student to grow in fields apart from college curriculum. The application system provides a platform where students who can learn about new technology in various field both theoretical and practical aspects.

II. MOTIVATION

There is a need for such application which will provide an easy access to different courses from beginner's level in a simple and interactive manner. Also the application will help students of other streams who want to learn computer courses and students can learn advance courses which are not provided in college.

III. PROBLEM DEFINITION

Due to the lack of money and bad financial condition so many students not able to get knowledge of computer or join institutes. If the above mentioned decides to join some institute the latter may encounter some experts that don't answer their basic questions. So this application is going to help those students to get education of computer from experts and they can start their education from very scratch.

IV. OVERVIEW

This project is an application based non-profitable system which is to be developed for students specially indigent students and rural students. Many Websites and applications are already available which provides education but not from scratch and on high cost. If students need any suggestions regarding to courses they can directly contact through the course instructor which can give them suggestions regarding to the course.

For using this application new student have to sign up and old student can register themselves for new course. This application system will recommend courses based on their fields of student interests. Voice based searching will be available for students by which students can search every information related to the courses. By using interactive GUI students will be able to check catalogue of every courses. The system provides a platform to conduct webinars and to publish papers in various publications.

V. CURRENT SYSTEM

The current system are all web based application which provides courses at high prices. Some application provides courses at low prices but the content is not good or it is so advance that a beginner cannot understand. Some application provides good content but they do not have interactive interface.

In current system the main problem is if anyone has any query the management system is not that interactive. There are platforms which provide courses but there platform give recommendation of courses it doesn't provide any suggestion for a naive user who doesn't know from where he / she should start or how he should approach a particular courses.

The other problem with this system is to have an active internet otherwise you can't work or access the courses, where as the main purpose of application is to provide courses to rural people so if they don't have internet than also they can access the course material.

In the previous systems if student have some queries he has to wait for replies of experts and if his queries are irrelevant to experts then students do not get replies from experts. If some students have language problem previous system cannot solve and it cannot provide any suggestions regarding the courses.

So the identified previous systems are lagging in solving these predicament of students. The system is focused on Indian students basically who are facing difficulties with English. The project goal is to provide computer education to indigent students so integration online and offline

platforms is considered.

VI. PROPOSED SYSTEM

In this project; after careful and relentless consideration; team have come to conclusion that system would be proceeding with SDLC, specifically pre-emitive SDLC or waterfall model. The reason to choose this specific model is due to its merits and current dynamics with the external institute the respective project is building for.

As per previous discussion with Onsite advisor of that institute, team was provided with required specification and overalls of the project and accordingly the task to develop the project was allocated.

This represents the structural design of system where the system is divided into two initial aspects; staff record only accessible by restricted personals and public data which can be accessed by student and other concerned peoples.

The structure suggest that system consist of four level; the first level is concerned with data accessibility in system and the second level is more concerned with the involved modules and the particular roles they play in next stage.

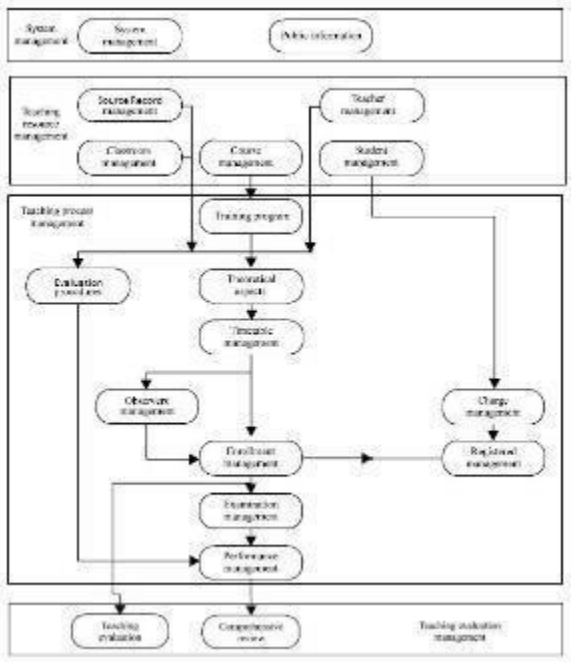


Figure 1: Block Diagram

The third level is the one which shows basic functionality of system and involving different modules as per their roles. And the final level is the evaluation stage based on functioning of involved parties in the previous level.

A. Flow of the System

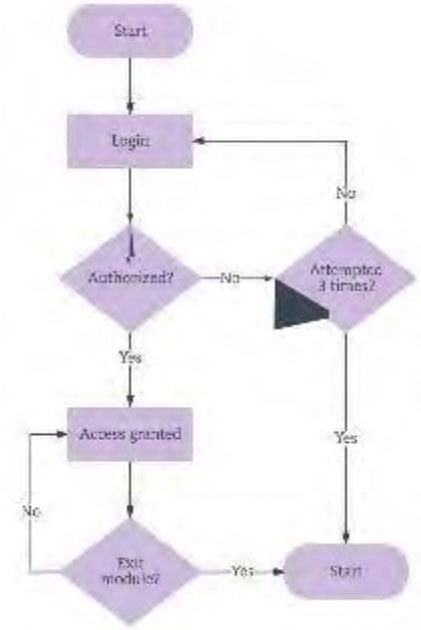


Figure 2: Flowchart of Login [8]

The initial stage is the login stage. Here user has to login with his unique id and password, if any of the above thing is wrong the system will ask again, and the system will give user three chances.

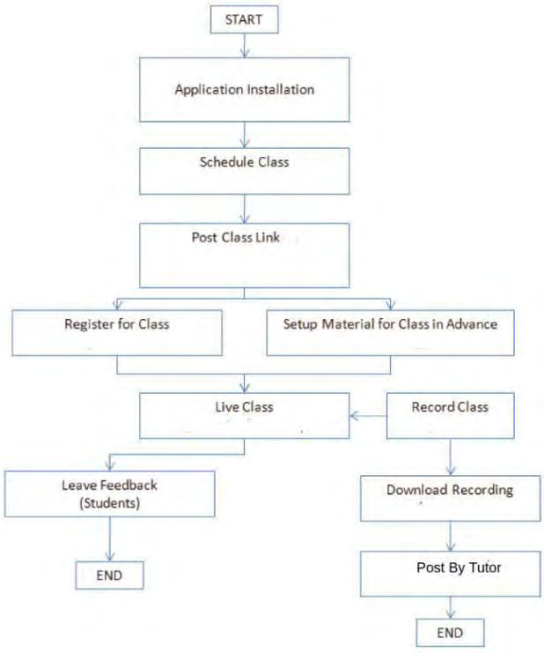


Figure 3: Flowchart of Course Upload [9]

This flowchart depicts how instructor will upload different assignment and course materials. Each course will be given unique id.

The algorithms starts with initial estimates for the K centroids, which can either be randomlygenerated or randomly selected from the data set. The algorithm then iterates between two steps:

In the data assignment step, each data point is assigned to its nearest centroid, based on the squared Euclidean distance. More formally, if ci is the collection of centroids in set C, then each data point x is assigned to a cluster based on

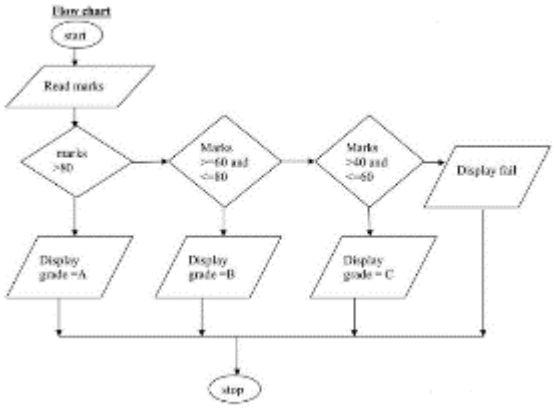


Figure 4: Flowchart of Grading [10]

$$\underset{c_i \in C}{\operatorname{argmin}}\ dist(c_i, x)^2$$

Where, "dist (.)" is the standard Euclidean distance. Let the set of data point assignments for each ith cluster centroid be Si.

Whereas, in the centroid update step, the centroids re recomputed. This is done by taking the mean of all data points assigned to that centroid's cluster.

$$c_i = \frac{1}{|S_i|} \sum_{x_i}$$

In this stage grading of student assignment and test will done. Based on how student has performed grades will be given.

VIII. CONCLUSION

areas. Areas like the administrating staff region where from the schedule calendar to student notification, lecturer's coordination and other aspects are handled. Then there is lecturer region where from test arrangement to reports of every student under them showing there performance analyses are handled by the respective lecturer.

Also, one takes a look at student region where every student is

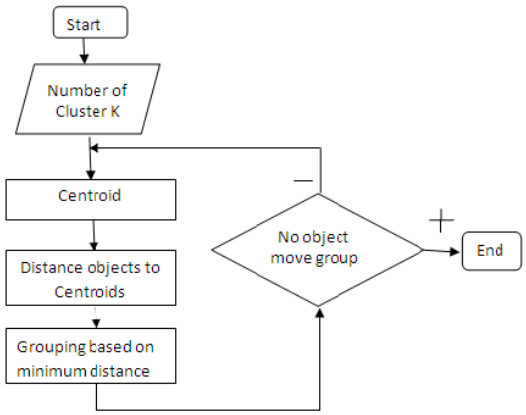


Figure 5: Flowchart of K-means Algorithm (Recommendation) [11]

The flowchart is of the algorithm we will use in our recommender system. The work of the recommender system is to recommend subjects to students based on their previous choices.

B. Recommender System

There are plans for implementation of clustering algorithm after completion of initial phase of project. The algorithm will be used for recommending courses based on student interest and answering their basic question. The specific algorithm which will be implemented for this will be k mean algorithm. It is an unsupervised learning algorithm.
The goal of algorithm is to find groups in the data with number of groups represented by K. The algorithm works iteratively to assign each data point to one of the group or cluster based on the features. Data points are clustered based on similarity.

The algorithm iterates between steps one and two until a stopping criteria is met (i.e., no data points change clusters, the sum of the distances is minimized, or some maximum number of iterations is reached).

This algorithm is guaranteed to converge to a result. The result may be a local optimum (i.e. not necessarily the best possible outcome), meaning that assessing more than one run of the algorithm with randomized starting centroids may give a better outcome. [12]

VII. CONCLUSION & FUTURE SCOPE

In this project, the team was able to develop an institute management system where the system handles certain specific

provided with very necessary aspect like the scheduled lecture updation and frequent exams. Also made available are the video lecture and other source material to read on. The system structure can be said as three tier system where administrating staff, students involved and lecturers all have separate rights to channel through and fulfill the necessary role.

Many students from a backward background are not able to

get the level of education to break the barrier. Even students of suburbs suffer from lack of in depth knowledge of the context. Here is where the institute will be able to make a difference. By providing a well-equipped system to a promising cause there

system plays a very crucial role to provide them with a stable backbone that can help them to get more organized with.

FUTURE SCOPE

In the starting phase the application will be deployed in some computer institutes under the NGO after success of first phase more number of computer institutes and individual experts can join by themselves.
Project will be deployed in all over in India under Pragati Computer Institute and The future scope of this application is to provide different courses other than computer courses.
In its future version it will be fully automated so that expert panel is not required anymore.
An android application is to be developed of this application.
Fully automated system is scheduled to be implemented in the next phase where input of admin will be lower.

REFERENCES

[1] “Small Business Data Storage Solutions” research paper by Synology A New NAS Experience, Synology America Corp. 2899 152nd Ave NE Redmond, WA USA 1-425-818-1587.
[2] D B Heras, D. Otero,and F. Arguello, ”An eco feedback system for improving the sustainability Performance of universities,” in Proc. 2011 IEEE International Conference on Virtual Environments Human –Computer Interfaces and Measurement Systems, Ottawa,ON 2011 , pp. 1 – 6
[3] S. Jeyalatha, B. Vijayakumar,and G.S. Wadhwa, “Design and implementation of web based application for relational data maintenance in an university environment,” in Proc. 2011 International Conference and Workshop on Current Trends in Information Technology , Dubai,pp. 105-112
[4] Rohan Bhirangi, Smita Bhoir “Automated Question Paper Generation System” International Journal of Emerging Research in Management & Technology ISSN: 2278-9359 (Volume-5, Issue-4),April 2016
[5] Veerle Van den Eynden, Louise Corti, Matthew Woollard, Libby Bishop and Laurence Horton “MANAGING AND SHARING DATA” UK Data Archive University of Essex Wivenhoe Park Colchester Essex CO4 3SQ ISBN: 1-904059-78-3, e 16 March 2011.
[6] Sawal Tandon Computer Science and Engineering, Lovely Professional University “Application of Incoming SMS to a Website, Control the Website to Send Bulk SMS” (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012,4287 – 4289, ISSN:0975-9646
[7] Hannah Bast and Claudius Korzen “The Icecite Research Paper Management System”, Department of Computer Science, University of Freiburg, Germany, X. Lin et al. (Eds.): WISE 2013, Part II, LNCS 8181, pp. 396–409, 2013. Springer-Verlag Berlin Heidelberg 2013.
[8]https://www.lucidchart.com/pages/examples/flowchart/system- flowchart-template
[9]https://www.researchgate.net/figure/260165534_Figure-1-Flowchart- for-conducting-WizIQ-virtual-tutorials-Note-ITS-Information-Technology
[10] https://sillycodes.com/c-program-to-calculate-grade-of-student/
[11] https://www.researchgate.net/figure/269800441_Fig-1-K-Means-Clustering-Flow-Chart
[12] Learn data Science, Machine Learning “Introduction to K-means Clustering” Andria Trevino 12.06.16 data sciencehttps://googleweblight.com/i?u=https://www.datascience.com/blog/k-means-clustering&grqid=qjW3m_1w&hl=en-IN

will be significant amount of impact on the society's culture and direct it towards a new direction of modernization. Under an institute which focuses of indigent people education and betterment, the

MyCityAct:Implementation of System to manage civic problem

Saurabh Singh, Adesh Borchate, Gaurav Shukla, Guide-Mrs. Harshali Patil

Department of Computer Engineering, Thakur College of Engineering & Technology, Mumbai, India

singhsaurabh2269@gmail.com, gauravshukla156@gmail.com, gauravshukla156@gmail.com, harshali.patil@gmail.com

Abstract **– MyCity Act: Implementation of System to Solve Civic Problem, proposed project is an automated system which will provide a better tech assistant for handling various civic problems in an efficient way. Our project will provide a common platform for citizens, authorities and the ground level worker who are responsible for implementation of commands. Mycity Act will serve as a single door solution for raising, following and solving concerned problems. In recent time smart phone technology and usage has grown exponentially and simultaneously. So keep in mind above fact we have aimed to implement our proposed project using android platform.**

The Idea of proposed project came from the situation faced by local people in day to day life. People usually came across situation where they find various type of civic related problems:-

Debris of garbage, Potholes, Damaged street lamps, Blocked sever line The citizens may have complaints with respect to their environment and city's infrastructure but they might not like the traditional complaining system in which they need to undergo a long tedious procedure like going to the office and standing there four hours in queue, wasting so much of their valuable time and efforts.

So, to gap the bridge, we came up with an online application introducing a new platform for sharing problems between Municipal authorities and the public, just in two clicks which can be easily used by the citizens in an optimal manner keeping them unaware of the background processes and details. Since it is a smart phone era where everyone possess smart phone. Among several existing platforms for mobile phones, Android is one of the largest platforms in the world that runs on several Smartphone's and tablets. Thus developing an android application to full fill this purpose will maintain a satisfactory relationship between citizens and governance and accelerate the process of civil development where all contribute to improve the condition and infrastructure of the city. India has third largest users of the smart phones. All these introduced services can be implemented with Android app for establishing digital community in the city. This app can promote digital literacy among the citizens.

I. INTRODUCTION

Since complaints are a valuable source of feedback to improve the infrastructure and conditions of our city. The citizens may have complaints with respect to their environment and city's infrastructure but they might not like the traditional complaining system in which they to undergo a long procedure like going to the office and standing there four hours in queue, wasting so much of their valuable time and efforts.

So, to gap the bridge, we came up with an online application introducing a new platform for sharing problems between civil service authorities and the public just in two clicks which can be easily used by the citizens in an optimal manner keeping them unaware of the background processes and details. Since it is a smart phone era where everyone possess smart phone. Among several existing platforms for mobile phones, Android is one of the largest platforms in the world that runs on several Smartphone's and tablets. Thus developing an android application to full fill this purpose will maintain a satisfactory relationship between citizens and governance and accelerate the process of civil development where all contribute to improve the condition and infrastructure of the city. India has third largest users of the smart phones. All these introduced services can be implemented with Android app for establishing digital community in the city. This app can promote digital literacy among the citizens.

II. RELATED WORK

In present time few of government and private application existed which are similar to few of module of proposed project:-

- Swachhgrahiis a cleanliness Application which is working towards ending the menace of open urination.
- "MeriSadak" is an android application by which any Indian Citizen can give his/her feedback on the pace of PMGSY road work, quality of PMGSY road work etc. (PMGSY stands for Pradhan Mantri Gram SadakYojana) to the related Departments in State Government or National Rural Roads Development Agency (NRRDA). This app also allows the user to take multiple photographs of the road and submit them along with the feedback.
- There is no present system which will do proposed work in same manner.

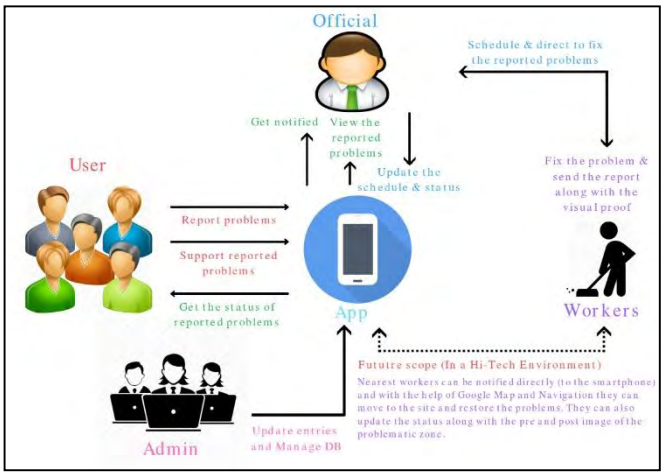
III. FEATURES

- Capturing and uploading the snap or video of the problematic zone as visual proof.
- Locating problematic zones by GPS and Google map
  Zip code: For GPS disabled services, map of that particular area will be displayed and user can easily locate the problematic zone. This enhances the flexibility of the system among the users.
- The integrated clock and calender will calculate the date and time automatically for every user activity.
- Under this section users can give possible suggestions to fix reported problemsas per their perspective.
- This could be helpful to prioritize the problems by giving likes and dislikes and/or commenting. It could also be helpful to ensure the relevancy.
- The user can track down the status of the reported problem such as "The problem is addressed", "Scheduled to resolve" etc.
- If the reported problem has not been addressed by the officials withing a specific span of time, then database itself generates the same problem and notifies the authority by placing that problem to the updated list. In this way the complaint can't be ignored or skipped without taking proper action.
- The server side algorithms automatically generate a predefined form consisting of all information entered by user and then send that report to concerned authority.

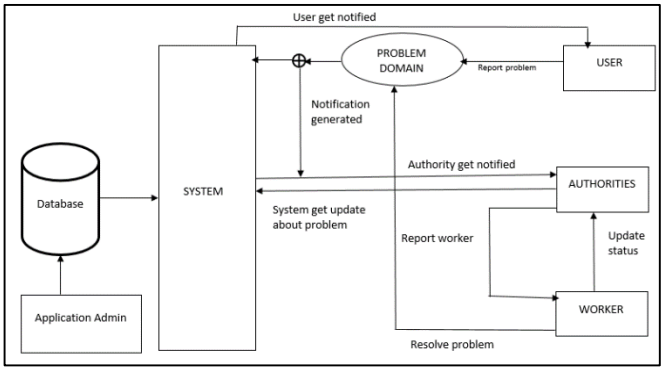
IV. DESIGN AND IMPLEMENTATION

1) User:

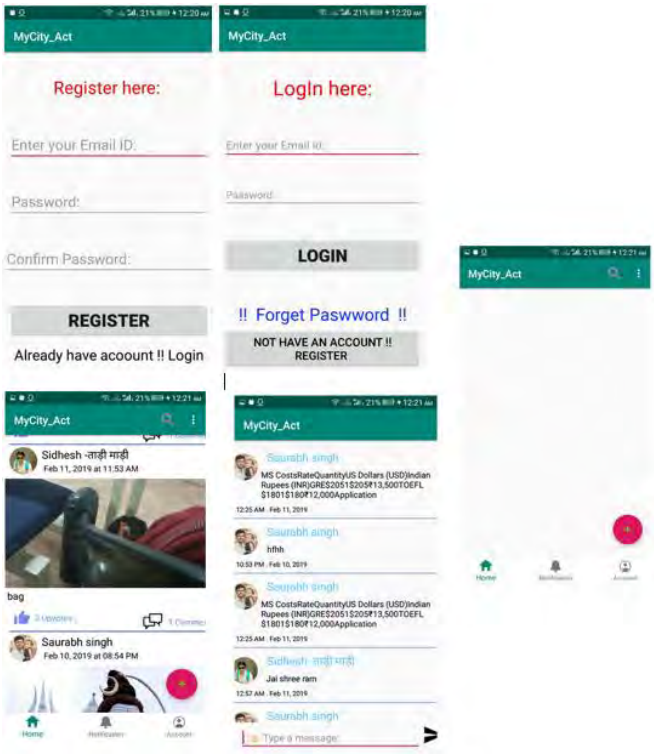
The user will primarily use the GUI for registering a new complaint by providing the necessary data. If user is not a registered user then he will have to register first and then sign in to the app and look for all registered complaints and their results. The user can also keep track of the registered complaints and its status.



2) Officials:



These are the actual people working for the app. When a complaint is registered, it will be notified to the officials of respective department. Accordingly they will inspect and



analyze the report and will take further action like addressing the problem, scheduling to fix it and directing the nearest worker etc.

3) App:

The application system has two major components, the first one is the server application and the second one is the mobile application. . The server application will run on the web server. The client application will run as a web application or an android application on a standalone PC or on an android based mobile phone respectively.

4) Admin:

The Task of admin is to deal with server side database. And maintain the complete record accordingly needed by user and authority.

V. CONCLUSION

Hence we conclude this paper by providing an immensely useful platform, encompasses new tools and technology, for the betterment of society with an ease of access towards problem reporting and quicker result than anticipation. With the onset of new technologies and innovations, human patience has shortened. Quick on-point result is expected with zero presence of human error. This App does quite the same thing providing several advantages over conventional manual complaint registration process with a greater efficiency, transparency and convenience. A Smart City can be taken according to six characteristics: Smart Economy, Smart People, Smart Governance, Smart Mobility, Smart Environment and Smart Living. The complete android app which will provide information services to the citizens. System will keep update to the citizens with information they need.

VI. REFERENCES

[1] Professor. Shinde R.S., Dept. Of Computer Science & Engineering, DACOE Karad, Maharashtra, India country.International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 03 Issue: 05 | “SmartCity(Karad)” Android Application . May-2016.

[2] Michael G. Wing, Aaron Eklund and Loren D. Kellogg.“Consumer-Grade Global Positioning System (GPS) Accuracy and Reliability,”Journal of Forestry, pp. 169-173, June 2005.Satish Kumar Prasad et al, / (IJCSIT)International Journal of Computer Science and Information Technologies, Vol. 7 (1) 402-406 ,2016.

[3] Easily create apps using the web technologies you know and love: HTML, CSS,and JavaScript; Internet: phonegap.com, [Apr. 25,2014].

[4] GoogleMapsJavaScriptAPIv3;Internet:developers.google.com/maps/web,[2014].

[5] Geolocation.Internet:docs.phonegap.com/en/3.3.0/cordova_geolocation_geolocation.md,html,[Apr. 25,2014].

[6] The Google Geocoding API Internet:developers.google.com/maps/documentation / geocoding/, [Apr. 25, 2014].

[7] Camera;Internet:docs.phonegap.com/en/3.3.0/cordova_camera_camera.md. html, Apr.25 ,2014.

[8] Vishesh K. Kandhari, Keertika D. Mohinani. “GPS based Complaint RedressalSystem”, IEEE Global Humanitarian TechnologyConference - South Asia Satellite (GHTC-SAS) September 26-27,2014

[9] Kim Nee Goh, Yin Ping Ng, KamaruzamanJusoff, Yoke Yie Chen and Yoon YehTan. “Architecture of a GPS-Based Road Management System”, World AppliedSciences Journal 12 (Special Issue on Computer Applications and KnowledgeManagement), pp.26-31, 2011.

[10] Umar Farooq, TanveerulHaq, Muhammad Amar, Muhammad Usman Asad andAsim Iqbal. “GPS-GSM Integration for EnhancingPublic Transportation Management Services&quot”, in Second International Conferenceon Computer Engineering and Applications, 2010

Building Machine Learning Model for Sentiment Analysis on Twitter information

Ms. Vani Singala, Ms. Nidhi Mishra, Ms. Aneri Moradia, Ms. Sonu Mehta, Mrs. MegharaniPatil

Department of Computer Engineering, Thakur College of Engineering & Technology, Mumbai, India

Abstract- **Continuous augmentation on a wide range region arrange association lays to gigantic expanded up and coming for union and asset sharing. Today, shifted socially associating locales like Twitter, Facebook, Myspace, and YouTube are to a great extent being used and it is for all intents and purposes resistant to expose them. These elements of social association and connection have a huge number of individuals sharing their contemplations every day in view of it's trademark short and effortlessly versatile and controllable way of articulation. It has turned out to be imperative to break down these casual articulations of discourse and recognize the intrinsic importance of a tweet and accordingly bifurcate the assessments of a lion's share agreement and use this unintentionally extremely valuable pool of absolutely described information to distinguish a specific target gathering of people .They are the majors of fundamental utilizations of Web 2.0. With the take off of person to person communication period, surge of client produced content has occurred. We hold up and review a worldview to mine the suppositions from a notable on-going small scale blogging application, Twitter. The administration on which the clients share their response on everything. Twitter contains a humungous number of content posts and it is thriving each day. The subsequent corpus can be incidentally substantial. Twitter's gathering of people contains a mixed blend of people, from ordinary clients to superstars, business delegates, government officials, and even presidents. Looking back assessment investigation of tweets isn't impeccable as formal dialect examination, as the human dialect is compound. Showing a machine to dissect the different linguistic varieties, slang and spelling botches that happen in online notices is a troublesome procedure. We offer a crossover technique comprising of corpus based and word reference based strategies to manage the semantic introduction of the sentiments in different tweets. A contextual analysis is exhibited to delineate the utilization and adequacy of the proposed framework.**

Keywords—Sentiment analysis; opinion mining; Twitter data analysis

I. INTRODUCTION

Sentiment Analysis (SA) along with its summarization has become the attractive area in the subject of data science, this is due to generalizing of online views are lucrative and desired in different applications. One such application which has been already developed is the pragmatic analysis of different user environment and preferences specified on Facebook, which helps other companies to deplete its target users and which specifies the probability success for an advertising company's product.

Microblogging sites have dozens of multiple thoughts shared every day of which it provides a platform for a simple way of declaration. The growing rate important to analyze these informal expressions of spoken words and identify the inherent meaning of a tweet and thus bifurcate the opinions of a majority consensus and take necessary actions based on it. In the proposed project, the pathway to a novel based solution to target-based sentiment conclusions and SA of informal texts subjecting the concentration of “tweets”, the twitter posts. To serve this, we have studied algorithms and methods for SA polarity detection and sentiment generalization.

This project is implemented using hybrid polarity detection. Also, learning method essentially uses both corpus based and word search-based book methodology to examine the semantic orientation of the described words tweeted [2]. Hybrid aspect - based learning that we have worked on in this project is inherently better than unsupervised or supervised learning based aspect detection system, as a supervised algorithm suffers from a large computer run time complexity due to its inability to automatically train itself on a large set of data, and unsupervised learning algorithms requires lot of efforts to acquire and maintain a lexicon, and our system has tried to reduce inconsistencies and ambiguous outputs, For instance, for the following tweet - "I absolutely detest Samsung galaxy note but like it's camera ", here a non-aspect -based system will straight away categorize this tweet as extreme negative as the presence of 'absolutely', but our aspect-based system will identify this tweet as negative for the product as a whole , but also categorize it's aspect(here camera) separately as positive, this gives the manufactures focused results to dwell on.

II. MARKET SURVEY

Opinion Extraction and Semantics Knowledge Discovery of Twitter Datais an aspect-based sentiment analysis project that aims to identify different important aspects of a company’s product and analysis the positive or negative tweets on that subject to identify the performance of the project in the market as well as separate the positive aspects of a system. Till today all the sentiment analysis projects under review have only been successfully developed using either supervised or unsupervised polarity detection methods, which has many drawbacks such as

The supervised polarity detection method uses Machine Learning (ML) approaches in which a classifier is trained based on a

feature set, using labeled training data, drawbacks of this approach is that for opinion polarity detection they are both domain (dataset topic) and temporally (datasets collected in different time periods on an annual basis) dependent.

The Unsupervised polarity detection methods are mostly based on a Sentiment Lexicon (SL), in which each sentiment-bearing word is associated with either a sentiment score or a set of sentiment bearing seed words as explained bellow, Lexicons are cumbersome and time consuming to maintain, furthermore they can be domain dependent.

The Hybrid approach uses both the above-mentioned polarity detection methods and thus removes the drawback, since our project is the first one to be developed and made commercially available using hybrid approaches it has an edge above other SA projects due to its higher accuracy and cheaper cost of development.

III. LITERATURE REVIEW

We have studied 4-5 papers similar to our project and have researched their methodology which can be utilized in our project. Also the development which can be included for enhancement of project [5]. It was develop using the approach called machine learning with SVM, Maxent and NB classifiers.Also, Sentiment lexicon which is directly created using both side of seedword. The further scope was sarcasm detection, along with, searching more features that could semantically conclude a document would be future works of this study [6]. It used build model's method using Support Vector Machines (SVM) and Naive Bayes, Maxent.

More Scope was stated as more research needs to be practiced to validate or invalidate these findings, using larger samples [7]. Method used was 6 state of art sets derived from semantic and synthetic analysis. And the Future Scope was increasing the scope by maximizing the tweeted datasets [8]. Sentiment analysis of twitter data done with Machine interaction technique and semantic analysis. It was developed using Maximum entropy, SVM with semantic orientation. The future scope of the project was better visualization of the content in a better manner useful for the users [9]. It was built using senticircle model mines the contextual semantics. Also Naïve Bayes and maximum entropy. It could be further developed by investigating the sentiment classification issue and its reflection on performance.

IV. BACKGROUND

PD of subjective opinions has been special focus of study in the project of literature. It is familiarly described as classifying subjective opinions into positive tweets opinions and negative tweets. Current PD methodology can be segregated as three main categories: supervised, unsupervised, and hybrid.

Supervised, Unsupervised, and Hybrid Polarity Detection

Supervised methods are Machine Learning method where a classifier is implemented on a feature set, which uses a labeled data. Pang et.al. were the initial ones to run SA on online movie reviews. They tested ML approaches, namely, SVM, MaxEnt and NB classifiers. Their research depicted how SVM is trained by molding of a unigram bag of texts, featuring set, standing out of all other approaches carried out in their research. The fault of formal ML methods for opinion polarity detection consists of domain (dataset topic) and temporarily (datasets collected in different time periods on an annual basis) dependent. Moreover, for classifiers to be engaged with advance large size feature sets (e.g. unigrams) long training are essential and due to the heavy size memory along with time complexity are required

Unsupervised Methods are majorly dependent on a Sentiment Lexicon (SL), where every sentiment-bearing word is correlated with a sentiment score or a sentiment set bearing seed words as depicted below. These technology approaches use varied algorithms to calculate a sentiment score for an opted portion of tweets. By using the sets of positive and negative seed-words in the literature study basically known to world as Semantic Orientation from Association, SL can be architect; or can be physically developed. Large number of manually built SLs. The SentiStrength lexicon specifically modeled for informal kind of texts (i.e. its words' stemming rules were adopted from the social networking website MySpace), is one such example. Faults of lexicon-based methods are investment required for capturing a lexicon and also maintaining them because of language changes resulted gradually. Additionally, lexicons could have domain dependency and the domain-independent ones may display less accuracy.

Hybrid methods adopt a combination of both of the above-mentioned categories to interpret opinion mining of tweets data. In majority of the system, a SL is subjected to launch features to practice an ML classifier which is referred to as sentiment features. For instance, adoption of an SL data to counter good and bad sides of tweets and utilized them to measures with different characteristic for training the classifier [1]. Our hybrid approach is one such example for such hybrid methods, which will significantly reduce drawbacks of the multiple methods. It resolves the faults of supervised methods, namely, the problem of large size feature sets and time and memory complexity. Contradictory, an essence for domain dependent lexicons will still stay. However, it displays to create many domain dependent lexicons and utilizing topic detection algorithms to investigate various domains and using the associated lexicon for SA of each domain may resolve the occurred problem.

V. SENTIMENT SUMMERIZATION

An aspect-based summarization system is a multi-documentsummarization system which summarizes documents based on different aspects of a target item. For example, an aspect-based sentiment summarizer summarises reviews about hotels' and restaurants' reviews, based on different aspects, namely, service, décor, value, etc. Analogously, an aspect-based sentiment summarizer, summarizing reviews about a Smartphone presents summaries regarding the Smartphone's battery life, camera, display, etc.

The approach presented by us for sentiment summarization consists of the following four main steps:

- Searching features of the intended item (e.g. smart phone's battery life) which can be studied in corpus of tweets.
- Clusters are formed by grouping the documents having same item features.

- By using polarity detection system we organize those documents.
- Based on the organized documents we output the textual and graphical summary

Each of the above mentioned steps will be explained in the following subsection:

A. The Aspect Detector Module

Using the multi-documents our system can summarize documents based on the most important features of the item. If there is no multi-document we can provide features of the item manually using plug-ins containing domain vocabulary. Conversely, they could be detected automatically, however, with lower precision.

B. Documents Clustering Module

In this step we cluster all the documents that have the aesthetics of the same aspect word in them. The clustering module is responsible for performing this task.

C. A Hybrid Polarity Detection System

Here in this section we will explain our new target-oriented hybrid system. This system is similar to that presented in [3]. However, here we propose some improvements to it that increases the classification accuracy over the previous accuracy baseline. Hybride Polarity Detection Systems architecture contains mainly three major modules namely: a preprocessing module, a lexicon-based module for sentiment generation and lastly a Machine Learning (ML) Classifier module.

VI. THE PREPROCESSING MODULE

The module executes following preprocessing steps explained below.

- "ATUSER" is added in place of @username.
- URLs are removed.
- "#word" is substituted with "word".
- The actual meaning of abbreviations or slangs is added instead of them. For it we can use a manually built slang dictionary
- Text snippets are formed from each document or tweets based up on '.' ',' '!' '?'. So, we can say a text snippet are words that appear in between two of the mention punctuations.
- The system can be provided with the query word using a target word that is found by using the sentiment or target word which is replaced by "TARGET".

VII. TECHNICAL FEASIBILITY

Our project is Opinion Extraction and Semantics Knowledge Discovery of Twitter Data, so to make use of this project the user needs to have a twitter account.

Method of Production:

We will follow the Spiral I.e. Iterative Model for software development.

We will be making use of Python NLTK (Natural Language Toolkit) or RapidMiner IDE to develop our project.

Production Technique:

- A sentiment analysis system is developed using sentiment summarization approaches, there are two main approaches namely, extractive-based summarization, which deals with extracting some supposedly important parts of the exact texts in a corpus of documents and presenting them as a summary of that corpus. The second method, known as abstractive summarization, deals with generating a textual summary in which the used words are not necessarily the ones used in the corpus. In this project we will be making use of extractive- based summarization and the system will consist of an aspect detector, an improved novel hybrid polarity detection system, as well as, a novel unsupervised polarity detection and ranking algorithm. In the process of summarizing multi-documents, the following steps are essential:
- Identifying products features (aspects) that public opinions are focused on. Detection of polarity of the sentences that have the occurrence of the important features identified in step one.
- Producing a numerical (in the form of graphics), as well as, a textual summary of those review sentences.

Project Requirements:

- Minimum RAM:1 GB
- 200 GB HDD.
- Intel 1.66 GHz Processor Pentium 4
- Windows XP, Windows 7,8
- Visual Studio 2010
- Access to Twitter API, Twitter account, access to internet at the time of accessing multi-document data from Twitter.

VIII. THE BLOCK DIAGRAM

The structure of the twitter analysis system is as follows:

- User: They are registered user of twitter who shares different views. The statement passed in these views needs to be analyzed. After User shares his opinion, the sentence goes to three modules for analyzing process.
- Retrieval Module: Here the sentences break into words for the analysis process. It acts as the first module.
- Preprocessing Module: This is the main module in analyzing process. It main function is to filter the data present in the Retrieval module. Firstly, it removes URL

such as @tags, #tags, etc. After which it removes errors. Then the filtered data enters to emoticon tagger from which the words related to opinions are extracted and forwarded to POS tagger. At the end of this filtering process a transaction file is produced where data are divided in the form of emoticons, verbs, adverbs and adjectives.

- Scoring Module: It the third module of the process where the results of the analyzing process are produced. It is basically bifurcated into two subparts-
- Word Seed List: Here the scores of verbs and adverbs are produced.
- Log Linear Regression Classifier: Here the scores of adjectives are produced

Twitter Sentiment Scoring Module: It is the final stage of the process and the scores for a opinions are generated

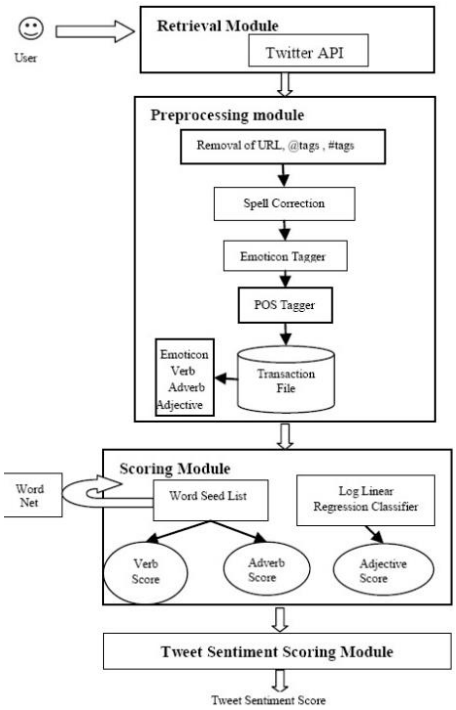


Figure 1. Block Diagram

IX. SENTIMENT FEATURE GENERATOR MODULE

This module proceeds by replacing the equivalence of slangs from the slang dictionary. We collect the slangs and their abbreviations through as many online sources as we could to create the slang dictionary. After this, the module tags all the sentiment-bearing words using the SentiStrength lexicon with the sentiment score of the following. In the same way, following a emoticon list, we tag the happy emoticon with a sentiment score of "+1" and the sad one with "-1". Additionally, it tags all the intensifiers (e.g. absolutely) and diminishes (e.g. might) with their corresponding scores. It also tags "NAGATE" to the negation words. Lastly it would add a score of "0" if the word does not along to any of the mentioned categories.

We can handle the occurrence of diminishes, negations and intensifiers once we are done with tagging all the words in the document. Using the score of the intensifier word we can change the strength of the sentiment-bearing words that are present in an intensifier. Oppositely, we lessen the strength of the sentiment-bearing word that is present after a diminishes, in case of diminishes, using the strength of diminishes. Finally, during appearance of negation which is to be handled, the polarity of the score of sentiment-bearing is flipped. After this we would like to weaken the above flipped score by 1, that is, if the score appears to be positive it will be subtracted by 1, conversely, if score is negative we add 1. During this process "0" tags are ignored which are found between sentiment-bearing words and shifters of a single text snippet while performing the computations mentioned above. Table 1 presents this feature set.

Table 1. Features used in our system

f1	Unsupervised polarity detection algorithm used for accounting the complete sentiments of twitter data.
f2	Positive word count
f3	Streak of negative feedback
f4	Number of blocks of bad words
f5	Count of cons followed by pros of product
f6	Amount of negation with a negation tweet
f7	Antithetical sentiment of words
f8	Positive word number guided by goal
f9	Negative word reference with pointer to aim
f10	Number of negation count followed by target
f11	Favorable trail by unfavorable count.
f12	Number of denial toured by acceptance
f13	Aim count treading heels with affirmative
f14	Goals number followed by disagreement
f15	Number of negation connected to beneficial finally to target.
f16	Count of negative heading with negative till the goal.

To get cumulative sentiment score we look up for feature f1. For further computation of this feature, the word's score and tagging process are aggregated which was explained in the engendering of the current subsection. Threshold of '0' for classifying words is defined when sentiment score of an entity or word is below '0' that word is tagged as negative however in other cases, if the score is above '0' it is denoted as positive. Then we calculate a comprehensive sentiment score for each document by calculating their aggregate.

The remaining features mentioned in the table are mostly Self-explanatory. The feature f7 called "inverse sentiment" makes use of some heuristics to find patterns that indicates a cluster of most of the words in a document stating the opposite polarity of

the actual cumulativel document polarity. These heuristics intend to find occurrences of words that often convey happening a situation or condition. For example, the words “if”, “after”, “before” and “until” belong to this category of words .If any of these words are detected in a corpus of a document previously analyzed , the algorithm used by f7, further analyzes that document on the basis of some predefined, subject specific heuristics to check if the sentiment words used in that document are indeed meant to convey their polar conflicting results. The heuristic regarding “if statements “are explained further in detail. Let’s take one example:

“If I don’t get an iPhone for Christmas, I would be sad.”

From the preceding example, we can observe that the actual polarity of the sentence regarding “iPhone” is positive, whereas, the sentence only contains words and patterns which indicate negative context - “don’t get an iPhone” is a negative statement regarding iPhone and its sets f10 to “1”.Similarly, the word “sad” is a negative word and its occurrence sets f3 and negation words are tagged with “Negation” f14to“1”.These features are often presenting negative sentences. Additionally, the presence of subsisting patterns, results in the inverse sentiment feature (f7) to be set to “1”. The heuristic explores the corpus document, to find out if both the “if clause” and the “main clause” in the example,include a negative sentiment pattern and if so it identifies this pattern as an inverse sentiment. Thus, our system is made so efficient that it easily detects and correctly classifies most such case.

Machine Learning Classifier

This module is a linear SVM takes input as the feature set which is described and explained in Table 1 and accordingly classifies tweets, which are documents, to separate class.

A. A Textual and Graphical Summary Generator Module

Here we will generate a textual summary as well as a graphical summary of the results obtained.

Non-Textual summarization

We can easily determine negative and positive points of document regarding each attribute of target item, which happens right after the documents are classified. Therefore, in such scenario, we have already come up with a numerical summary of the entire corpus regarding most important attributes of our target item.

X. TESTING

In this application there are certain requirements which need to be dealt with. Requirements like its users/customers account, credentials etc. For that application has to have a “Login” page that has these elements:

- Account/Username
- Password
- Login/Sign in Button

Unit testing of above mentioned case can be as follows:

- Field length – username and password fields
  - In here the module is concerned with the field length of certain attributes like password which should be according to mentioned description (special characters).
- Input field
  - Value of input field should be valid like in case of user name should be appropriate.
- Login button will forward user to the home page only if the credentials match. If not, then a pop up message of “invalid credentials” will appear.

Integration testing, of particular case can be as follows:

- Form Login
  - User sees the welcome message after entering valid values and pushing the login button.
  - User should be navigated to the welcome page or home page after valid entry and clicking Login button.

In this particular case considered for functional testing, description can be as follows:

1. The expected behavior is checked, i.e. the user will be able to login on clicking the login button after entering valid username and password values
2. After successful login user is navigated to home page.
3. There should be an error message that should appear on invalid login.
4. To check whether there is any stored site cookies for login fields.
5. Confirmation of registration is send to respective student.

XI. APPLICATION SCOPE

Search Engine Optimum searching and analysis of data from Twitter UI for efficient related results. Advertising Identifying target group of people for a particular product based on their twitter activity and tweets and retweets. Product Development Discover the performance of different aspects of a product based on user reviews and identify the aspects that need to be improved. Complaint Processing Getting ready to take appropriate action to satisfy the customers based on any probable negative reviews identified through the analysis of tweets. Sales Department Identifying the region-specific sales of a product and it's correlation with those region's product specific tweets.

XII. RESULTS

The outcome mainly depends on datasets which is created manually consisting 940 tweets. This tweet is collected by a group of 22 humans which has both negative polarity and positive polarity. We collect datasets by picking a particular word which is seen repeatedly. For example, let us take word “Samsung”. Aspects of this word are observed by 3 persons voting process. Out of collection of all tweets the word “Samsung” is checked, to which each tweet is read and labeled positive and negative by the selected three people. In this two out of three people will agree on same label. Positive and negative points are given to each of the aspects of the Samsung. For example, Samsung’s camera, battery life, side buttons and remaining components of Samsung. The points are given by looking at positive words such as nice, good, best and negative words such as don’t, unlikely, hate. We overcome challenges for dataset of a target word by performing sentiment analysis to compare two different Smartphone’s and informality of text

statements. Evaluation of words is done based on a formula which results single class accuracy and overall accuracy by calculating number of true positive and negative words and false positive and negative words.

Table 2. Review Scores for Product Searched

Product Searched	Summarized Reviews
Micromax	Positive: 70%, Negative: 30%
Vicco	Positive: 20%, Negative: 80%
OnePlus	Positive: 100%, Negative: 0%
IPhone	Positive: 80%, Negative: 20%

XIII. POLARITY DETECTION ALGORITHMS

The observations of our experiment are given in the following subpart. Reference performed an experiment; it trained and tested the SVM, Naïve Bayes (NB) and even Maximum Entropy (MaxEnt) classifiers based on the same dataset explained in the start of the current section with presence or absence of stop words and emoticons in the unigram feature set [4]. By performing the experiment, we concluded that the SVM trained algorithm with unigram feature set with stop words and emoticons added, performed better than all others with an overall accuracy of 86.69%. Lastly, I would like to say that we will be using SVM as our baseline for the experiments.

XIV. CONCLUSION

The rise of micro-blogging has changed the means of communication nowadays. The humongous repository of information made available by twitter has deemed this website as an attractive source to get data from various different entities, to perform public opinion mining.

In this project, we have built a sentiment summarization system by combining methods and algorithms, which allowed us to summarize opinionated texts in the domain of consumer-products. Additionally, we have illustrated that our unsupervised PD and ranking algorithm could be utilized to generate extractive summaries of a corpus of multi-documents for an efficient aspect-based summary of data. Our experiments have lead us to believe that in the context of SA, moving towards sentiment features that require cognitive thinking rather than conventional text processing features would be a promising solution to this problem. Detection of sarcasm, as well as, finding more features that could semantically interpret a document are also implemented with the most accuracy provided till date. Additionally, building a multi-domain context dependent lexicon would be a future direction of this project.

XV. FUTURE SCOPE

Our project has already covered most aspects of opinion extraction as well as sentiment discovery. We have also tried to implement the sarcasm feature in the application. So, the future direction of our project is to build a multi-domain context dependent lexicon that will enhance the sentiment scores of the tweets obtained also increase the number of inputs taken into consideration for analysis. Further we would like to move the project to various other applications other than twitter.

REFERENCES

[1] Hogenboom, A.; van Iterson, P.; Heerschop, B.; Frasincar, F.; Kaymak, U., "Determining negation scope and strength in sentiment analysis," Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on , vol., no., pp.2589,2594, 9-12 Oct. 2011.

[2] Lai, P., “Extracting Strong Sentiment Trends from Twitter”, Jan 2011.

[3] O'connor, B.; Balasubramanyan, R.; Routledge, B.; Smith, N. From “Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series.” International AAAI Conference on Weblogs and Social Media, North America, May 2010.

[4] Pak, A., and Paroubek, P., "Twitter as a corpus for sentiment analysis and opinion mining" in Proceedings of LREC, May 2010.

[5] “Sentiment analysis and summarization of twitter data” [online].Available: http://ieeexplore.ieee.org/document/6755222/

[6] “Sentiment Analysis on twitter” [online]. Available: https://www.ijcsi.org/papers/IJCSI-9-4-3-372-378

[7] “Semantic pattern for Sentiment analysis of twitter” [online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-11915-1_21

[8] “Sentiment analysis of twitter data using Machine learning approaches and semantic analysis” [online]. Available: http://ieeexplore.ieee.org/document/6897213

[9] “Semantics pattern for sentiment analysis on twitter” [online]. Available: https://www.researchgate.net//268133308_Semantic_Patterns_for_Sentiment_Analysis

Designing a Intelligent Chatbot for E-commerce

Mr. Prakhar Agrawal, Mr. Dhruvil Parekh, Mr. Rohan Gupta, Mrs.Megharani Patil

Department of Computer Engineering, Thakur College of Engineering And Technology Mumbai, India

Abstract: These days, organizations are gradually beginning to convey chatbots as another technique for correspondence with its clients. Because of the subject's earliest stages and absence of research regarding the matter, the reason for this examination is to investigate the idea of portable ambassador chatbots and an endeavor is made to decide the Dutch Millennial' goal to utilize chatbots as the following interface for versatile business. An exploration show is proposed dependent on the Technology Acceptance Model (TAM) and Innovation Diffusion Theory (IDT). Information is gathered by methods for an online review among 195 members. The proposed research show is tried by methods for straightforward relapse investigation and results are cross-approved utilizing IBM Watson Analytics. All proposed speculations are upheld. In any case, there is no unambiguous response to whether Dutch Millennial have the aim to utilize chatbots as the following interface for business. Regardless, the greater part of the respondents express a positive initial introduction towards chatbots. This examination realizes a few impediments with respect to outer legitimacy and the exploration show is restricted to five autonomous builds. Extra builds or estimation instruments could be utilized to get a more profound comprehension in regards to the subject. In addition, utilizing a genuine trial may produce unmistakable outcomes. Associations needing to send flag-bearer chatbots, advertisers and chatbot engineers ought to think about similarity, the customers' way of life and shopping inclinations, for a fruitful usage. So also, the purchasers' security concerns and protection from nosy versatile commercial are imperative points to be considered.

I. INTRODUCTION

Chatbots are software agents that interact with the user in a conversation. The main goal of their creation was to resemble a human being in the way they perform said interaction, trying to make the user think he/she is writing to another human being. This has been implemented with varying degrees of success. One of the most popular languages for the definition of a chatbot knowledge base is artificial Intelligence Mark-up Language(AIML). This thesis focuses on the implementation of an AIML interpreter written in JavaScript to allow for a web-based client-side specific usage of AIML chatbots .The interpreter must guarantee the compliance of properly formed AIML documents, perform all the necessary pre-processing duties for the correct usage of the chatbot and ensure the correctness of both pattern matching of user input and chatbot response. The interpreter fully exploits the Document Object Model(DOM) tree manipulation functions of the jQuery library to achieve said goals ,treating AIML files as if they were normal XML files The result is a well performing ,fully functional AIML interpreter tailored around AIML 1.0 specification.

A chatbot is software that is used to interact between a computer and a human in natural language. Naturally, it can extend daily life, such as help desk tools, automatic telephone answering systems, to aid in education, business and e-commerce. Have you ever had a problem with a service or a product and ended up contacting the company? Many of us have and chances are that the experience wasn't pleasant. The process can be very frustrating and in most cases, those customer support agents aren't trained very well and have limited knowledge and resources to actually help solve your issue .It's time to move on from those pesky agents! It's 2017 and technology is advancing.

II. LITERATURESURVEY

Ecommerce web companies spend a lot of their investment maintaining their call centres. These call centres mainly handle queries of the users, regarding their products. This traditional method costs both time and money to the companies. One of the first and main goals of chatbots has always been to resemble an intelligent human person and make it hard or impossible for the other party of the conversation to understand their real nature (as in artificial). A Chat-bot is a conversational agent that interacts with users using natural language.

There are many applications that are incorporating a human appearance and intending to simulate human dialog, but in most of the cases the knowledge of the conversational bot is stored in a database created by a human experts.

However, very few researches have investigated the idea of creating a chat-bot with an artificial character and personality starting from web pages or plain text about a certain person. This paper describes an approach to the idea of identifying the most important facts in texts describing the life (including the personality) of an historical figure for building a conversational agent that could be used in middle-school CSCL scenarios[1]

This paper describes a flexible method of teaching introductory artificial intelligence (AI) using a novel, Java- implemented, simple agent framework developed specifically for the purposes of this course. Although numerous agent frameworks have been proposed in the vast body of literature, none of these available frameworks proved to be simple enough to be used by first-year students of computer science. Hence, the authors set out to create a novel framework that would be suitable for the aims of the course, for the level of computing skills of the intended group of students, and for the size of this group of students. a synthesis of the traditional objectivist approach and a real-world-oriented, constructivist approach to teaching programming to novices. The main aim of implementing such a pedagogy was to engage the students in learning to which they personally relate while attaining intellectual rigor. Classroom experience indicates that students learn more effectively when the traditional objectivist approach is combined with a constructivist approach than when this orthodox approach to teaching programming to novices is used alone [2]

A chatbot aims to make a conversation between both human and machine. The machine has been embedded knowledge to identify the sentences and making a decision itself as response to answer a question. The response principle is matching the input sentence from user. From input sentence, it will be scored to get the similarity of sentences, the higher score obtained the more similar of reference sentences. The sentence similarity calculation in this paper using bigram which divides input sentence as two letters of input sentence. The knowledge of chatbot are stored in the database. The chatbot consists of core and interface that is accessing that core in relational database management systems (RDBMS). The database has been employed as knowledge storage and interpreter has be unemployed as stored programs of function and procedure sets for pattern-matching requirement. The interface is standalone which has been built using programming language of Pascal and Java. [3]

Chatbots are software agents that interact with the user in a conversation. The main goal of their creation was to resemble a human being in the way they perform said interaction, trying to make the user think he/she is writing to another human being. This has been implemented with varying degrees of success. One of the most popular languages for the definition of a chatbot knowledge base is artificial Intelligence Mark-up Language(AIML).This thesis focuses on the implementation of an AIML interpreter written in JavaScript to allow for a web-based client-side specific usage of AIML chatbots. The interpreter must guarantee the compliance of properly formed AIML documents, perform all the necessary pre-processing duties for the correct usage of the chatbot and ensure the correctness of both pattern matching of user input and chatbot response. The interpreter fully exploits the Document Object Model(DOM) tree manipulation functions of the jQuery library to achieve said goals, treating AIML files as if they were normal XML files The result is a well performing, fully functional AIML interpreter tailored around AIML 1.0 specification. A chatbot is software that is used to interact between a computer and a human in natural language. Naturally, it can extend daily life, such as help desk tools, automatic telephone answering systems, to aid in education, business and e-commerce[4]

Chat bots typically provide a text-based user interface, allowing the user to type commands and receive text as well as text to speech response. Chat bots are usually statefull services, remembering previous commands in order to provide functionality. When chat bot technology is integrated with popular web services it can be utilized securely by an even larger audience. The college enquiry chat bot will be built using artificial algorithms that analyzes user‟s queries and understand user‟s message. This System will be a web application [3] which provides answer to the query of the student very effectively. Students just have to put their query to the bot which is used for chatting. The system will use the artificial intelligence algorithms to give appropriate answers to the user. If the answer is found invalid, then some system to declare the answer as invalid can be incorporated. These invalid answers can be deleted or modified by the admin of the system. The student will not have to go to the college for enquiring something. Student can use the chat bot to get the answers to their queries. Students can use this web based system for making enquiries at any point of time. This system may help students to stay updated with the college activities[5]

III. PROPOSED WORK

AI has reached a stage in which chatbots can have increasinglyengagingandhumanconversations,allowing businessestoleveragetheinexpensiveandwide-reaching technologytoengagewithmoreconsumers.

Chatbots are particularly well suited for mobile — perhapsmoresothanapps.Messagingisattheheartofthe mobile experience, as the rapid adoption of chat apps demonstrates.

The chatbot ecosystem is already robust, encompassing many different third-party chat bots, native bots, distribution channels, and enabling technology companies.

We need a trained professionals to answer the queries on the answering side on the other side this system understands the queries and answer accordingly, which saves a lot ofman-power

IV. FEASIBILITY STUDY(FAD)

As businesses continue to automate their sales and customer support services in order to reduce costs, chatbots and other intuitive artificial intelligence programs are becoming more popular.
Chatbots are capable of helping businesses significantly cut labor costs.While complete automation of the customer service workforce is not feasible, automating a portion of the customer management and sales positions in the US is possible through chatbots and other automation technology, and this would result in considerable savings forbusinesses.
Many companies have invested more resources in better training for their customer service representatives, but the recent wave of artificial intelligence (AI) programming is changing the game, mainly throughchatbots.

V. FEASIBILITY STUDY(FAD)

As businesses continue to automate their sales and customer support services in order to reduce costs, chatbots and other intuitive artificial intelligence programs are becoming more popular.
Chatbots are capable of helping businesses significantly cut labor costs.While complete automation of the customer service workforce is not feasible, automating a portion of the customer management and sales positions in the US is possible through chatbots and other automation technology, and this would result in considerable savings forbusinesses.
Many companies have invested more resources in better training for their customer service representatives, but the recent wave of artificial intelligence (AI) programming is changing the game, mainly throughchatbots.

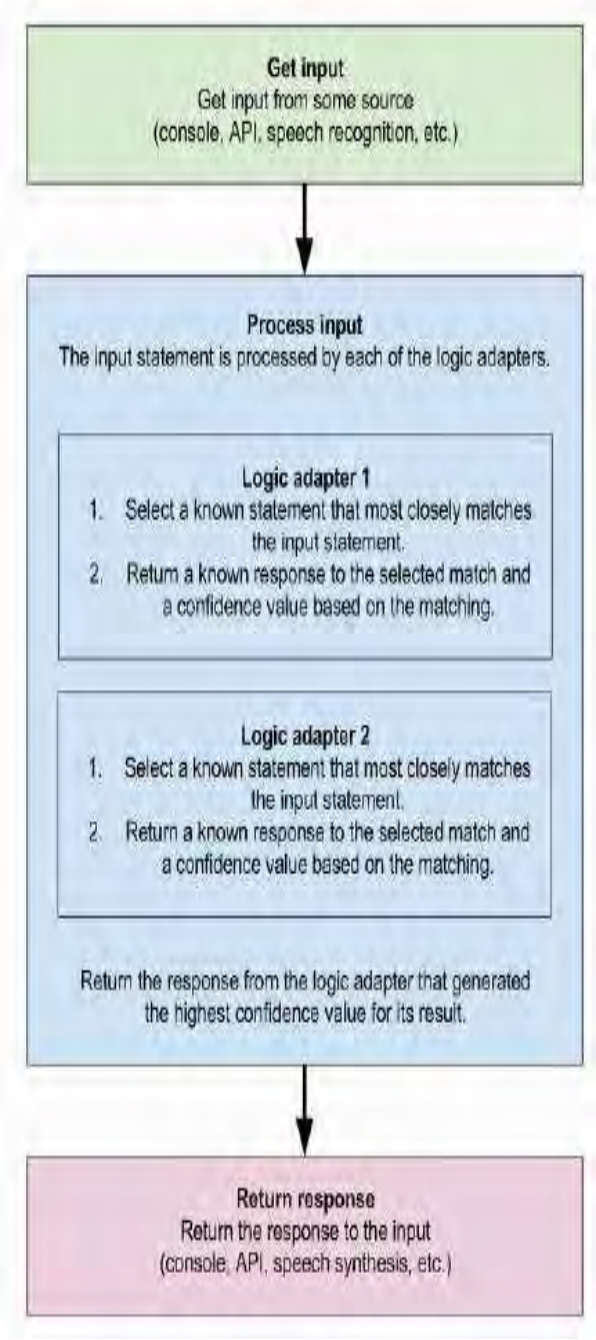


Fig-1

Algorithm

Step 1: Start.
Step 2: Get the user query. (INPUT)

Step 3: Pre-processing of the query
Step 4: Fetch the remaining only keywords from the query
Step 5: Match the fetched keywords with the keywords in Knowledge base, and provide an appropriate response. The keywords will be matched with the help of keyword matching algorithm.
Step 6: Return the query response as an output to the user.
Step 7: Exit.

VI. RESULTS

This is the home pages when you visit the site

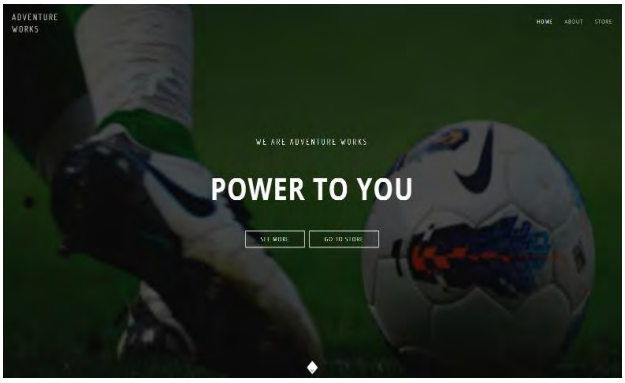


Fig-2.1



Fig-2.2

The registration page. User needs to enter his email id, username and a password to create a new account and to view the store.



Fig2.3

The congratulation page after you have registered



Fig2.4

If the user has an account, he can enter his login details to login and view the store.

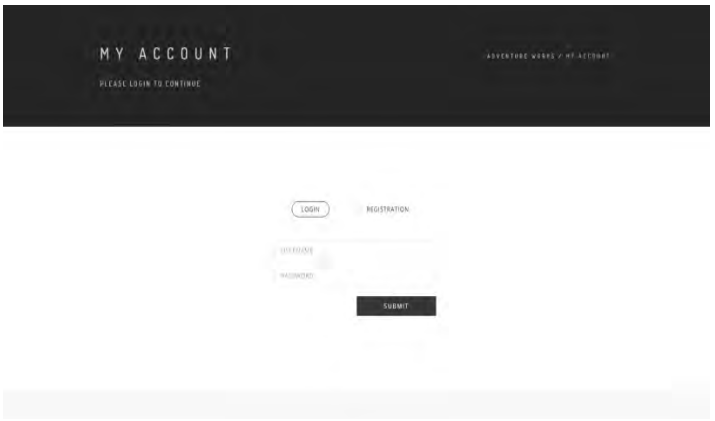


Fig2.5

Error page is you have not successfully logined

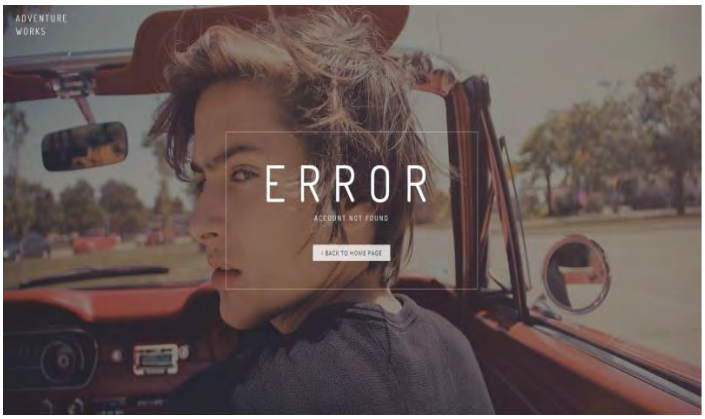


Fig2.6

Database Screenshots

The data base of the project where the registration is stored.

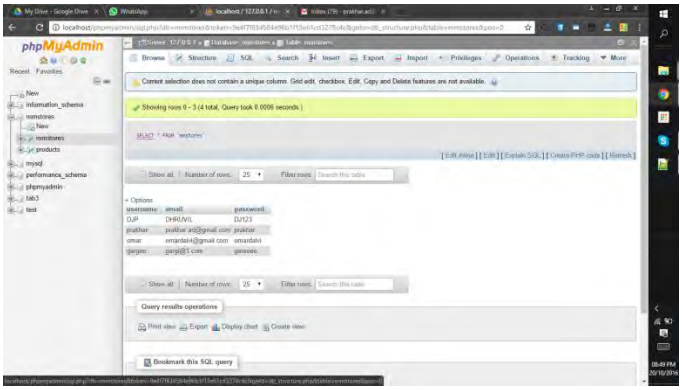


Fig-2.7

CART DATABASE: The items added to cart are listed in a table.

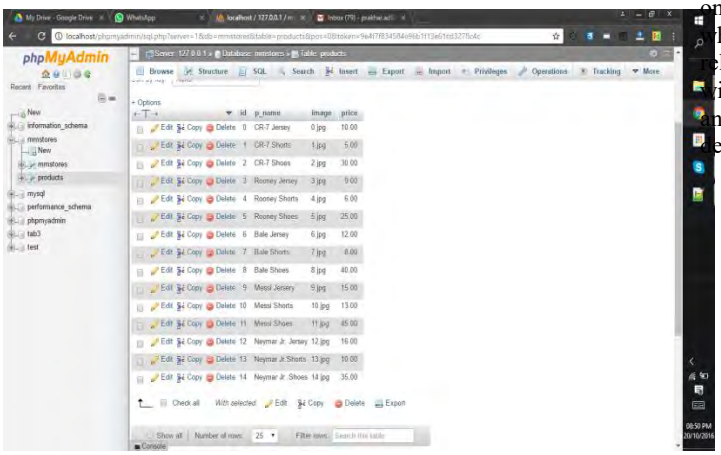


Fig-2.8

User Interface: When you visit the chat-bot for the first time

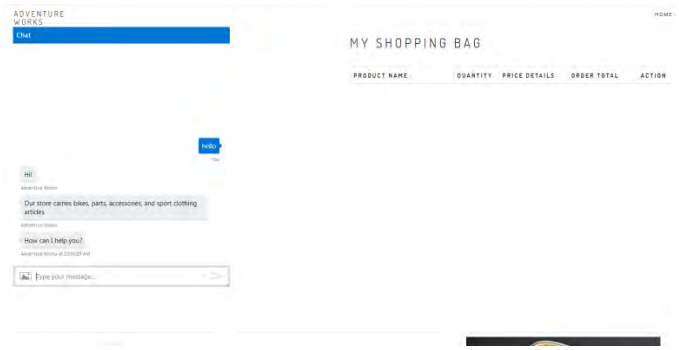
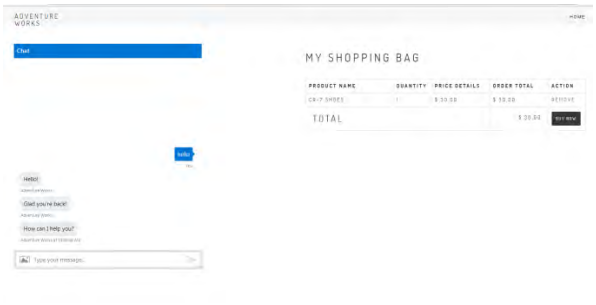


Fig-2.9

When you revisit the Chatbot after ending your first session



VII. CONCLUSIONS

The main objective of the project is to develop an algorithm that will be used to identify answers related to user submitted questions. The need is to develop a database where all the related data will be stored and to develop a web interface. The web interface developed will have two parts, one for simple users and one for the administrator. A background research took place, which included an overview of the conversation procedure and any relevant chat bots available. A database will be developed, which will store information about questions, answers, keywords, logs and feedback messages. A usable system will be designed, developed and deployed to the web server.

REFERENCES

[1] Emanuela Haller and TraianRebedea, “Designing a Chat-bot that Simulates an Historical Figure”, IEEE Conference Publications, July2013

[2] MajaPantic, ReinierZwitserloot, and Robbert Jan Grootjans, “Teaching Introductory Artificial Intelligence Using Asimple Agent Framework”, IEEE Transactions On Education, Vol. 48, No. 3, August2005.

[3] S. J. du Preez, M. Lall and S. Sinha, "An intelligent web-based voicechat bot,"EUROCON 2009, EUROCON '09. IEEE, St.-Petersburg,2009.

[4] Y. Chen, W. Wang and Z. Liu, "Keyword-based search and exploration on databases," 2011 IEEE 27th International Conference on Data Engineering, Hannover,2011.

[5] BayuSetiaji& Ferry WahyuWibowo “Chatbot Using A Knowledge in Database”2016 7th International Conference on Intelligent Systems, Modelling andSimulation.

[6] Artificial Intelligence: A Modern Approach(By: Stuart Russell, Peter Norvig)

[7] Introduction to Artificial Intelligence(By: Philip C. Jackson)

[8] Designing Voice User Interfaces: Principles of Conversational Experiences By CathyPearl

[9] Bot Business 101: How to start, run & grow your Bot / AI business By EkimKaya

[10] exam. Rocky Nook, Inc., 2014.

UI based Content Management System

Saurabh Singh, Shubham Vishwakarma, Ajaykumar Verma, Rahul Singh, Dr. Sheetal Rathi

Department of Computer Engineering and Technology, Thakur College of Engineering and Technology, Mumbai, India

saurabhasingh1997@gmail.com, vshubum@gmail.com, ajayverma1244@gmail.com, rs315@gmail.com, sheetal.rathi@thakureducation.org

***Abstract*—Content management is a collaborative approach where multiple users can log on and contribute or edit content to be published. It reduces its reliance on front-end engineers for making changes in website. The backend services are provided by CDA that support management and delivery of the content once it has been created isn the CMA. The CMS is being considered as a platform, and it could help organizations use the web which the easiest way to reach users in challenging times and even give them a competitive advantage over competitors. Both enterprise content management and web content management systems have two components: a content management application (CMA) and a content delivery application (CDA). However, there are certain challenges that need to be addressed like Large files slow down site performance, Images and videos don't look right as users switch from device to device; Content users upload varies in style and quality etc. Being an asset to the company, it is still overlooked as a source of competitive advantage. In this paper we use web development technology so that we can optimize website performance, minimize bandwidth usage and give users have a top-notch experience.**

I. Introduction

A content management system (CMS) is a software application or set of related programs that is used to create and manage digital content over web. CMS's are typically used for enterprise content management (ECM) and web content management (WCM) whereas it has other uses as well. But ECM and WCM are the most important aspects due to which CMS are in use nowadays. An ECM provides facilities as such collaboration in the workplace which can be achieved by integrating document management, digital asset management over web which is of great importance to the company and records retention functionalities, and also provide end users with role-based access to the organization's digital assets. Whereas WCM provides facilities such as collaborative authoring for websites and others as well which are of great importance to the enterprise using the CMS for their company. ECM software often includes a WCM publishing functionality which is an add-on with other features, but ECM webpages typically remain behind the organization's firewall.

There are various business benefits of using CMS. The business benefits include:

1) Reduce the site maintenance costs
2) Basically increase on content publish efficiency
3) Very-less long-term dependencies on specialist skill sets
4) Greater visual consistency due to templating
5) Distribution of content authoring to a wider team with minimal training

II. Background

Different features of CMS vary from CMS to CMS, but the underlying principle of operation remains the same such as indexing, searching, retrieving, managing different formats and publishment.

Indexing, retrieving and searching features index all data so that they can be easily accessed by search functions and let users to search usually done by keywords such as dates, attributes or author name.

Format management simplifies the process of transforming scanned documents and electronic documents into HTML or other format documents such as PDF.

Revision features deals with periodic updating and revision of the content after initial publication. Revision control keeps track of changes made by user.

Publishing functionality permits users to use a provided template or a set of templates authorized by the organization, as well as wizard and other corresponding tools for creating or modifying content.

Other popular features of CMS's include:

1) Search Engine Optimization supported website URLS
2) Online help facility is integrated along with the privilege of discussion forums.
3) Permission modules based on groups.
4) Supports customized & non-customized templates.
5) Simple to install (like wizard installation process)
6) Admin community with bi-lingual support.

III. Problem Identification

There exist many advantages of using Web based CMS technology over other existing systems or techniques in corporate world. Companies find it very pleasing and profitable to use due to the advantages listed below. These basically sums up what actually a web based CMS can provide to the existing Systems. These are the benefits which leads to enforcing a company to use a CMS system. Content management systems are basically often considered as a silver bullet which will eventually solve all your content problems. In reality existing/ other CMS is not enough as it comes with its own limitations which is been improved upon or being solved in our web based CMS.

The following is the top limitations which is been solved in our web based CMS.

a) A lack of an editorial control: It has consequence such as lack of central control which will ensure the quality and accuracy of copy produced. Which will eventually lead to contradictions and difference in styles of writing across the site.

Our Solution: Get an editor to control editing

b) Reputation of contributors: The greatest and biggest selling points of having a content management system is that they eventually allow anybody to post to your website which is one of its selling points and it comes with a disadvantage. Unfortunately, just because your staff can edit the site, does not guarantee that they will definitely do so and there is a danger of repudiation as well.

Our solution: Recognize the importance of the web

c) Poorly trained authors: Whenever an organization buys out a new content management system they almost always offer some form of training which is pretty essential with the complexity of existing CMS to use for the members of the organization. However, in many cases it is not enough.

Our solution: We will provide video training material

d) No clear calls to action: As we have already said, almost all the content providers are focus on conveying the information to others rather than meeting user needs. Until the user is ready to take those major steps they are left to wander around the website. However, they are also neglecting the business needs too. Even when content providers are thinking about calls to action, they are focusing on the big actions such as "contact us."

Our solution: Always guide the user to the next action

e) No sense of community: Increasingly, content management systems come with some great community tools. They have forums, comments and integrate with everything from Facebook to Twitter. However, great technology does not build great communities.

Our solution: Build relationship not functionality

IV. Literarture Survey

It is clear from the society survey and feasibility study of the system that this idea will majorly attract the people who are into their own startup firm, blogger, etc. And those who are not into development i.e. not having that much technical skills or do not want to invest their time into development can use this web based CMS. These are the target audience which we have aimed through web based CMS. In CMS there is no involvement of technical tools and it is very easy to deploy so startup idea can be converted into firm with no delay.

Any development of a system is directly or indirectly related to society so Society Survey is an essential part of the development phase. The development of a system may either affect the society in a positive way or a negative way.

a) Moving from CaaS to EaaS: There was a rage of headless CMS tools in 2016, especially among enterprise architects who were looking for maximum front-end flexibility. But the future trend involve that the intelligent WCM architectures will supply EaaS as well. It sounds easy but it is not easy as it sounds.

b) Shifting of Marketplace towards Complexity: This is current trend in the world of CMS but it will really boast in the upcoming future as digital increasingly becomes an arena of experts.

c) Entire Cloud shift towards PaaS: The current trend clearly shows that cloud has hit WCM. Nowadays most organizations are looking to implement CMS technology purely based on CLOUD.

d) WCM vendors offering Marketing Data Warehousing: The scope of WCM is a debatable topic. One key question that arises that "Should our WCM vendors should offer marketing or Not?" But, in reality, almost every enterprise will want to keep their data and content and also their engagement tiers distinct and separate which will eventually offer more sort of flexibility to the organization using the CMS.

e) Growing Emphasis on Sharable Micro Experience: It can be said to be a sort of prediction but in current trends in enterprise world regarding CMS.

If the development is not acceptable to the society then there is not much scope of it being accepted in the world or at larger scale so on basis of which it can be concluded whether to undergo development phase or not. Thus, a survey must be carried out before hand of what they feel of this development and take their considerations into account before starting the project development. Hence, we too asked people in our neighborhood about our project whether they would like to use a system which will help them to exhibit their products and services to the end users and users will get notifications of the same and the response was positive. In today's world, where majority of people use mobile applications for their day to day needs, the application for advance marketing will be a success.

V. Proposed Model

Our proposed model will overcome the following drawbacks as we have studied in the literature survey. Our proposed model is moving from CaaS to EaaS which is a very big advancement over the existing CMS Systems. The current trend in the world of CMS but it will really boast in the upcoming future as digital increasingly becomes an arena of experts.

It is clear from the society survey and feasibility study of the system that this idea will majorly attract the people who are into their own startup firm, blogger, etc. And those who are not into development i.e. not having that much technical skills or do not want to invest their time into development can use this web based CMS. These are the target audience which we have aimed through web based CMS. In CMS there is no involvement of technical tools and it is very easy to deploy so startup idea can be converted into firm with no delay. These observations from the Literature Survey can be used as the features that we will incorporate in our model.

Our model that is our website is focusing on incorporating the features we found to be unavailable in the current systems but may be able to see it in near future. Our proposed model will offer a great user experience for both the users using it and the client using our CMS to build their own site which is our goal.

With further release we are focusing to incorporate the future features which are not available currently in any CMS or if available then only in few of them which will be our selling point. Or the point which will make the entrepreneurs who are not willing to invest a much in setting up the online business.

We will also provide the basic or core features provided by most of the Content Management Systems. Which include:

1) Search Engine Optimization supported website URLS

2) Online help facility is integrated along with the privilege of discussion forums.

3) Permission modules based on groups.

4) Supports customized & non-customized templates.

5) Simple to install (like wizard installation process)

6) Admin community with bi-lingual support.

Following is the block diagram of our Content Management System:

The below block diagram gives a gist of information about the proposed model. To be true it acts as a skeleton to our Content Management System.

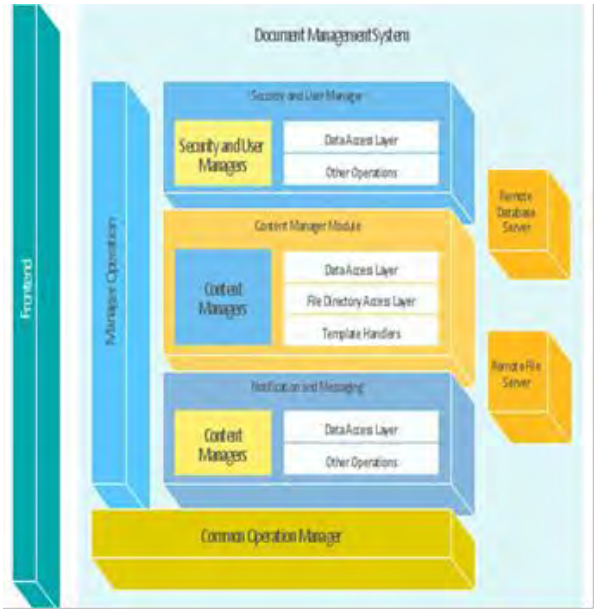


Fig. 1. Block diagram of our Content Management System

VI. Results And Discussions

There exist many advantages of using Web based CMS technology over other existing systems or techniques in corporate world. Companies find it very pleasing and profitable to use due to the advantages listed below. These basically sums up what actually a web based CMS can provide to the existing Systems. These are the benefits which leads to enforcing a company to use a CMS system.

1) It Improves Digital Marketing.

2) Reduces the time required and money as well.

3) Provides different access levels to various users.

4) Convenient and easy to use and content can be updated regularly without the involvement of the development team

5) Ability to make design changes.

6) CMS web design enables easy customizations and enhancements.

7) Ability to add new features via plug-ins.

8) Enables workflow management.

The end product will be almost same as the proposed model. But more likely that new features will be added as soon new technology is been introduced in the market. But our main objective/ focus is to provide the clients a free experience of Content Management System with great user experience.

The greatest and biggest selling points of having a content management system is that they eventually allow anybody to post to your website which is one of its selling points and it comes with a disadvantage. Unfortunately, just because your staff can edit the site, does not guarantee that they will definitely do so and there is a danger of repudiation as well. But our System does provide ability to control and make the employees of the stakeholders to make changes or updates to the System, so that the system can be up-to-date.

The following are the images of a dummy blog website been built on our Content Management System,

- The first image is the dashboard of the Blog website been built on our CMS.
- The image clearly shows that it provides various features and controls to the user.



- The second image is the image which clearly shows that our system also provide profile view feature which is rarely found in a CMS.
- It also provides basic information about the user and activity performed by that particular user.



VII. Conclusion

A content management system (CMS) is a software application or set of related programs that is used to create and manage digital content over web. CMS's are typically used for enterprise content management (ECM) and web content

management (WCM) whereas it has other uses as well. But ECM and WCM are the most important aspects due to which CMS are in use nowadays.

While of all the existing CMS WordPress seem the best CMS platform compared to other CMS such as Drupal and Joomla platforms, there still are situations wherein you might not find WordPress a worthy choice.

Our proposed system stands out at this point. It offers both continuity and reliability which is rarely available together. But the success of any project considering the fact depends on human factors, proposed CMS is undoubtedly the ideal choice as a reliable CMS platform. After all, it is very straightforward and extremely easy to operate and thereby reduces the chances of the human bottleneck and on the other hand it increases the productivity to a great extent.

VIII. Future Scope

Based on our Literature Survey the CMS can see a variety of change in near future.

1) Moving from CaaS to EaaS: There was a rage of headless CMS tools in 2016, especially among enterprise architects who were looking for maximum front-end flexibility. But the future trend involve that the intelligent WCM architectures will supply EaaS as well. It sounds easy but it is not easy as it sounds.

2) Shifting of Marketplace towards Complexity: This is current trend in the world of CMS but it will really boast in the upcoming future as digital increasingly becomes an arena of experts.

3) Entire Cloud shift towards PaaS: The current trend clearly shows that cloud has hit WCM. Nowadays most organizations are looking to implement CMS technology purely based on CLOUD.

4) WCM vendors offering Marketing Data Warehousing: The scope of WCM is a debatable topic. One key question that arises that “Should our WCM vendors should offer marketing or Not?” But, In reality, almost every enterprise will want to keep their data and content and also their engagement tiers distinct and separate which will eventually offer more sort of flexibility to the organization using the CMS.

5) Growing Emphasis on Sharable Micro Experience: It can be said to be a sort of prediction but in current trends in enterprise world regarding CMS. This will eventually present two major challenges to the existing web based CMS systems in order to prevail and sustain in this trend. Which can be summarized as

a) Opening up new platforms on their own for the other non-enterprise contributors.

b) Providing omni-channel capabilities to preview the marketplace.

References

[1] https://www.trainingindustry.com/glossary/content-management-system-cms/

[2] http://www.accuraty.com/web-content-management-systems-intranet

[3] https://snap360.ca/website-design-and-development/content-management-systems-giving-you-control/

[4] https://www.smartbugmedia.com/blog/4-guidelines-for-choosing-the-best-content-mansagement-system-cms-for-a-business-website

[5] https://pimcore.com/nl/resources/blog/why-you-should-be-using-a-content-management-system-cms_a981

[6] https://www.makeuseof.com/tag/10-popular-content-management-systems-online/

[7] https://www.everyinteraction.com/definition/content-management-system-cms/

Webpage ranking Search Engine with SEO Suggester

Mr. TejasSutone
Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India
tejassutone27@gmail.com

Mr. VikasKaul
Assistant Professor Dept. of IT
Thakur College of Engineering and Technology
Mumbai, India
Kaul.vikas@thakureducation.org

Mr. Harsh Bharucha
Department of Information Technology
Thakur College of Engineering and Technology
hrshbharucha@gmail.com

Mr. S. K. Narayankhedkar
Professor
MGM COET , Navi Mumbai

Abstract—*Search Engine Optimization (SEO) Techniques often come under the scrutiny regarding their viability and ethical acceptance in the web community. Ever since internet marketing and web analytics became key components of marketing budget of a business, obtaining higher ranking in the Search Engine Result Pages (SERPs) has become highly competitive and consequently outsourced to professional firms. SEO is an evolving methodology which adapts itself to the latest algorithm used by the search engines. In this paper, we discuss the ever-changing SEO guidelines with reference to Google search engine. Current viability of old-school SEO techniques – keywords, backlinks and meta descriptions is examined with reference to regular search algorithm updates by Google. The system provides accurate results and suggestions as per users need. It is easy way to find relevant information about any subject at very low cost. Every user can login and then search for particular results.[1]The user gets a list of URLs according to the query asked that have been generated by the system. : The system has an advance feature that suggests website's owner for suitable keywords, meta descriptions for their webpage's SEO. The search engine is first fed with URLs of various website and stores this data in its database. After that the search engine allows user to search for particular data.*

Keywords: web mashup, meta data ,google updates, page rank , SEO , keyword , crawling , indexing .

I. INTRODUCTION

A search engine is an information retrieval system that has various website data stored in it. It shows user appropriate results on search. It also consists of website data. The search engine records no of times the url was clicked on, its goes to te web url to fetch its meta data and gives points according to page errors. Anseosuggester used with it allows user to enter a live url. Our system goes and scans the webpages extracts its meta features and provides seo suggestions.[2] The search engine is first fed with URLs of various website and stores this data in its database. After that the search engine allows user to search for particular data. The user can enter the data or text lines (query) to be searched. The system then uses this text and matches it with the content provided in the URLs fed in the database. The user query is matched with these contents and finally it generates a list of related URLs. The user can them click on the URLs thus generated and it will take them directly to the website.[3] The search is based on users demand. Once user clicks on the links he sees the web page and all the data on it. It also calculates and displays the web page rank based on website meta score and number of visits. The system has an advance feature that suggests website's owner for suitable keywords, meta descriptions for their website's SEO. Search Engine Optimization is constantly changing as new aspects come into play, and others go. Fundamentally, businesses need to recognize the two crucial fundamentals for high-quality SEO, on-page and off-page. On-page SEO refers to what a publisher can control directly, whereas off-page SEO basically relies on user behaviour, social engagements, visitors, and other publishers.[4]

II. BACKGROUND

A search engine results page (SERP) is the page displayed by a search engine in response to a query by a searcher. The main component of the SERP is the listing of results that are returned by the search engine in response to a keyword query, although the page may also contain other results such as advertisements.

Now customers don't just use search engines to find websites to buy things – they use search engines throughout the whole consideration and purchase process:

- For Awareness – with categorical & problem-based search queries like “Link Building Tools” or “Measure Social Media ROI”
- For Research and Product Evaluation – with queries about specific products and problems, like “ERP for Manufacturing”, or “Facebook Analytics for Retailers”
- For Final Product Selection and Reviews – with queries about specific brands, reviews, and comparisons.

III.IMPORTANCE

Search engine optimization (SEO) is the process of improving the visibility of a website or a web page in a search engines results page. Businesses operating in an online environment typically strive to promote their products and services on search engines. To gain more lead generation andultimately customer’s ongoing effort to improve a domain name, page rank and search engine performance is required.The system provides accurate results and suggestions as per users need.It is easy way to find relevant information about any subject at very low cost. Every user can login and then search for particular results .The user gets a list of URLs according to the query asked that have been generated by the system. : The system has an advance feature that suggests website's owner for suitable keywords, meta descriptions for their webpage's SEO. The search engine is first fed with URLs of various website and stores this data in its database. After that the search engine allows user to search for particular data.

IV. MODULES OF THE SYSTEM

User Login: Every user can login and then search for particular results.

Admin Login: Admin has a separate login and password to enter the admin area.

Website URLs: Admin can enter the site URL and Meta data into the database.

Search Box: The system provides a search box for user to write down their search query.

Real time Search List: The user gets a list of URLs according to the query asked that have been generated by the system.

Page Rank Calculation: The system automatically scans pages and records number of visits on pages as per their usage in search engine and allocates page rank to them automatically. This data is shown to the admin.

SEO suggestions: The system has an advance feature that suggests website's owner for suitable keywords, meta descriptions for their webpage's SEO.

V.PROBLEM DEFINITION

A search engine is an information retrieval system that has various website data stored in it. It shows user appropriate results on search..The search engine records no of times the url was clicked on, its goes to web url to fetch its meta data and gives points according to page errors.Anseosuggester used with it allows user to enter a live url. Our system goes and scans the webpages extracts its meta features and provides seo suggestions. The search engine is first fed with URLs of various website and stores this data in its database. After that the search engine allows user to search for particular data. The user can enter the data or text lines (query) to be searched.. The system then uses this text and matches it with the content provided in the URLs fed in the database. The user query is matched with these contents and finally it generates a list of related URLs. The user can them click on the URLs thus generated and it will take them directly to the website. The search is based on users demand. Once user clicks on the links he sees the web page and all the data on it. It also calculates and displays the web page rank based on website meta score and number of visits. The system has an advance feature that suggests website's owner for suitable keywords, meta descriptions for their website's SEO.

VI. METHODOLOGY

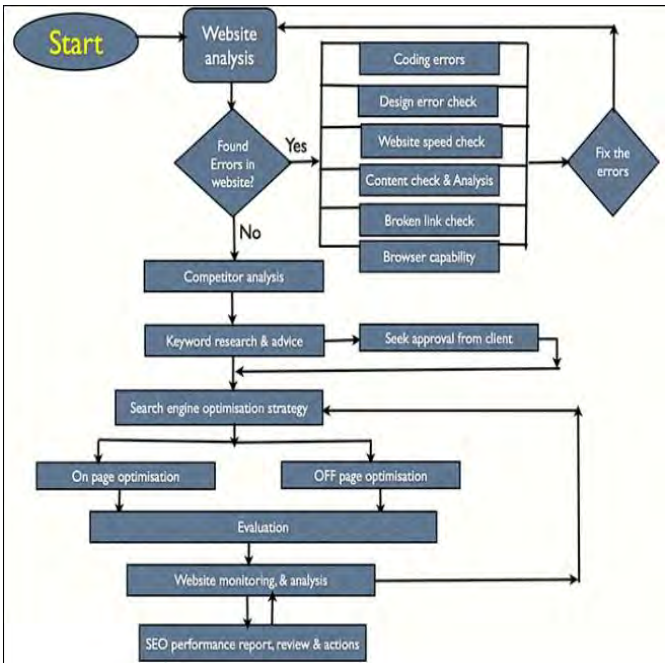


FIG 1-BLOCK DIAGRAM

Step-1: Website analysis
1.competitor analysis

2.keyword research and advice

Step-2: Installigvisual stdio10 windows Based system
1.First double click on setup.exe

2.After clicking you can click on install Microsoft visual stdio 2010

3.click next to start installation

4. Then I accept the licence term

5.install full feature

6.Then setup will start to install its components
7. Reboot ur computer and installation is complete

Step-3: create websiteseo optimization
1.website monitoring and analysis

Step-4: Seo performance review and actions

1. After installation, the website need to provide the following information:
   - Seo audit score
   - Keyword Discovery
   - Client Referrals
   - Page level competitive Analysis
   -

2. show search results pages input given to user

3. Get output as required to user

VII. HARDWARE REQUIREMENT

1. Need PROCESSOR – I3

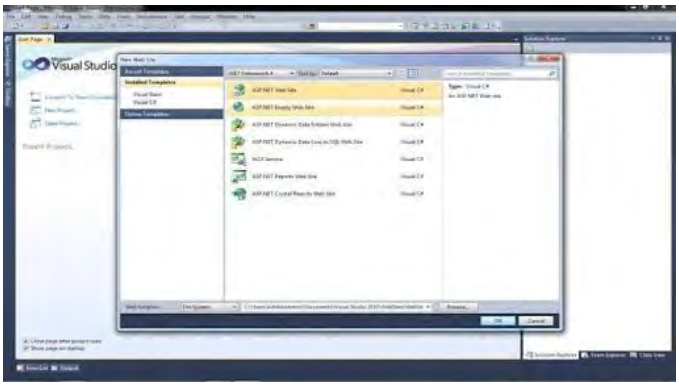
2 Required HARD DISK – 5 GB

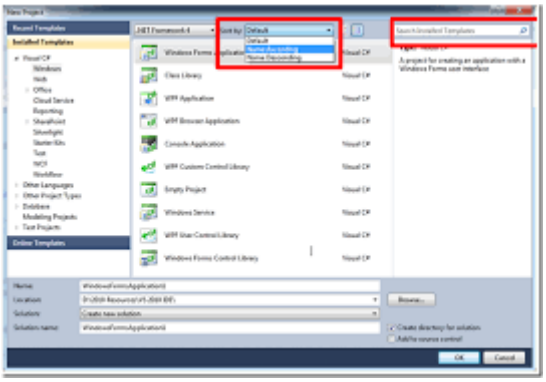
3. NEED MEMORY – 1GB RAM

VIII.GUI DESIGN (SCREEN SHOTS)









IX.CONCLUSIONS

Search Engine Optimization is constantly changing as new aspects come into play, and others go. Fundamentally, businesses need to recognize the two crucial fundamentals for high-quality SEO, on-page and off-page. On-page SEO refers to what a publisher can control directly, whereas off-page SEO basically relies on user behaviour, social engagements, visitors, and other publishers.The understanding the working of Crawling, Indexing, and Ranking is needed for Seo Practitioner for improve recognizing the Visibility of Website.

.

X. FUTURE SCOPE

The project has a vast scope in future. The system can be used by any organization to provide user the convenience while accessing the website.

It can also be used by educational institutes, hospitals or companies in their software or applications to be used by employees.It can be very well used by site owners for SEO suggestions.

XI. ACKNOWLEDGMENT

The systemprovides accurate results and suggestions as per users need .It also precisely describes the information that users are searches. The system gives results according to the best relations between the query asked and the information. Especially for website owners this system has a new feature whereby they get suggestions from it for various suitable keywords, titles, content etc for SEO purposes. Saves time as it provides results instantly.

XII. REFERENCES

[1] NirmalaShinge,Dr.NIleshMahajanAn Analysis Study of Web Page Ranking Algorithms 2016

[2] Sanjay Kumar, Dharmender Kumar A Review Paper on Page Ranking Algorithms 2015

[3] SeemaRani ,UpasnaGarg A Review Paper On Web Page Ranking Algorithms 2014

[4] Dilip Kumar, A.K.SharmaA Comparative Analysis of Web Page Ranking
Algorithm

Opinion Mining On Customer Feedback

Sanjana Puthran Thakur College of Engineering
and Technology India,Mumbai.

Isheta Tilara
Thakur College of Engineering and Technology India,Mumbai.

Shweta Mishra Thakur College of Engineering
and Technology India,Mumbai.

Purvi Sankhe
Thakur College of Engineering and Technology

Abstract—In an e-commerce environment where millions of transactions take place between the providers and users, a need for establishment of validity of the service provided arises. A customer feedback system has been provided by the marketplace operators in order to fulfill such need. But the feedback generated may not be always relied upon. The feedback may positively or negatively affect its sales, instead of showcasing the actual genuineness of the product or service, in customer's point of view. Our work proposes an enhancement to traditional feedback system by introducing an Trust Reputation System (TRS) which helps filtering out the valid consumer feedback using a set of algorithms, thereby creating a trust degree for the user.

Keywords **: Trust Reputation System, Opinion Mining, Sentiment Analysis.**

I. INTRODUCTION

The consumers in the online market face the problem of filtering out the best products from a list of variety of options.There are various marketplace operators who provide feedback system to help customer identify quality products, by reviewing the customer opinion and accordingly choose the product. Most of the consumers buy products based on product reviews. This either negative or positively affect the sale of the products. Also, this paves a way for spammers for decreasing the sale of the product . To eliminate this, the paper focuses on enhancing the feedback system by introducing the concept of trustworthiness .This can be done through Trust Reputation System. TRS are programs that allow users to rate each other. Using such methods can help decrease the number of spammers, thereby potentially increasing the amount of genuine reviews. The advantage of such reviews is it helps determining the genuiness of the product.

II. RELATED WORK

Sentiment analysis has been studied in wide area of domain such as movie review, teaching review, product review, e-learning, hotel review and many more. Most scholars focused to quantitative data analysis. However, some studies have been done on qualitative data using sentiment analysis, we found six works that mentioned the idea of using opinion mining and sentiment analysis in education. Algorithms such as Naive Bayes, k-means and Support Vector Machine are used in opinion classification. The paper also focuses on the truth reputation system. There exists several truth reputation system architectures having different algorithms to calculate the reputation score related to the product. Many authors [1],[2],[3],[4],[5],[6] have proposed in their work several TRS architectures with different algorithms to calculate reputation score related to the product. Also, a few academic work on Truth reputation system has been devoted to the inclusion of the semantic analysis of feedbacks in the calculation of the trust score of the product and specially the trust degree of the user. Even in studies attempting to provide more complex reputation methods, some issues are still not taken into consideration, such as the credibility of referees, the update of the trust degree of the user at any intervention, the age of the rating and the feedback or the concordance between the given rating which is a scalar value and the textual feedback associated to it. In contrast to the mentioned TRS, our

proposed design overcomes these issues and makes use

of an algorithm which includes analysis of textual feedbacks in order to calculate the trust degree of the user giving the feedback and a trustful reputation score for the product.

III. OVERVIEW

The consumers in the online market face the problem of filtering out the best reviews or feedback for purchase of the products. We try to eliminate the problem by listing out the best reviews so that it becomes easy for the customers to decide on a product by analysing other consumer experiences, by allowing them to post their reviews. Consumers dealing with the online market might sometimes buy substandard products. Though the e-commerce company provide facilities like return and exchange of products, the process becomes a tedious task sometimes. The project aims to provide the costumers an opportunity to select the desired products based on the rating of the item they wish or plan on to buy, which has been evaluated on the basis of rating and reviews contributed by the consumers with the help of a Truth Reputation System (TRS).

IV. FEATURES OF THE PROJECT

The Opinion Mining of our project will be based on Sentiment analysis algorithms & methods and also on Truth Reputation System algorithm .Trust Reputation Systems (TRS) will provide the necessary information to support relying parties in taking the right decision in any electronic transaction. In fact, as security providers in e-services, TRS have to faithfully calculate the most trustworthy score for a targeted product or service. Thus, TRS must rely on a robust architecture and suitable algorithms that are able to select, store, generate and classify scores and feedbacks.

V. PROPOSED WORK

In the proposed architecture, for each user who wants to leave a rating (appreciation) and a feedback (semantic review), we analyse the customers attitude towards a number of short and selected feedbacks and stored by product in the knowledge base. This user's review is going to be reached by any other user. Then, we suppose that we have a path relaying all the users (the nodes). As a result, we need to know the trust

degree of the user and determine the trust degree of the feedback.[4]

Trust Reputation System Design

A. Algorithm Description

The customer starts by giving a rating and a textual feedback about a specific product. When they click on submit, in order to validate the given information, we are going to redirect the user to another interface showing this message for example: "please give us your opinion about the following feedbacks before validating the information you gave below:" In this interface we will find chosen feedbacks from the database from different types. Those feedbacks can be fabricated in order to summarize numerous users feedbacks stored in the database. The generated feedbacks can be stored in another knowledge base. So as much as we add feedbacks in the ordinary database, we will fill the knowledge database with prefabricated feedbacks using text mining algorithms and tools. However, some users can give already summarized feedbacks that can directly be included in the knowledge database. Indeed, there are many text mining and data mining algorithms and tools that could search the most appropriate feedbacks that are first of all related to the product and that can recapitulate and summarize most of each type of the users‟ feedbacks.

Actually, before sending the customers feedback and appreciation about the product to the trust reputation system, we have to verify the concordance between them in order to avoid and eliminate contradiction or malicious programs attacking our system. In the redirected interface, we will display several feedbacks from different types. However, the user can specify the number of feedbacks to be liked or disliked. Of course, we can also specify the minimum and the maximum number of feedbacks to be displayed by the user.

In fact, we are trying through this redirection to detect and analyse the user intention behind his intervention on the e- commerce application. Hence, we examine and evaluate his intention using other pre fabricated feedbacks with different types. Of course, we have already the trustworthiness of each feedback. Consequently, we use our reputation algorithm studied in section [4.2] in order to generate the user trust degree which plays the role of a coefficient and then rectify his appreciation according to his trust degree and generates the score of the feedback. Indeed, each feedback has trustworthiness in a threshold [-5,5]. The closest is the trustworthiness to 5, the most trustworthy

the feedback is. The closest is the trustworthiness to - 5, the very untrustworthy is the feedback. If the feedback is trustworthy its score would be included in [0,5] else it would be included in [-5,0].[4]

B. TRS algorithm

Reputation algorithm used in this TRS is using semantic feedbacks analysis in order to generate a trustful reputation score for the product.Actually, we have 3 types of feedbacks:

** Positive feedbacks: represent opinions that expressing a positive point of view about the product. Those ameliorative opinions contain a positive content concerning the product. Then, the adjective positive is referring to the nature of the content of the feedbacks not its trustworthiness. However, each feedback whatever is its type can have either a positive trustworthiness or a negative trustworthiness. Either positive trustworthiness or negative one, it is gradual: it has degrees as float in a threshold of [-5.5].

**Negative feedbacks: represent opinions talking negatively about the product. Logically, the users giving such opinions are not satisfied of the commented product. This feedback could be telling the truth or a part from the truth or could be far from the truth. That's why, each feedback has its trustworthiness represented by a float number between
-5 and5.

**mitigated feedbacks: represent feedbacks that are talking positively about some aspects of the product and negatively about other aspects. They are also characterised by trustworthiness included in [-5.5].

**contradictious feedbacks: represent feedbacks with a contradictious content for example a feedback where the user is not talking about the specified product but another one or he/she is affirming that the camera of a mobile phone is great and later in the same opinion is saying that the camera is very bad. In fact, we have to start by detecting the contradictious feedbacks. Then we are in need of a semantic analysis algorithm and tool that can detect the contradiction in a specific content related to a product. We can personalize the analysis according to the product. For instance, if the user says that "the swimming pool of the hotel which does not afford one is not clean", the algorithm must be able to detect this great contradiction. We can give to the algorithm for each product as an input the property of the algorithm; if there is no similarity we can consider it as a contradiction. But the agreement includes the meaning of course. Because if the customer writes that the negative thing about this hotel is that there is no swimming pool. He is telling the truth then obviously the presence of an absent property in a feedback doesn't mean that there is a contradiction. Actually, before sending the customers

Feedbacks from other users, on which the current user approves of like or dislike. Thus at the end a trust score is created for the user, which in turn decides whether user's feedback should be posted or not postedAfter measuring feedback and appreciation about the product to the trust reputation system, we have to verify the concordance and the alliance between them so we don't have contradiction. After verifying the concordance between the appreciation and the textual feedback we are going to redirect the user to the selection of prefabricated feedbacks. Then the user is going to click on like or dislike according to each feedback. The event of click will be managed in order to get some information needed in the calculus of the trust degree of the user. The function uses as a parameter the id of the feedback in order to get from Knowledge base its trustworthiness. We need to get also the previous trust degree of the user if he has been already engaged in a transaction or he has used the application for rating purpose. The user choices either "like" or "dislike" is an important parameter to determine his trustworthiness.[4].

VI. RESULT

Initially, the user gives a rating and a textual feedback about the purchased product. Then we validate the information provided through an interface. In fact, in this interface we will find chosen feedbacks from the database from different types. The feedbacks can be used to summarize numerous users feedbacks stored in the database. The generated feedbacks can be stored in other knowledge base. So as much as we add feedbacks in the ordinary database, we will fill the knowledge database with prefabricated feedbacks using text mining algorithms and tools. Some users can give summarized feedbacks that can directly be included in the knowledge base . Before sending the feedback to trust reputation system it has to go through concordance test, so that no contradiction of rating and feedback occurs.

the concordance the feedback is sent to Trust Reputation System for further processing. Then if concordance test is satisfied, the user is redirected to another page containing a set of

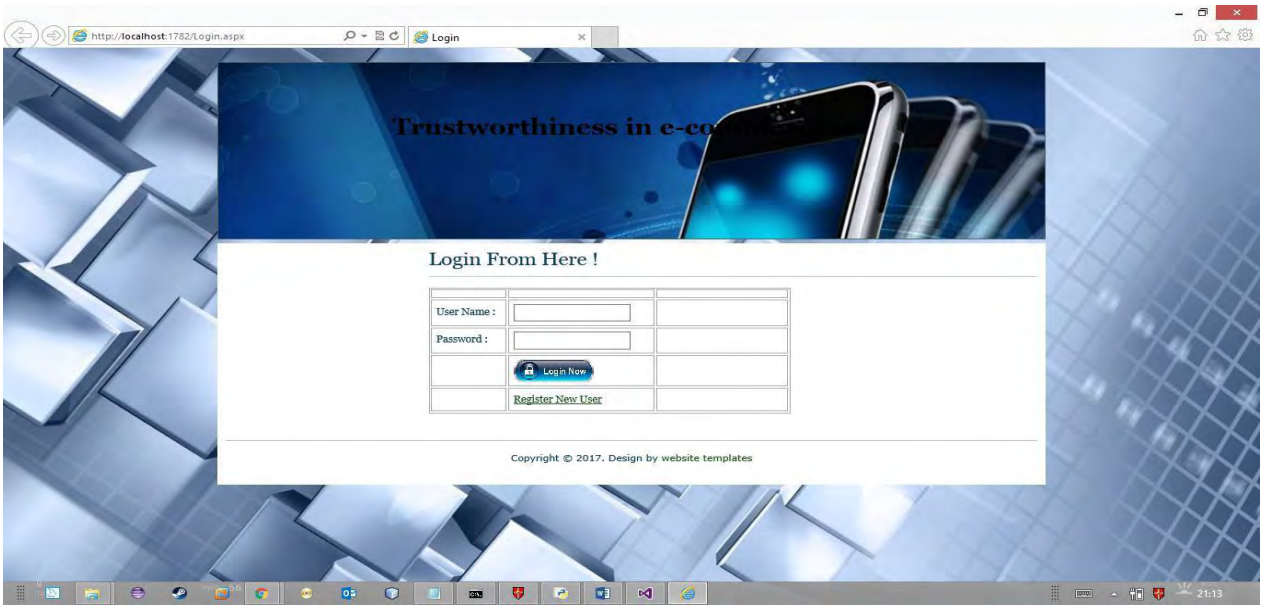


Figure1:Login Page

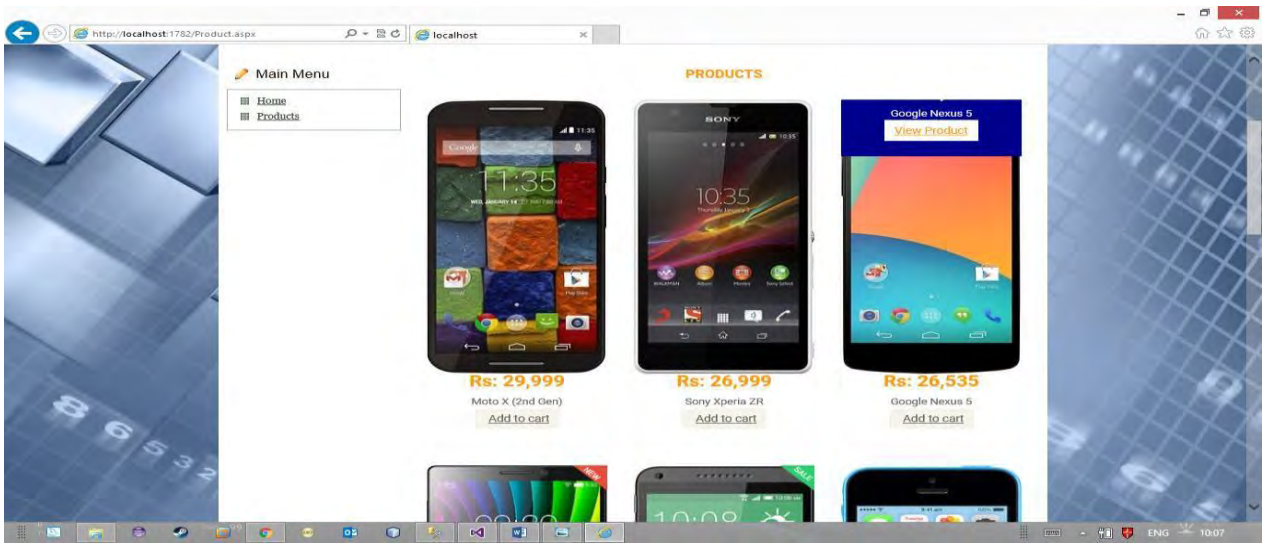


Figure 2.Product Details Page

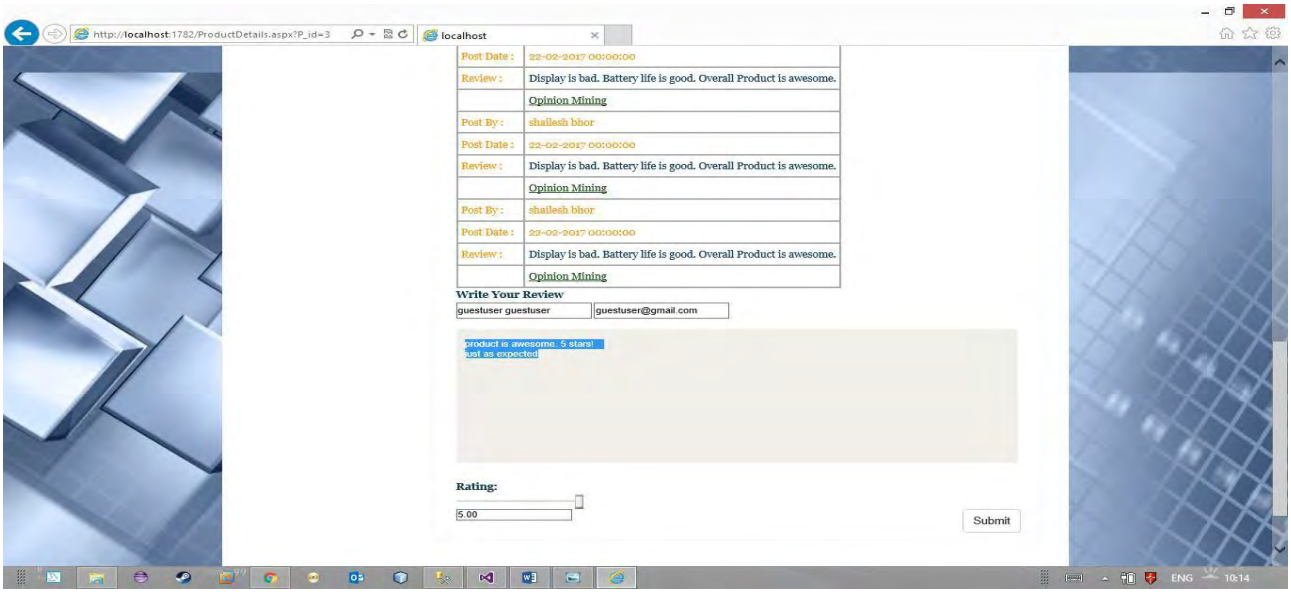


Figure 3:Feedback Page

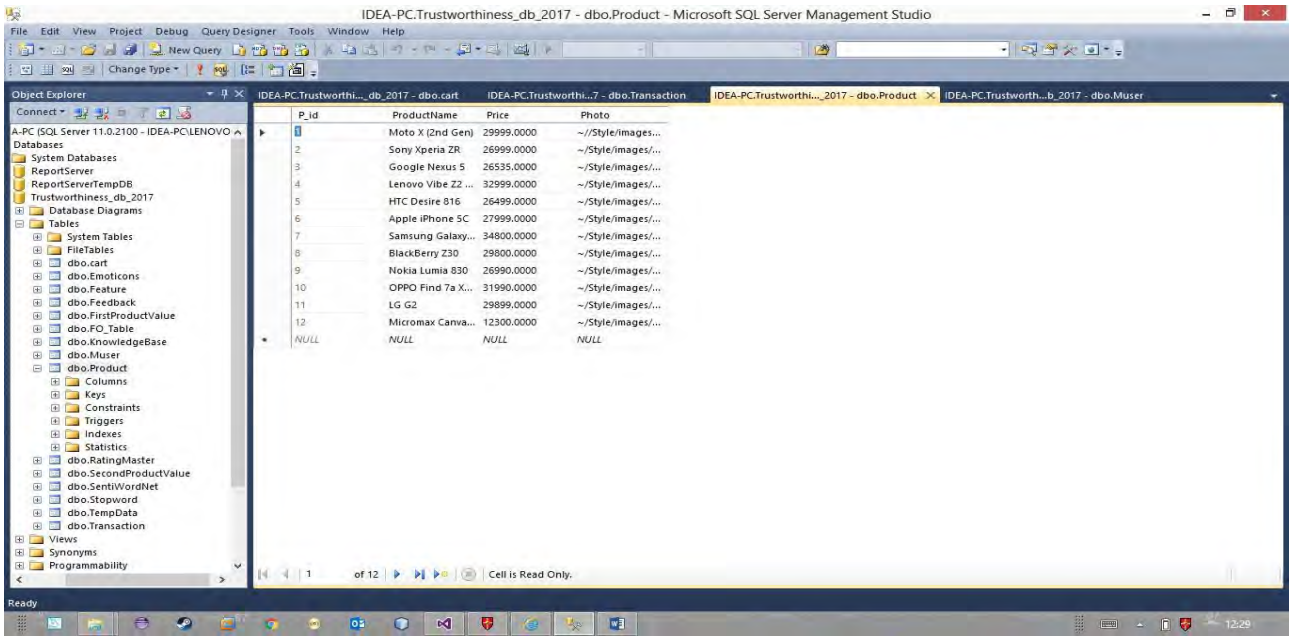


Figure 4:Database table for Product.

VII. CONCLUSION

Lack of information regarding particular products leads to wrong selection of product which in turn leads to huge holes in pockets of the customers. Thus we aim to provide the accurate and true reviews about the particular products which will help customers in picking up the right product. We attempt to calculate the trust degree of the user according to his subjective choice either “like” or “dislike” and according to the feedback.Those results such as trust weight and scores help users making a decision about purchasing or not a product from an e-commerce application. However those scores are not always truthful. Then, they can falsify the

weight and the ratings. Semantic feedbacks are more meaningful than single scores.

VIII.FUTURE SCOPE

The consumers dealing with our website would be able to access precise data and reviews of the consumers feedback and use it intelligently for product selection and for buying of it as well . This software would be useful for any similar e-commerce business dealing with problems regarding the issues of trustworthiness of reviews.The provision of visual representation can be used by customers to buy genuine products. On some extent it would also help the marketplace operators and vendors to filter out their potential customers. In today's time data is said to be the biggest asset for any company or organization. Thus, it is of immense importance to analyses the data and get some results out of it.

IX. ACKNOWLEDGEMENT

We sincerely thank to our guide Mrs. Purvi Sankhe, our HOD Dr. Rajesh S. Bansode, our Dean Dr. Kamal Shah and our principle Dr. B. K. Mishra for his/her guidance and support for carrying out our project work.

. REFERENCES

[1] The Analysis and Prediction of Customer Review Rating Using Opinion Mining Wararat Songpan Department of Computer Science, Faculty of Science, Khon Kaen University Khon Kaen, Thailand.

[2] A. Jøsang R. Hayward Simon Pope: Trust Network Analysis with Subjective Logic. Proceedings of the Second International Conference on Emerging Security Information, Systems and Technologies (SECURWARE 2008), Cap Esterel, France, August 2008.Mandalay, Myanmar

[3] Fereshteh Ghazizadeh Ehsaei, Ab. Razak Che Hussin: Acceptance of Feedbacks in Reputation Systems: The Role of Online Social Interactions Information Management and Business Review Vol. 4, No. 7, pp. 391-401, July 2012 (ISSN 2220-3796).

[4] A New Reputation Algorithm for Evaluating Trustworthiness in E-Commerce Context Hasnae RAHIMI1, Hanan EL BAKKALI2 Information Security Research Team (ISeRT) Université Mohammed V- Souissi, ENSIAS

[5]Co-Extracting Opinion Targets and OpinionWords from Online Reviews Based on the WordAlignment ModelKang Liu, Liheng Xu, and Jun Zhao

[6] A. Gutowska and A. Sloane: Modelling the B2C Marketplace: Evaluation of a Reputation Metric for e- commerce. Proceedings of Web Information Systems and Technologies - WEBIST , pp. 212-226, 2009.

Android Based Complaint Management System

Sangeeta Vhatkar
Assistant professor,IT Department Thakur College of Engineering And Technology University of Mumbai.

Mohitraj Shah
TEIT B(Pursuing)Thakur College of Engineering And Technology University of Mumbai.

ShashiKumar Thakur
TEIT B(Pursuing)Thakur College of Engineering And Technology University of Mumbai.

Shubham Singh
TEIT B(Pursuing)Thakur College of Engineering And Technology University of Mumbai.
.
.

Abstract-***In today's scenario, the complaint system is paper based which is time consuming. Also the paper based system does not track the status of the complaint. So in order to overcome this, complaint management system is based on android platform. The android is the handiest system in today's world. The light weight API supported by android system is Google Firebase API. So this makes the Complaint Management System more reliable and light weight. Application will help the college student as well as working staff of college to register their grievances about day to day problems in their class or college premises through a mobile application. It will provide students to deliver their complaints and problems to college authority as well as let the authorities to address the problem in a short period of time. The main objective of the Complaints Management System is to make complaints easier to coordinate, monitor, track and resolve, and to provide institute with an effective tool to identify and target problem areas, monitor complaints handling performance and make institute better working and studying place for college staff and students respectively . This application provides an interface to register one's complaint and follow it up. It provides an inbuilt camera module to click the picture of the problem as soon as detected.***

In this paper we are proposing the complaint management system on the android platform using firebase API.

Keywords: ***Android, online, complaint, management, responds, students, staff.***

I. INTRODUCTION

A complaint system is a set of procedures used in organizations to address complaints and resolve disputes. In most of the colleges in Mumbai university there is no direct communication between the college authorities and students in efficient way for solving their problems related to infrastructure and many other things i.e. for getting a problem solved one need to complain for same thing many times and yet there is no confirmation that problem will be solved or not.

Online Complaint System is software developed for any issues related to infrastructure**.** It facilitates to access the information of a particular problem to not only to staff but also for student. There is another part which is feedback, the student can give the feedback at anytime from anywhere to faculty. This feedback can be reviewed by the admin or the management committee of the institute through which the confidentiality of the feedback of the faculty can be maintained. It can also be useful in an organization or company at a certain limit not the whole

The proposed app uses android camera driver for the images to be clicked. The architecture of android which includes camera driver in kernel [1] has been used. The backend of the app is on firebase API provided by the Google [2], also since it is light API and provides various features like authorization, messaging, notification, storage.

II. RELATED WORK

In 2016, Mr. Ganesh R. Vishwakarma[1] proposed the system which gives various reports on student attendance. Our system will be able to keep track on various infrastructure problems with staff as well as students.

Online Complaint Management System [5], is based on the web portal. In today's scenario every system is on the android platform. So the proposed system is based on the android platform which is more user friendly and with rich GUI.

The study of Google Firebase API for Android [3] gives the information how to add firebase to our android project and how to use different features in our android application according to our need. Also our backend is based on firebase and we are using features like e-mail verification, storage, notification, real time database.

In complaint management system for Municipal Corporation [4], there is a provision for people to text there message and upload it for its solution by respective bodies and also keep a follow on it.
Since there is no mandatory of image upload, this application will provide a image upload as a mandatory way to complain. So that the admin can get the clear picture of the problem.

III. PROPOSED WORK

The main purpose of this project is to help the student in getting good infrastructure and getting their problems solved in online without going to the admin regularly until the problem is solved. By this system the student as well as admin can save his time and paper work. Its main purpose is to provide a smart and easy way through Android. It can also enhance in completion of work in constraint of time. It will also make complaints easier to coordinate, monitor, track and resolve. The requirement that are needed for building up the system in hardware aspects are Intel quad core 1.7 GHZ Processor or above, minimum 100 GB HD and 4GB of RAM . The software required is Android studio, firebase API, GCM server.

IV. MODULES OF THE SYSTEM

User Module: In this module we are authenticating the users by providing g-mail authorization through firebase. If username and password is valid then they will be taken to their screens. To identify user we have get detail of student and teacher during registration. The teachers are uniquely identified using their unique id number assigned by the college.

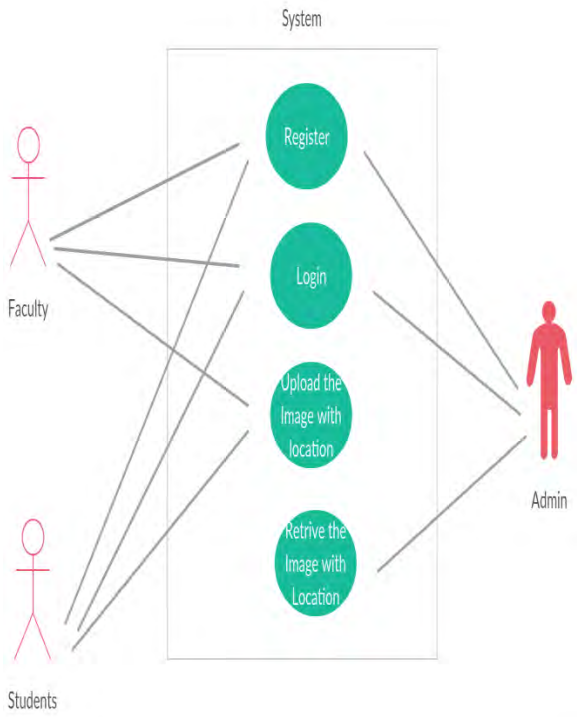


Fig. 1 Use Case Diagram of the Complaint management System

Database Module: The proposed system uses sql lite as its database because of their simplicity and flexibility. This module store every single information about students, faculty and model their data on specified operations. These operations can be storing images address in database or can be authentication credentials.

Admin Module: This module is designed for administrator of college, which receives all the complaint from the students and staff members through database. Admin also have their respective login and registration screen. The complaints that are seen by the admin are notified to the student and staff.

V. ALGORITHM FOR COMPLAINT MANAGEMENT SYSTEM

Step1. Registration.

The first step in this application is to get the student and teaching faculty to register. They need to first register their g-mail id in the database. The respective person will then provide his or her username and password for registration. An OTP would be then sent via e-mail address on the phone by the admin or faculty. The unique identity of the student is there class and roll number whereas for faculty is there id number. The admin can login using there g-mail id with their unique id as their unique identity. Now the users can login using their credentials.

Step2 Select image

The system provides two ways to select the image of the problems. The user(student/faculty) can either directly click the image from the in-app camera or he can get the image from gallery. The purpose behind the selection of image from gallery is that the user can click the image and then when he get the internet connection he can register the complaint, which enhances the flexibility of registering the complaint from any location.

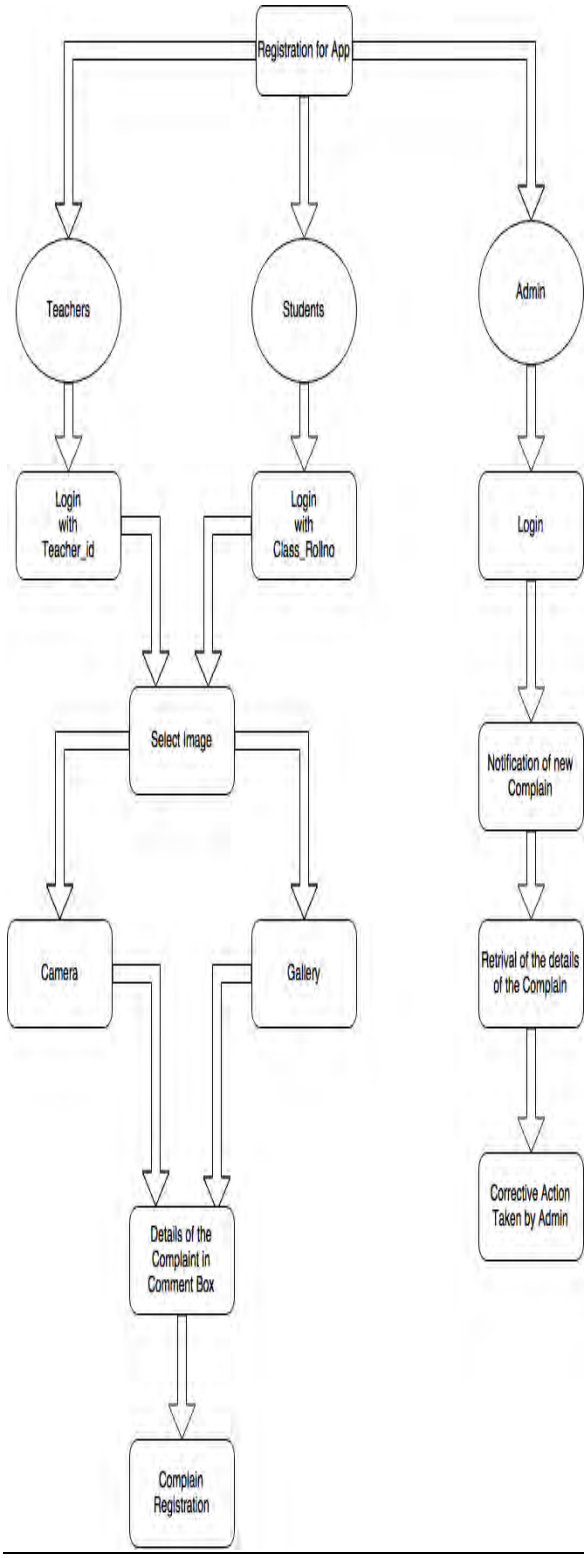


Fig.2.Flow-chat of the complaint management system

Step3 Complaint Description:

The app provides a comment box, to give the details of the problems or the images. It is provided because the admin will get the clearer picture of the problems. The comment box can also be used to give identity of the person, location of the problems or it can be used for time and date.

Step4 Registration of Complaint:

The complaints are registered using the image of any broken or not working appliances or furniture and its location. Once the complaint is ready the user requires an internet connection to upload the images which can be done in a single click. While registration of the complaint the database stores the unique identity of the person who complains, images url and the comment is stored as the string.

Step5 Notification to Admin

If the new complaint is entered in the database the admin is notified through app notification manager. If in case there is no internet connection in admin app then as soon as he connects to internet he gets notified.

Step6 Retrieval of the complaints:

Once the admin is notified for the new complaint the app will retrieve all the data of the complaint in the app and display it in the gallery format.

Step7 Corrective Actions for the complaint:

The admin can now process the complaint one by one and can take respective actions for the complaints registered.

VI. RESULT

The work focuses on making a paperless complaint management system on android platform using firebase API as a server.

Expected activities to be carried are as follows:
1) Registration and Login for the app
2) Click the picture of the problem(either broken/not working)
3) Add the comment with respective to the complained image and Upload it.
4) Complaint stored in database with Image and their respective id
5) Notification received to the admin.
6) Retrieval of the details of the complaint and corrective actions taken to resolve the complaint.

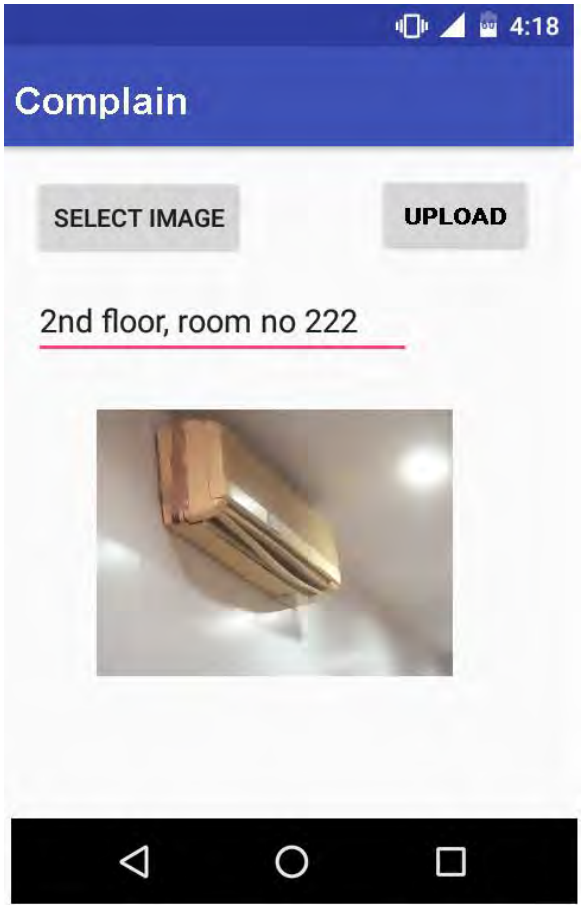


Fig 3.Teachers/Student App Interface

In the above figure(Fig 3) the AC is not in working condition so the user has clicked the image of the AC which is seen in the figure. The user has also provided the location of the same AC. The user clicks the upload button which then stores the image, unique id of the user and the respective comments.



Fig.4 Complaint Structure in Database

Whenever a new complaint is registered automatically the URL of the image is fetched and is posted onto the database along with its id and the comments related to it. Since the database is a real time database whenever a new post is uploaded automatically a notification is send to the admin.

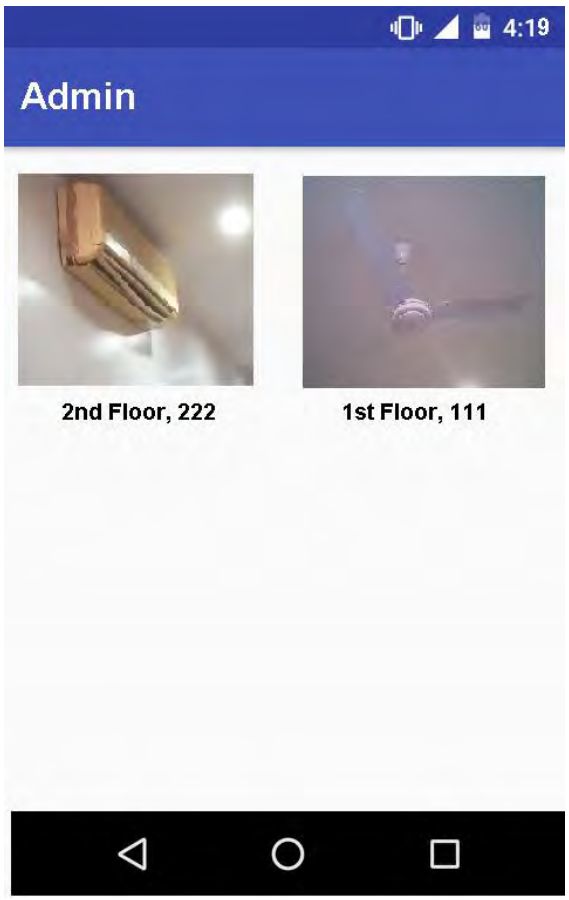


Fig.5 Admin App Interface

On receiving the notification of the new complain the admin gets the list of the complaint with the image and its location in the gallery format. In order to get the full details of the complain the admin needs to click on the respective images which will then show the complete description of the same.

VII. CONCLUSION

An Android based mobile application for College Management System is presented. The application offers reliability, time savings and easy control. It can be used as a base for creating and enhancing applications for resolving issues for colleges or any workplace. It provides high security and a system that reduces the work and resources required in traditional process. The proposed system provides the new way of computing and displaying an operation with responsive and attractive user-interface.

In future a real-time notification can be added, for every new complaint that is resolved by the admin at that time itself and a rating system for the solution of complaint. Also the structure of app can be modified to fit in any management system like hospital, companies, malls etc.

REFERENCES

[1] Li Ma, Lei Gu, Jin Wang, "Research and Development of Mobile Application for Android Platform" , *International Journal of Multimedia and Ubiquitous Engineering* Volume 9, No. 4(2014).

[2] Ganesh R. Vishwakarma , "Android College Management System" , *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* Volume 5, Issue 4, April 2016.

[3] Navdeep Singh, "Study of Google Firebase API for Android" ,*International Journal of Innovative Research in Computer and Communication Engineering.* Vol. 4, Issue 9, September 2016.

[4] Trupti Bomble, Ritika Raut, Ruchi Kanekar, Prof.Shekh AhmadHusen, "Android Based Complaint Management System For Municipal Corporation" , *Trupti Bomble et al. Int. Journal of*

Engineering Research and Applications ISSN : 2248-9622, Vol. 5, Issue 4, (Part -3) April 2015.

[5] Osman Nasr, EnayatAlkhider, "Online Complaint Management System" *International Journal of Innovative Science, Engineering & Technology*, Vol. 2, Issue 6, June 2015.

[6] Melina Fernandes, Merika Britto, Festo Francis**,** **"**Android Based Complaint Management System**:** People's Corner", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 3, Issue 4, April 2015.

[7] Android guide- Firebase (https://firebase.google.com/docs/)

[8] Connect to Firebase | Android studio (https://www.developer.android.com/studio/write/firebase.htm

Android Based Voting Application

Mary Margarat Valentine
Information Technology Dept
Thakur College of Engineering
Mumbai, India
e.marymargret@gmail.com

Rupesh Vishwakarma
Information Technology
Thakur College of Engineering
Mumbai, India
rupeshvishwakarma.mum@gmail.com

Aanchal Singh
Information Technology
Thakur College of Engineering
Mumbai, India
singhaanchal36@gmail.com

Abhishek Singh
Information Technology
Thakur College of Engineering
Mumbai, India
Abhisheksingh.xav@gmail.com

Abstract- **This project is about Online Voting System that is automation of present voting system which is done manually. According to this system, people who are citizens of India and age greater than 18 years can directly vote online instead of going to polling booths. For the process of automation, a database is used to store the details of voters regarding their names, ages and addresses and is maintained by election commission of India.**

The main objective of the proposed system is to increase the voting percentage and to provide facility at door steps, even to improve the security of voting system with valid votes. In this process of online voting, a person needs to fill the registration form to register. The details filled in the form are checked with details loaded in the database, if the match is found then username and password will be given to user to utilize his voting else if the fields of registration and database are not matched then transaction is discarded.

I. INTRODUCTION

Election allows the public to choose their representatives and express their selection for how they will be governed. Information and communication technology is linked for the proper execution of democratic rights. The principle of election process is fundamental to purity of democracy itself. The election system must be sufficiently sensible to withstand variety duplicitous behavior and must be appropriate so that voters and candidate can accept result of an election. A paper ballot is used to cast the vote in election and may be a piece of paper or small ball is used in voting. Actually it was a small ball used to record decisions made by voters. Each voter use only one ballot and ballot cannot be shared.at polling station voter cast their vote on ballot and put that ballot in box. Another voting system is EVM. The full form of EVM is “electronic voting machine”. Votes are stored by using Control unit and it is used by poll workers. Voters used Ballot unit which is placed at election booth. This system uses a six volt alkaline battery so that it can be easily used in rural area “Online voting” or “E-Voting” is also called as Internet voting. In this system by using internet connection voter can cast his/her votes. Gujarat was the first Indian state which implements this system in April 2011.Many people find the EVM system time consuming. This system is developed with the help of TCS. As in rural area, internet facility is not up to the mark as compare to the urban area because of the limited number of towers. So people cannot use internet for voting purpose. The current voting system is time consuming as voters have to wait in a long queue to cast their votes.

Voting was, is and always will be the most crucial aspect of any and every democracy. There exist many ways to cast vote for elections. These involve traditional means of voting which includes manual voting process that is using ballot papers. In ballot paper system user used to cast his vote by imprinting the stamp of his desired candidate on a paper and put it in a secret box. The other voting method is using electronic voting machine (EVM). EVM are majorly used when the population is large and it is practically impossible to count so many papers. So an EVM is

like remote where there are buttons or every candidate on a single machine itself, user just has to cast his vote by pressing he button in front of candidate. Advantage of using EVM over ballot paper system is votes need not be counted after wards instead they are counted while vote casting is done within machine itself. Security is an important part of voting. But, the issues of cheating in election and other malpractice leads in many difficulties and challenges.

I. PROPOSED SYSTEM

E-voting overcomes every major hurdle involved in the traditional system. Adapting to e-voting also increases the response of people for elections due to increased convenience. It is quicker and private than the traditional method. E-voting is also cost effective because it replaces the EVM machine and ballot papers. Also, the task force involved for conducting elections gets replaced except the higher officials. Replacing the paper ballot system also makes it eco-friendly. Authentication and cross verification becomes easy when the data is recorded digitally. The E-voting application also gives the real time results. Thus, by reducing the errors involved in counting.

The application of online voting system is android application designed with frontend technology Android Studio and the database used to store details is Google Firebase. The users in this project include voters and even administrators for handling system. The main goal of the proposed system is accuracy in the voting system, reduce the work pressure of maintaining details manually, make it more eco-friendly, to improve performance of system, to increase response, to cover people from every corner of country to use the application, and also to maintain security.

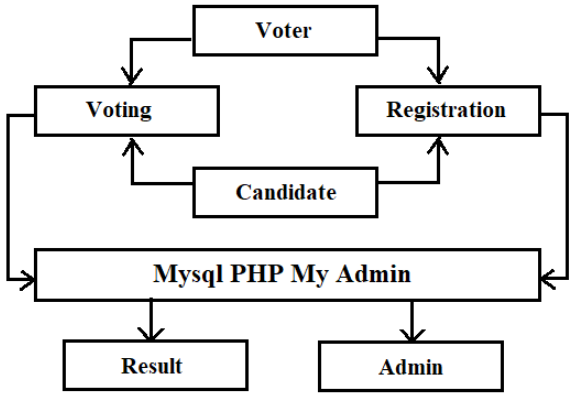


Figure 1: System Architecture

II. SCOPE OF THE PROJECT

Presently the voting consumes a lot of money which would be reduced in considerable amounts. Furthermore the flexibility for voter increases as they have to vote from any place, any time of the day when the election is scheduled. The project has various modules like authentication, registration and voting and displaying results. For voting and authentication fingerprint scanner is used to maintain the secrecy. The application will have its own database and maintain information about voters, candidate and all necessary information about upcoming elections.

III. PROBLEM DEFINITION

The proposed system encompasses the voting feature for an institution over the application. It has processes like registering, authentication, voting and displaying the results. Registration should include signing up of new user and should not allow existing users to create a new account. Registration should be done using email id phone number and fingerprint. In the proposed system authentication of the user should be done by OTP which will be sent over SMS and email. Also the fingerprint should be recorded using the cell phones fingerprint sensor to maintain the uniqueness. Phone number and email-id of each and every user will be unique and also as the user-id should be generated and provided by the application. Thus, use rid also remains unique. While voting, authentication should be done by users fingerprint to make sure the user as unique. As the voting process continues, simultaneously the votes will be recorded and counting will be done. After the completion of

voting process, the results should be declared on the application.

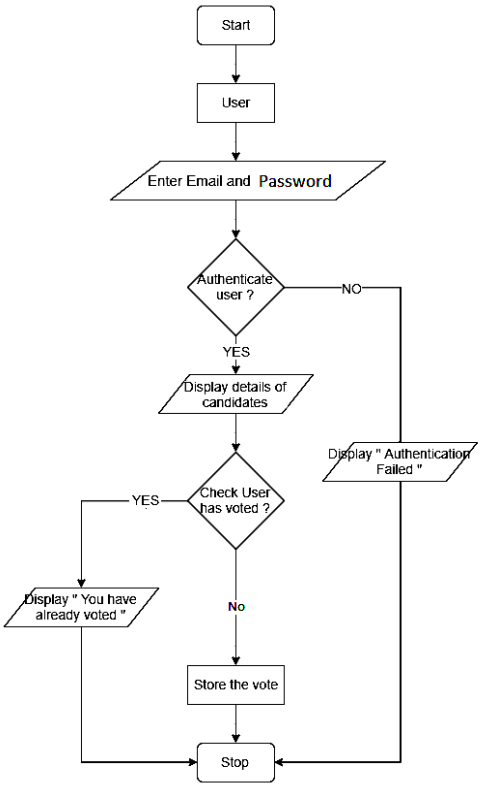


Figure 2: Flow of Proposed System

IV. METHODOLOGY

Initially user will be registering to our application using his/ her contact no. and email then application would be checking whether the email and mobile no already exists. If yes alert message is given to the user. If no, he can proceed to further authentication. Eventually user ID and password would be generated for the user. Now for voting process user has to log in to the application where he can read about the candidates information. While voting will be done using fingerprint and if the user has already voted he would be alerted about the same. After entire voting process is completed the votes will be counted and results will be displayed

i. Android application is to be created with the help of android studio which will contain different pages like login page, registration page, voting page, home page, Notification page, and Profile page of user as well of candidate.

ii. Google firebase is used to store, collect, analyze, process, modify, retrieve, and many other operation.

iii. With the help of Google firebase we will maintain the data of user as well as candidate, votes of the user will be stored on firebase, it also helps in counting the votes and provides the result instantly.

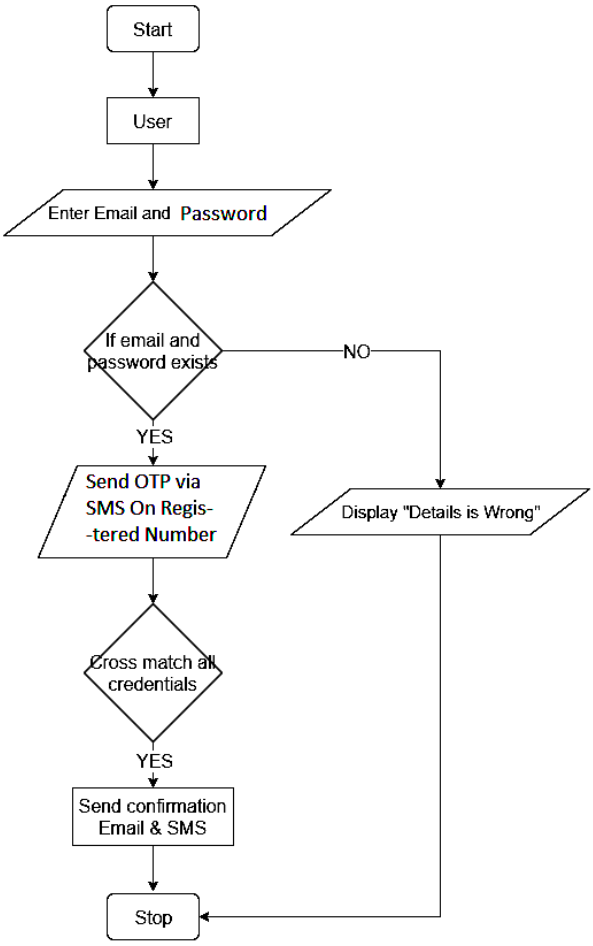


Figure 3: Flow of Registration System

V. Results and Discussion

A. Results

The details filled in the form are checked with details loaded in the database, if the match is found then username and password will be given to user to utilize his voting else if the fields of registration and database are not matched then transaction is discarded.



Figure 5.1.1 Login Activity

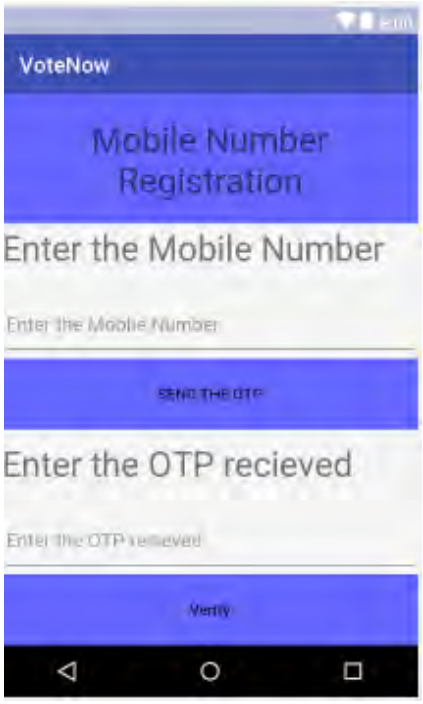


Figure 5.1.3 OTP Verification Activity

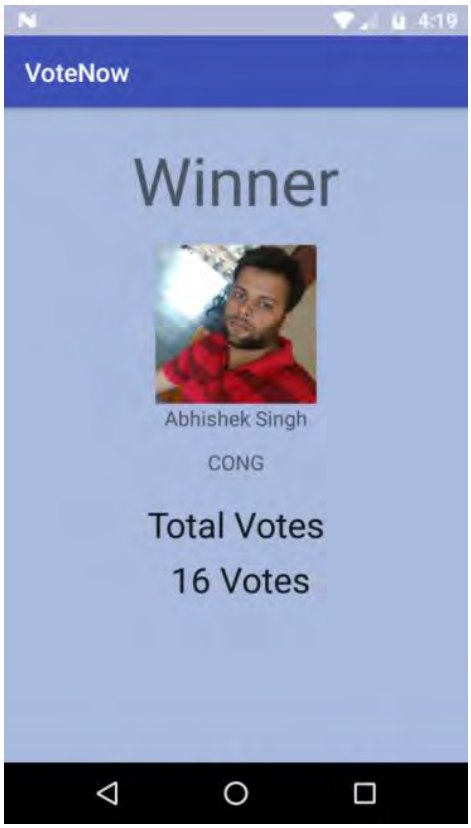


Figure 5.1.3 Result Activity

The authentication process uses email verification. Email provided by the user during sign up will receive a link which will make him a valid user. If email verification isn't done then he will be invalid user and his vote will be nullified.

The voting is done using fingerprint and votes are stored in firebase. Counting of votes is done simultaneously as the voting is done. If incase a valid user fails to vote between the allotted time then he will be sent an email and SMS just before 15 minutes the process ends, to remind him about that he has not yet exercised his power of voting.

B. Future Scope

With the help of this project one will be analyzing the traditional voting system and suggest a new and flexible way for voting. Thus, by this project there will be an alternate fast way to cast the vote via this application. With the help of the application it is expected that the cost and time required for conducting the election will get reduced. It is expected from the above project that it will help the institute and the organization to conduct the election as per there requirements in a very convenient way.

Currently the application is suitable to use only in the urban areas as many of people in our country lack basic digital services like data connection and slow data rates. Moreover, people also don't have mobile phone which has a fingerprint sensor in it. Many people living in rural areas also digitally illiterate. Same is not the case in urban areas as many have such phones embedded with fingerprint sensor and continuous data connectivity. So at initial stages it can be applied in urban areas.

With current advancement in technologies and reducing cost of devices it is expected that within few years everyone will be able to afford and buy a mobile phone which has an embedded fingerprint sensor. So the process of voting can be completely digitized across entire country.

VI. CONCLUSION

This project focused on the analysis and development of e voting application on an android platform which can tackle all earlier issues that were encountered in a conventional voting system. It will be definitely helpful for the users who wish to vote and the voting process will be made very easy by using this application. E-voting application systems have many advantages over the traditional voting system. Some of these advantages are less cost, faster generation results, easy accessibility, accuracy, and low risk of human and mechanical errors. It is very difficult to develop e-voting system which can allow security and privacy on the high level. Due to option of viewing candidate profile the voters get easily acquainted with all the candidates easily. Using android voting system percentage of voting increases and cost and time of voting process is decreases. It is easy to use and it is less time consuming. The system is first analyzed for

both SMS and internet based voting. It's then developed by implementing both techniques using android platform. Future development focuses to design a system which can be easy to use and will provide security and privacy of voters and votes on acceptable level by proper authentication and processing section. Fingerprint recognition is used as a form of biometric to recognize identities of human beings. Thumb impression of every individual is unique. Instead of providing authentication based off a single thumb print using both the thumbs guarantees greater amount of security and makes the process of authentication more efficient and evolutionary. Hence, it helps in maximizing the security. Our system maintains integrity of the voting process from the minute the voter casts his/her vote until the cast vote is registered. While observing fully fledged voting transparency, the system is capable of denying access to any unauthorized voter and also prevents multiple votes from same voter. With the use of the e-voting system, many of the issues that have challenged traditional voting systems in the past are thus resolved. We can conclude that through the online voting system the voting percentage increases and even the cost and maintenance are controlled and reduced. But at the same time voters need to be educated so that the present system can be utilized properly else complexity arises. And security should be implemented properly. Application should be adaptable and responsive during crash of system. As of now the target audience for this application are people who are digitally literate, have the access to unhindered data connectivity and also the mobile phones which have a fingerprint sensor, basically people living in urban areas as all of the above things are easily available with current advancement in technology and devices enabled with fingerprint sensor are affordable for many.

VII. ACKNOWLEDGMENT

Before going on with the contents of our project we would like to extend our gratitude to the people without whom this project would not have been possible. We sincerely thank to our guide Mrs. Mary Margarat for his guidance and support for carrying out our project work. We also thank the project coordinators Mr. Aditya Desai and Mrs. Neha Kapadia for arranging the necessary facilities to carry out the project work. We also thank our respective Dr. Rajesh S. Bansode (HOD – IT), Dr. R.R. Sedamkar (Dean Academic), Dr. B.K. Mishra (Principal) and the college management for their support.

VIII. REFERENCES

Journal References:

[1] Aashish Pedbulkar, Pranav Rajhans, Shweta Joshi, "Empirical study for e voting system" at IJARCETVolume 4, Issue 3, March 2015

[2] Dr. Aree Ali Mohammed, RamyarAbdolrahmanTimour, "Efficient E-voting Android Based System" at IJARCET Volume 3, Issue 11, November 2013.

[3] Kirti Autade, PallaviGhadge, Sarika Kale, Coauthors- Prof. N. J. Kulkarni, Prof. S. S. Mujgond,"Evoting on Android System", ISSN 2250-2459, Volume 2, Issue 2 in February 2012.

[4] Mary Maragarat, Hitendra Bhandari, Parth Modha," E-Voting System", ISBN 978-0-9972393-4- 8,"Multicon-w 2017" ICCCV in February 2017.

[5] https://pdfs.semanticscholar.org/881b/625d2d036cda a98a6ee332bd451ad403f3.pdf

[6] http://www.ijetae.com/files/Volume2Issue2/IJETAE_ 0212_40.pdf

[7] http://www.ijsr.net/archive/v4i6/SUB156101.pdf

[8] https://tcetmumbai.in/multicon/2017/ICCCV.pdf

College Management System And Data Analysis With Machine Learning

Meet Gadoya
Department of Information Technology,
Thakur College of Engineering & Technology,
Mumbai meetgadoya996@gmail.com

Raj Asha
Department of Information Technology,
Thakur College of Engineering & Technology,
Mumbai rajasha1711@gmail.com

Saurabh Maydeo
Department of Information Technology,
Thakur College of Engineering & Technology,
Mumbai saurabhmaydeo@gmail.com

Mary Maragarat
Department of Information Technology,
Thakur College of Engineering & Technology,
Mumbai mary.margarat@thakureducation.org

***Abstract*—In the process of college management, there are lot of challenges that are faced by the teachers and staff such as informing all students about some important event or taking attendance accurately. Also, there is no central storage where all student data is stored and organized. Also, there is no way to analyze the student data and take important decision based on the analysis. With the help of our College Management System, lot of benefits can be achieved by the college. First of all, student information will be stored in a single location and organized in a way which helps to access data easily. The process of taking attendance will be quick as well as done accurately and the students could also view their attendance at any time. Also, time of the teacher will be saved by easily posting notices to students as and when needed directly from his/her laptop or PC. Therefore, with the help of the college management system lot of time of teachers will be saved as well as data stored will be more accurate than before.We have created the user interface for the user to easily interact with the system. The system provides web pages with very simple forms for students and teachers. The database has been created with the tables linked together for easy access of data from the tables. Attendance module for taking and viewing attendance is also created along with the notice module for posting notices. Also, the profile page for the students and teachers has been created which provides general information about students and teachers.**

I. INTRODUCTION

College Management System(CMS) provides almost all of the important features required in the college management. It helps in carrying out major functions for day-to-day college life. It provides a different teacher, student and staff login. Each of them would have different access rights. It also provides a facility to take and view attendance as well as provide a facility of notice board. Students and teachers can communicate with a messaging module. A module is been used to analyze the student's performance based on overall attendance, pointers, co-curricular and extra-curricular activities. Any person can store and edit the profile. Notice can either be college related or specific department or class related.

EXISTING SYSTEMS

The traditional college management is a very tiring and time- consuming process. So, the basic motivation was to insert all of the college management functionalities into a single portal with better usability and efficiency.

In traditional management, teachers have to carry large book for attendance and manually find out the appropriate page for the class and the subject which could be a hectic task.

Here the notices are being displayed on notice board which has limited space.If there are multiple notice than some recent notice is been removed and student could be unaware of it.

In traditional management, there is no specific and proper mechanism to analyze a particular student performance. There is also no message functionality available.

All the college functions which are distributed in traditional approach are been combined into a single portal for better functionality.

LITERATURE SURVEY

We went through many previously done works related to this topic.

Paper published in International Journal of Advanced Research in Computer Engineering & Technology on April 2016 with title "Android College Management System" by author Vishwakarma Ganesh was useful. The major key findings were this system use data mining for remedial students and deals with IMEI number for better security.

Paper published in International Research Journal of Engineering & Technology on May 2016 with title "College Management System" by authors ShrikantPatnaik, KhushbooKumari Singh, RashmiRanjan, NikiKumari. The major key findings were this system also has hostel management.

Paper published in International Journal of Advanced Research in Computer & Communication Engineering on June 2013 with title "Student Information Management System" by authors S. R.

Bharamagoudar, Geeta R. B, S. G. Totad was useful. The major key findings were this system can represent progress of students. Paper published in International Journal of Computer Applications on July 2015 with title "College Management System" by author Lalit Joshi was also helpful. The major key findings were this system provide facilities such as attendance of teachers, college feedback.

PROPOSED SYSTEM

College Management System would help to carry out college functionalities properly and all activities under single portal will make it easier to use. It helps is better attendance viewing function and various functions can be applied on the

attendance database to find defaulters.It has proper module to analyze a student performance and provide some help to improve the overall performance.For analyzing, classification algorithms are being used. It requires active internet connection for working and the information about teachers, students and staff is been stored in the database.Users can communicate with others using a message module provided.

II.METHODOLOGY

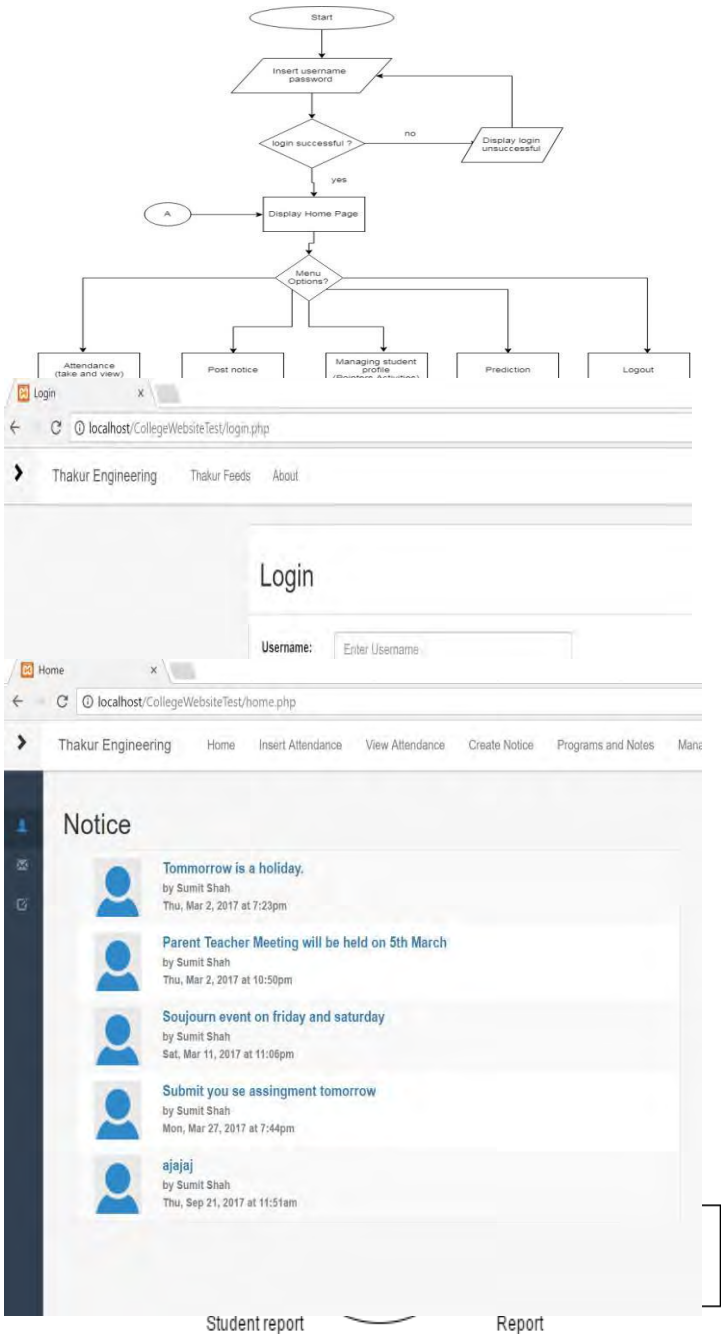


Fig 2.2-Block Diagram
Fig 2.3-Block Diagram

FEATURES OF THE PROJECT:

1. Authentication for students and teachers through a login system. So, users have to enter their username and password that is stored in the database.
2. Teachers can take attendance of students through a very simple user interface. Teacher have to select from multiple drop-down lists, the class, date, time, subject to display list of students to which he/she can give attendance by selecting a checkbox.
3. Teachers and students can view attendance through a simple interface by simple selecting roll number of students or name of student. Attendance of student can be displayed in two views, day- to-day attendance and aggregate attendance of a selected date range.
4. Teacher can post notices to a class or a group of classes to inform about some important event or task that is to be done by the students. Teacher simply has to write a title, message and select the classes to which he/she wants to send notice. Only the selected class students will receive the notice.
5. Student's result can be stored and viewed by teachers and students. Term test 1 and 2 as well as the semester results of each student will be stored in the student's profile page in textual as well as visual form such as bar graph.
6. Data analysis using machine learning algorithm will allow college to analyze the data of students and make various inferences from the data to take better decisions.
7. Classify data using classification algorithms in order to classify new data to make predictions which has better probability

III.DESIGN/RESULTS (SCREEN SHOTS)

Fig: 3.1: Login page

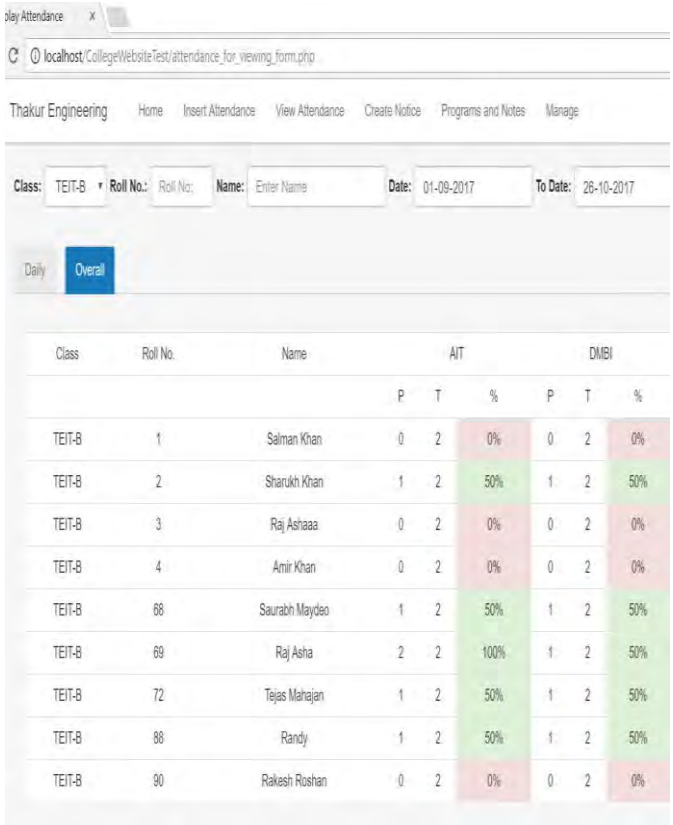


Fig: 3.3: Attendance page

IV.CONCLUSIONS

Colleges Management is a very crucial component for a college and it plays a very important role for the smooth functioning of the college and allows the college to manage things such as attendance of the students, informing students about important events and notices as well as communication between students and teachers etc. Therefore, for a smooth management of the college as well as ease of work and communication between students and teachers, college management system will provide a wide range of features and functionalities such as taking of attendance by teachers, viewing attendance by teachers and students, posting notices by teachers, sending of private messages between students and teachers, viewing results of students etc.Data Analysis component which includes prediction of pointer of students, classification of students into various grades as well as clustering of data allows the college to analyze the data and make various decisions for the students.

V. FUTURE SCOPE

The college management system can be further enhanced for carrying out almost all activities related to college. It can be used to keep track of all submissions of the students such as assignment, experiments and remedial submission.

A dashboard can be used by the teacher to get the overall idea of the progress of each of the student from the student's attendance, marks, pointer, submission status along with the student's participation in co-curricular and extra-curricular activities.

The entire website can have a mobile version in the form of an application which can be used by the students and teachers. This

will provide mobility and ease of access to information. Students can also get notification on their mobile based on events like when a notice is posted or his/her attendance is taken.

VI. ACKNOWLEDGMENT

We have taken up the implementation of this System. However, it would not have been possible without the kind support and help of many individuals and organizations. We would like to extend our sincere thanks to all of them. We are highly indebted of Mrs. Mary Margret for her guidance and constant supervision as well as for providing necessary information regarding the project and for her support in completing the project. We would also thank the college for providing the necessary facilities that was used by us during our project. We would like to express our deepest gratitude to our HOD Dr. Rajesh Bansode who managed to help us despite his busy schedule and took out time to provide guidance and inputs which were helpful in building this project. We would also like to thank our parents who gave us the moral support which motivated us to work even harder. Special thanks go to our classmates who have extended their help in accomplishing this undertaking.

VII.REFERENCES

[1] Tom M. Mitchell "Machine Learning"Publisher McGraw-Hill, 1997 ISBN 0071154671, 9780071154673

[2] W. Jason Gilmore "Beginning PHP and MySQL: From Novice to Professional" PublisherDreamtech Press; Fourth edition ISBN- 10: 8184897456 ISBN-13: 978-8184897456

[3] Sadeg BafandehImandoust And Mohammad Bolandraftar "Application of K-Nearest Neighbor (KNN) Approach for
Predicting Economic Events: Theoretical Background"ijeraVol. 3, Issue 5, Sep-Oct 2013, pp.605-61http://www.ijera.com/papers/Vol3_issue5/DI35605610.pdf

[4] SankarRajagopal "CUSTOMER DATA CLUSTERING USING DATA MINING TECHNIQUE" IJDMS Vol.3, No.4, November 2011
http://airccse.org/journal/ijdms/papers/3411ijdms01.pdf

An Intelligent College Enquiry Voice Chatbot

Vivek Kumar, Bharat Paliwal, Vedprakash Jha, Neha Patwari
Department Information Technology , Thakur College of Engineering and Technology,Mumbai

Abstract **- There are hundreds of different bots, developed for a variety of reasons, they range from hardwired programs with simply coded patterns to systems build upon embedded learning algorithms that continuously expand their language knowledge base. Bots are created purely for fun or as a part of interactive games, internet information services, websites guides e-commerce agents or more. Traditional technologies just used to respond to queries fired by the user by retrieving the answers from their databases but we need to develop a system which would interact with the user in an interactive manner as any other human will interact. If the response to the query fired by the user is not present in the database of the system then the system should respond to the user's query using internet. Chatbots typically provide a text-based user interface, allowing the user to type commands and receive text response as well as speech response. Chatbots are usually a stateful services, remembering previous commands and perhaps even conversation in order to provide functionality. When chatbot technology is integrated with popular web services it can be utilized securely by an even larger audience. Chatbots are simple to use, it usually introduces itself and we respond with a statement or question. The bot consults its knowledge base or programming languages and replies. The conversation continues as long as it is interesting or useful for human. This system is for college purpose and will respond to user query in an interactive manner. It will be having an ability to react according to the user emotions like if the user is sad then it will respond in a happy manner using happy type emoji's and messages. There will be an integration of the concepts of learning and emotions to a chatbot system incorporating spoken language system to address some of the inefficiencies found in these. This concept is important according to our vision, because the main goal of a chatbot should be to mimic real human conversation. More or less, we visualize this character as a personalization of someone who has some knowledge about something but is constantly learning everything that is said to it. This chatbot is having a simple user interface which responds to basic queries of users. There is no login require for using the chatbot to access the information. So there won't be any Registration and Login page for user since it is very boring procedure but there will be an admin login page through which an admin can put new queries and their answers in the database. A good database with queries and answers related to college as well as with some keywords tables will be created. There will be an addition of emoji's section in this chatbot through which it could easily respond to thequeries according to the user emotions.**

Keywords - *chatbot, chatterbot, pattern matching, keyword matching.*

1. I. INTRODUCTION

Chatbot is a conventional agent that is able to interact with users in a given subject by using natural language [1]. Normally, chatbot has the ability to answer questions from the user, provide comments, or bring a topic to be discussed with the user. Many chatbots have been deployed on the internet for the purpose of education, customer service, site guidance, or even entertainment functions. Existing famous chatbot systems include ALICE, SimSimi and Cleverbot. Chatbot can keep running on neighborhood PCs and telephones, however more often than not it is gotten to through the web. Chatbot is ordinarily seen as drawing in programming substance which people can converse with. It can be fascinating, rousing and captivating. It shows up all around, from old antiquated HTML pages to present day propelled long range informal communication sites, and from standard PCs to stylish brilliant cell phones. Chatbots talk in relatively every real dialect. Their dialect (Natural Language Processing, NLP) aptitudes differ from to a great degree poor to extremely shrewd wise, supportive and amusing. The same counts for their graphic design, sometimes it feels like a cartoonish character drawn by a child, and on the other hand here are photo-realistic 3D animated characters available, which are hard to distinguish from humans. And they are all referred to as "chatbots". College Enquiry Chatbot project will be built using artificial intelligence algorithms that will analyze user's queries and understand user's message. This system will be a android application which will provide answers to the queries of the users. User's just have to go to the chat screen to ask any query related to the college. The system replies using an effective Graphical user interface and a voice message which implies that as if a real person is talking to the user. The user don't have to register himself and login to the system since this is very overwhelming process. There is an Admin screen which will help admin to add new keywords and their appropriate response to the database. The user can query about the college related activities through online with the help of this application. The user can query college related activities such as date and timing of annual day, sports day, and other cultural activities, admission process and fees. This system helps the student to be updated about the college activities.

2. II. MOTIVATION

Artificial intelligence systems technology offers an alternative way to tackle complex ill-defined problems. They particularly useful in system modeling as in implementing complex mappings and system identification. Artificial intelligence (Al) systems comprise areas like, expert systems artificial neural networks, genetic algorithms, fuzzy and various logic and various hybrid systems, which combine two or more techniques. It gives us a touch of reality. It's advanced features like lip synchronization and audio-video-text response helps itto appear more real.

3. III. PROBLEM DEFINITION

Traditional technologies just used to respond to queries fired by the user by retrieving the answers from their databases butwe need to develop a system which would interact with the user in an interactive manner as any other human will interact.Also if the response to the query fired by the user is not present in the AML database of the system then the system should respond to the user's query using internet.

Phase 1(Information Gathering) : Study different papers related to chatbot and getting more and more information about it. To know about the Artificial Intelligence and types of algorithms which is required in creating chatbot.

Phase 2(Analysis& Design) : Develop the Graphical User Interface and analyze the internal structure and working model of the chatbot and how it will react to the user query. Making a list of different features that we can implement in our chatbot.

Phase 3(Implementation & Testing) : Developing an application in android and connecting it with a backend using volley library. Backend will be made using JSP and servlets. We will be creating Java Server Pages at the backend and different algorithms will be used for processing the user query and fetching the appropriate response for it from database. The code will be tested on real application and accuracy of responses will be measured. Then some methods and algorithms will be used for increasing the accuracy of response

Phase 4(Deployment) : Once the bot is built, it will be deployed to hosting environment which is our college website. The hosted environment must be stable and needs its own monitoring and develops support.

4. IV. RELATED WORKS

The biggest challenge of developing a successful question answering system may be the ability to translate question from natural language into a formal representation or equality. Utilities named element acknowledgment and connection extraction to separate bits of a formal portrayal, for example, a constrained arrangement of kinds of substances and a couple of sorts of relations communicated between them. Generally, conventional inquiry noting framework does not get a total importance portrayal of an inquiry and afterward attempt to answer it; rather, talk bot discovers bits of information from inquiries and tries to misuse what sources it has as confirmation for answers.Importantly, it does this in the presence of precision and recall errors that normally confound representation and reasoning systems [2]. Traditional question answering has three components such as question classification, information retrieval, and answer extraction. Question classification is used to classify the question based on the type of its entity. Information retrieval technique is employed to extract out relevant answers with query posted by their intelligent question answering system. Finally, answer extraction component is emerging topics in the question answering system where these systems are often requiring ranking and validating a candidate's answer[3]. Deep QA system goes further and discards the question classification operation, largely because it aims to deal with open domain problems. Instead, it puts emphasis on the question analysis and answer filtering[4]. In our opinion, that is a solution which combines bottom-up and top-down processing strategy. To be more specific, constructing query and issuing it to retrieval document is a bottom-up process; filtering the candidate's answer out with the information extracted from question is a top-down process. In the computer vision discipline, the appliances of bottom-up & top-down theory to some problems have achieved a higher level understanding [5]. Those researches inspire us to apply those theories to the research and implementation of a dialog system.

5. V. METHODOLOGY

Focusing on the interaction problem between chatbot and users, the basic challenge is how to understand commands that users issue to the chatbot. For a given user command, different chatbot systems have different solutions to parse. For example, some pretty basic systems can only distinguish "0" and "1" commands. To understand complex user commands expressed by natural language is still a terrible challenge for the natural language processing industry. But that doesn't prevent us from proposing our system framework to integrating more knowledge into chatbot. Our system architecture is showed in Figure 1.

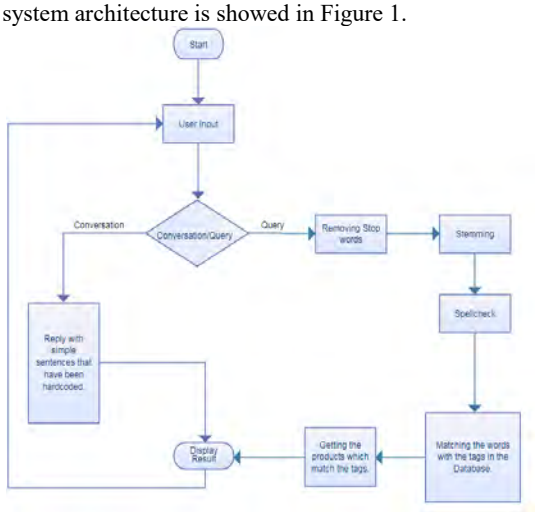


Fig.1. Chatbot system architecture.

The basic algorithm that will be implemented for working of this proposed system is as follows:
Step 1: Start the chatbot application.
Step 2: Get the query from the user in the text form.
Step 3: Send that user query to the application server.
Step 4: Pre-processing of the query will be done on server side using different algorithms.
Step 5: Fetch keywords from the query using algorithms.
Step 6: Process the keywords using keyword matching algorithm.
Step 7: Get the appropriate data from database.
Step 8: Return the data as a response of query to the application.
Step 9: Either finish the conversation or go on with another query.

6. VI. RESULTS

A usable system will be designed, developed and deployed to the college web server. A database is developed, which stores information about questions, answers, keywords, logs and feedback messages. The interface developed has two parts, one for simple users and one for the administrator.

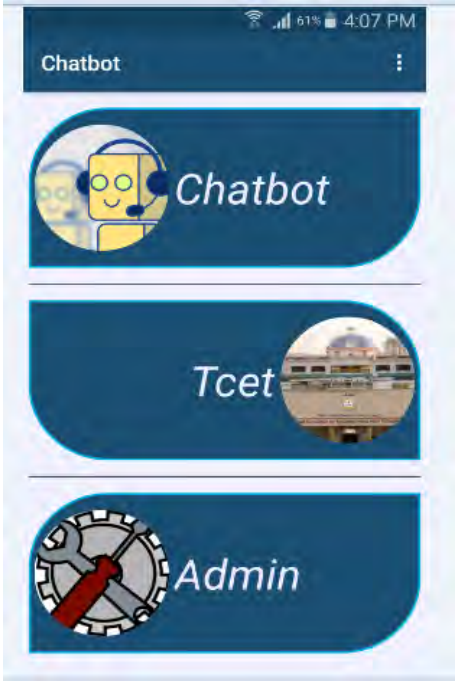


Fig.2. Home screen.

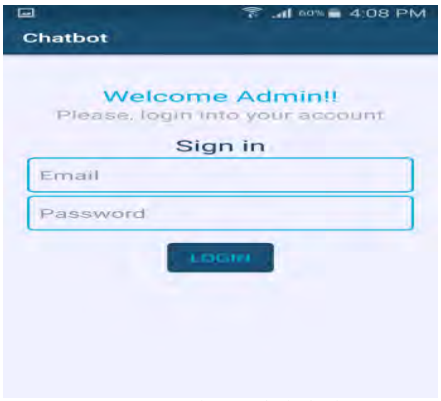


Fig.3. Admin login screen.

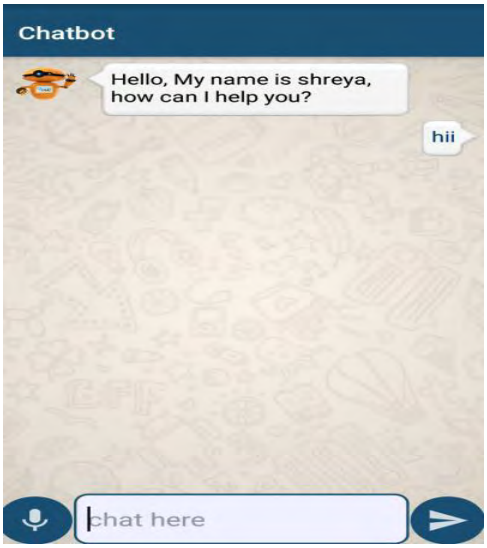


Fig.3. Chat screen.



Fig.4. Admin Screen.

7. VII. CONCLUSION

The main objectives of the project is to develop an algorithm that will be used to identify answers related to user submitted questions. A background research took place, which included an overview of the conversation procedure and any relevant chatbot available. This chatbot design may be able to have another direction of interface between human and

computer such as customer service. Another direction that can be made is to explore this design with additional tool such as web camera to make the agent can analyze the user's emotion and reaction.

Future Scope

Ranging from the fields of education to entertainment, from interactive conversation to work-related information, from admission procedure enquiry to comprehending quantum mechanics, Chat bots have considerable potential in various vital areas involving communication and sharing of information in an efficient manner.

REFERENCES

[1] Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with bilingual parallel corpora. In Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL), pages 597–604.

[2] Antoine Bordes, Nicolas Usunier, Sumit Chopra, and Jason Weston. 2015. Large-scale simple question answering with memory networks. arXiv preprint arXiv:1506.02075.

[3] Peter F Brown, Vincent J Della Pietra, Stephen A Della Pietra, and Robert L Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. Computational linguistics, 19(2):263–311.

[4] Zongcheng Ji, Zhengdong Lu, and Hang Li. 2014. An information retrieval approach to short text conversation. arXiv preprint arXiv:1408.6988.

[5] Yishu Miao, Lei Yu, and Phil Blunsom. 2015. Neural variational inference for text processing. arXiv preprint arXiv:1511.06038.

[6] Wenpeng Yin, Hinrich Schutze, Bing Xiang, and ¨ Bowen Zhou. 2015. Abcnn: Attention-based convolutional neural network for modeling sentence pairs. arXiv preprint arXiv:1512.05193.

Adaptive Automated Attendance Analysis For Education System

Kameshwar Gupta, Pratikkumar Mishra, Akanksha Das, Bhushan Nemade

Information Technology Department,

Thakur College of Engineering & Technology,

Mumbai, India

kameshwargupta710@gmail.com, pm24496@gmail.com, akankshadas654@gmail.com, bnemade@gmail.com

Abstract − **Adaptive Automated Attendance System is an advancement that has taken place by replacing the traditional marking attendance system. The proposed system aims at simplifying the attendance system and its management by face detection and recognition method. This system includes detection of faces through a high definition camera, the camera captures several faces of students within the class further these images are recognized by comparing them with the existing features in database thereby conducting the attendance. Existing algorithms for multiple object recognition like the Support vector machine, Trace transform, Independent component analysis and Naïve Bayes classification have multiple drawbacks such as extensive memory is required, lower speed, order of components is difficult to determine and loss of accuracy. The proposed system uses deep learning algorithm such as Convolutional neural network cascade which is used for face detection and Linear Discriminant Analysis is used for face recognition. This system is capable of detecting face in various angles, lighting conditions and has the ability to detect face even if there is an occlusion in the face up to some extent. Most of the existing systems are time consuming and require some kind of semi manual work from students or teachers. The proposed system aims at removing such work and providing better accuracy. The benchmark parameter is the number of successfully detected faces, previous system could achieve this parameter up to 70% whereas proposed system achieves it up to 80%. The proposed system provides a detailed reporting system with more accurate results and also records the monthly attendance and semester attendance which simplifies the task of identifying the defaulters.**

General Terms

Face Recognition, Viola-Jones Face Detection Algorithm, Convolutional neural network Algorithm.

Keywords

LDA, Ada-Boost, cascading classifier, background regions.

I. Introduction

Face recognition is an essential field in many applications, one of which is Attendance Management System .The process of taking attendance of the student in the classroom is a tedious job for teachers like calling out the names waiting for response and also maintaining the attendance till the month to generate the report. Analyzing them manually can lead to wastage of time and thus in order to save time and improve the accuracy, we have used the concept of face recognition based attendance system. Thus face detection and recognition module detect the faces from the image captured by the camera , and the image of the face is stored.

In Adaptive automated attendance analysis for education system, database of students with their corresponding faces along with different angles and variation is stored and trained over a period of time. All this can be done through various face detection techniques such as Viola-Jones and Convolutional Neural Network Cascade. Image of a class containing multiple faces is taken with the face being detected and recognized by comparing their extracted features in the database, thereby making the attendance of the student. The main objective of this project is to design and implement a large scale attendance system of educational institute or organizations. In this project we plan to design and implement a software application that allows college to detect and recognize upon database history. Further we try to scale our application to achieve high throughput and scalability.

II. Related theory

The proposed system aims at automating and simplifying attendance management system by using face recognition. The objective of the project is to design and implement a large scale attendance system of educational institutes or organization. the plan is to design. and implement a software application that allows the college to detect and recognize upon database history.

The system consists of a camera that captures the images of the classroom and sends it to the image enhancement module. After

enhancement the image comes in the Face Detection and Recognition modules and then the attendance is marked on the database server. At the time of enrolment templates of face images of individual students are stored in the Face database ,Here all the faces are detected from the input image and the algorithm compares them one by one with the face database. If any face is recognized the attendance is marked on the server from where anyone can access and use it for different purposes.

Algorithms for face detection

- **Viola-Jones face Detector**

In Viola-Jones system a simple feature is used, with relation to the feature sets.Viola and Jones make note that the fact the choice of features instead of a statistical pixel based system is important due to the benefit of ad-hoc domain encoding. In the case of face detection this is particularly important. Features can be used to represent both the statistically close facial information and sparsely related background data in a sample image.

In its simplest form the features can be thought of as pixel intensity set evaluations. This is where the sum of the luminance of the pixels in the while region of the feature is subtracted from the sum of the luminance in the remaining gray section. This difference value is used as the feature value, and can be combined to form a weak hypothesis on regions of the image. Within the implementation four of the Haar-like features are chosen, the first with horizontal division, the second a vertical, the third containing two vertical divisions and the last containing both the horizontal and vertical division. The features are called Haar-like because of their resemblance to Haar-basis functions.

- **Neural Network-Based Face Detection**

In neural network-based face detection system. A retinally connected neural network examines small windows of an image and decides whether each window contains a face. The system arbitrates between multiple networks to improve performance over a single network

It simplifies the difficult task of manually selecting nonface training examples, which must be chosen to span the entire space of nonface images. Simple heuristics, such as using the fact that faces rarely overlap in images, can further improve the accuracy.

- **Boxflow algorithm**

The algorithms aims to further enhance the discriminative capability of face detectors in the deployment domain through unsupervised detector adaptation. BoxFlow – a new unsupervised detector adaptation method that can effectively adapt a face detector pre-trained static images to videos. BoxFlow unsupervisedly adapts face detectors through fully exploiting the motion contexts across video frames. In particular, BoxFlow introduces three novel components: Generalized heat map representation of face locations with augmented shape flexibility; Motion based temporal contextual regularization among adjacent frames for unsupervised face detection refinement.

A self-paced learning strategy that adapts face detectors from easy data samples to challenging ones progressively. The viola-jones face detector algorithm which is the best one is only giving 60% precision. The experiment conducted can be estimated on the basis of the values of precision and recall hence we are using convolutional neural network to improve the efficiency by 10%. And it will further make the system more efficient.

III. Motivation

Attendance is very important for a student, a single absent adds a big difference in the performance of the college. Mostly students of high school are prone to absences, it is because of the laziness to attend the class, some students prefer going to computer shops playing games rather than entering the class and some student cannot refuse the influence of a friend inviting to go with them during class period. Some of this reasons are not reported to the parents or guardians because the way of informing them is the traditional way which is inviting the parents through students and the parent need come to college and communicate with the teacher about the absenteeism of the student. the proposed system takes care of these flaws present in the manual based attendance system.

IV. Problem definition

Most of the time of lecture or laboratory starts the lecturer or teaching assistant delays the lecture to record students attendance. This is a lengthy process and takes lots of time and effort, especially if it is a lecture with huge number of students. It also causes a lot of disturbance and interruption when an exam is held. In this project, we intend to find the best model and set of features to design and implement a large scale attendance system of educational institutes using face recognition. In these project we plan to design and implement a software application that allows college to detect and recognize upon database history.

V. Methodology

The Proposed methodology is illustrated in Figure. It contains following steps :- Data Extraction (Image input of class)**,**Data Preprocessing (Face detection)**,** Data Integration and Transformation(Recognition and Comparison)**,** Attendance Database

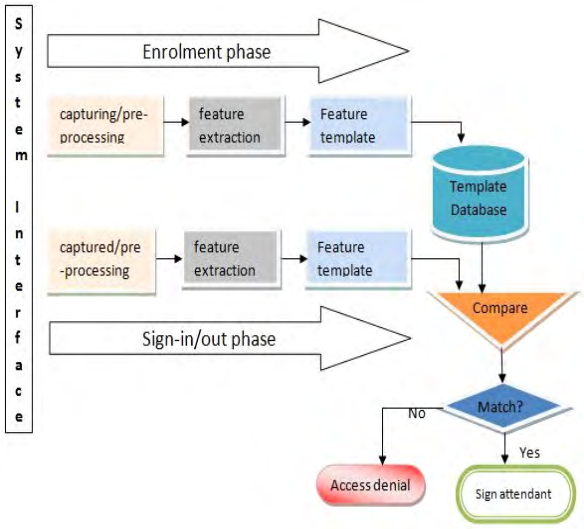


Figure 1 : Implementation Method Adaptive Automated Attendance Analysis for Education System.

- **System Architecture**

Phases of our system are:

- Image input of class
- Face Detection using Viola Jones
- Face Recognition using LDA
- Feature Comparison
- Face recognition using CNN
- Attendance Database
- Monthly Report

1. Image input of class
Image input of class will be taken from camera. This image will consist of all the students present in class for that particular lecture.

2. Face Detection using Viola Jones algorithm
This algorithm is used to detect the face of each student. This algorithm consists of four phases integral image, Haar- like Features, AdaBoost, Cascading Classifier. In integral image we assign each pixel a value. And these values are converted by summing up all the pixel values present above and at the left side. Haar-like features consists of special pattern which is compared with image pixels and accordingly 1 or 0 is assigned. In AdaBoost only some necessary features are selected removing redundant ones. In Cascading classifiers weak classifiers are cascaded to make a strong classifier. The detailed method is explained in later section.

3. Face Recognition using Linear Discriminant Analysis
In this algorithm image is divided into several parts and on each part LDA is applied. Initially 3 X 3 top left corner window is taken and each pixel is converted into equivalent grey level. After this centre pixel value is compared with all the neighborhood values and if centre pixel value is greater than surrounding pixel then it is 1 else 0.After this, the 8-bit code is converted into decimal equivalent and this procedure is repeated for the whole image. After this all the decimal values are plotted on histogram.

4. Feature Comparison
In this phase the calculated features are compared with stored features. In this for testing we have used SQL database. This includes variation in gestures, luminosity, expression, surrounding, race, ethnicity, age, camera quality, color saturation, focus, and other parameters.

5. Face Recognition Using Convolutional Neural Network
If the results are still not matched we then use the Convolutional neural networks. Convolutional neural networks (CNNs) have very powerful discriminative capability, while maintaining high performance. The proposed CNN cascade operates at multiple resolutions, quickly rejects the background regions in the fast low resolution stages, and carefully evaluates a small number of challenging candidates in the last high resolution stage

6. Attendance Database
If features match in comparison phase, then attendance for particular student is registered in the database. This database is stored in the backend of the system and the results drawn are all entered in the same.

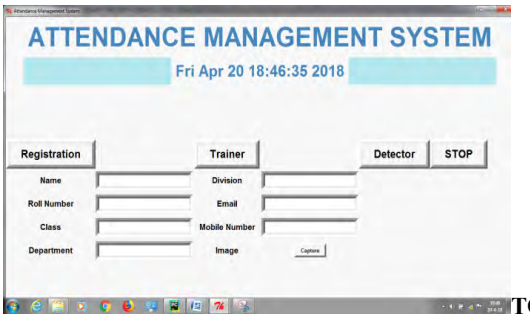
7. Monthly Report
At the end of each month monthly report will be generated and will be mailed to HOD of department. The report can be generated as per the institute norms. This reduces or finishes off the entire work of the faculties of the institute. Any alterations required to be done in the final report can be done by the respective authorities.
In this way, the proposed system does the work of attendance marking using face detection and face recognition respectively and generates the final report as per the required institute norms

VII. Results

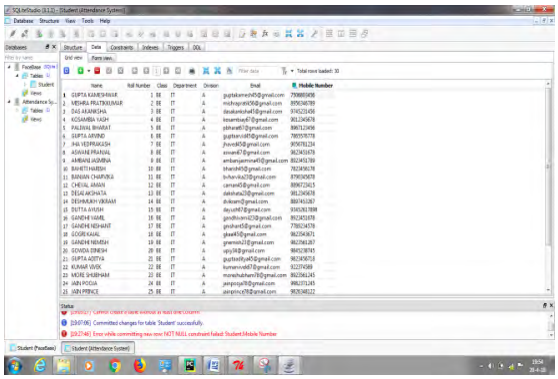
- **Outputs**

The system is able to detect faces through a high definition camera, the camera captures several faces of students within the class further these images are recognized by comparing them with the existing features in database thereby conducting the attendance. The system detects and recognizes up to 80% of faces of students present in the classroom thereby recording the attendance

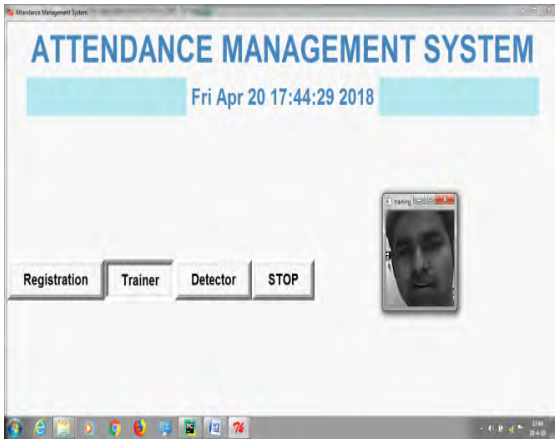
1. DATASET CREA



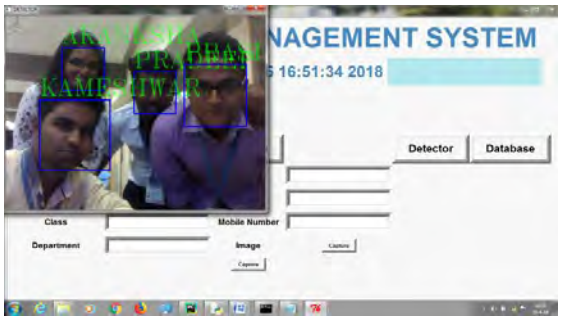
TOR



2. TRAINING DATASET



3. DETECTOR



4. IMPROVED EFFICENCY OF THE SYSTEM



- **Percentage improvement in results.**

Existing algorithms for multiple object recognition like the Support vector machine, Trace transform, Independent component analysis and Naïve Bayes classification have multiple drawbacks such as extensive memory is required, lower speed, order of components is difficult to determine and loss of accuracy. In the proposed system deep learning algorithm such as convolutional neural network cascade is used for face detection and Linear Discriminant Analysis is used for face recognition. This system is capable of detecting face in various angles, lighting conditions and has the ability to detect face even if there is an occlusion in the face up to some extent.

Previous system could achieve accuracy up to 70% whereas the current system increases its accuracy up to 80%.

VIII. Future Scope

A. The project can be enhanced for processing 3D images.

B. Authentication can be embedded into some kind of web application which will be an added advantage for providing the login for the website.

C. The project could be further enhanced and developed on an android platform so that the students and even their parents could keep a track of their attendance.

D. It can also be used in several multinational organizations, offices and other giant government organizations for attendance as well as security purposes.

E. For security reasons, we can use detection and recognition system. To identify culprits on railway station, bus stations and other public places we can use this system. It will be a helping hand to the police.

IX. Conclusion

In order to reduce the faculty effort and to manage the time effectively the authors proposed automated attendance system base on face recognition in schools/colleges. In this system we have implemented an attendance system for a lecture, section or laboratory by which lecturer or teaching assistant an record student's attendance. It saves time and effort, especially if it is a lecture with huge number of students. Most of the existing systems are time consuming and require for a semi manual work from the teacher or students. This approach aims to solve the issues by integrating face recognition in the process. Even though this system still lacks the ability to identify each student present on class, there is still much more room for improvement. Since we implement a modular approach we can improve different modules until we reach an acceptable detection and identification rate. Another issue that has to be taken in consideration in the future is a method to ensure users privacy. Whenever an image is stored on our servers, it must be impossible for a person to use that image. computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

X. References

[1] B.-C. Chen, Y.-Y. Chen, Y.-H. Kuo, T. D. Ngo, D.-D. Le, S. Satoh, and W. H. Hsu. Scalable face track retrieval in video archives using bag-of-faces sparse representation. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.

[2] F. Comaschi, S. Stuijk, T. Basten, and H. Corporaal. Online multiface detection and tracking using detector confidence and structured SVMs. In *IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS)*, 2015.

[3] S. Xiao, S. Yan, and A. A. Kassim, "Facial landmark detection via progressive initialization," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 33–40.

[4] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp.5325–5334. S. Ren, X. Cao, Y. Wei, and J. Sun, "Face alignment at 3000 fps via regressing local binary features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp.1685–1692.

[5] S. Ren, X. Cao, Y. Wei, and J. Sun, "Face alignment at 3000 fps via regressing local binary features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp.1685–1692.

Android Based Car Tracking System

Karan Pawar, Aditya Gupta, Dr. Radhika Kotecha (Professor)
Department of Information & Technology
Thakur College of Engineering & Technology
Mumbai, India
karanpawar163@gmail.com, Kuku.g.r@gmail.com, Kotecha.radhika7@gmail.com

Abstract-In this paper, we present a low-cost, simplified yet versatile and optimized vehicle diagnostic system that is compatible with all vehicles manufactured after 1996 (OBDII compliant).The graphical user-interface (display and command) is Andriod-based and utilizes the popular personal area network (PAN) communications standard Bluetooth to facilitate extraction and relaying of readings, diagnostic trouble codes (DTC), and commands. The paper targets smart phones as the computing device for the obvious growth and demand on such mobile devices, in addition to reducing the overall system cost by utilizing the built-in functionalities that is integrated in such compact devices. When installed properly, such devices can be a low cost alternative to integrated navigation systems. Moreover, our choice of Android as our operating system platform is consistent with current market trends and shares and user acceptance of such platform. Through the approach proposed in this paper, new business opportunities that provides customer-oriented vehicle diagnosis and remote maintenance can are offered. Also, extraction of fault detection and sensor output signals from the engine control unit (ECU) according to the OBD-II standard, vital vehicle information can be displayed to users, and/or communicated to remote maintenance departments via cellular networks. Hence, if a correct diagnoses (or prediction) of a defect is available, instructionsand commands can be sent to the driver on how to proceed in that situation. The results shown in the paper prove the same.

Keywords
Optimization, Defect Diagnosis, Decision Making, Android

I. INTRODUCTION

While your vehicle is being driven on to the road, anything unfortunate can happen to the vehicle; it may meet some unwanted accident or may be theft. So, it is very necessary to track your vehicle. Keep i.e. track on the status and exact location of vehicle carrying valuable goods at anytime and anywhere is essential for both personal as well as business purpose. You can track your vehicle by calling to the driver over mobile phone but this not convenient as your driver might give you false information. So, at any time you want to be 100% sure about the location and status of the vehicle. Here, various vehicle-tracking technologies based on the Global Positioning System (GPS) or radio frequency medium can track your vehicle location in real-time and provide better control over your assets. Built-in automobile navigation systems are being equipped with various multimedia capabilities to satisfy the customer expectations of modern functions and features and offer diverse services in the current greatly competitive market state. This reflects on the overall system cost and limits the advantages to high-end vehicle models. New business opportunities that provide customer-oriented vehicle diagnosis and remote maintenance can be offered. Further, extraction of fault detection and sensor output signals from the engine control unit (ECU) according to the OBD-II standard, vital vehicle information can be displayed to users, or/and communicated to remote maintenance departments via cellular networks where relatively newer cellular access technologies offer future solutions endless choices for high-end applications Hence, if a correct diagnoses (or prediction) of a defect is available, instructions and commands can be sent to the driver on how to proceed in that situation. Also, if the system has knowledge of the location, in addition to an error category, the system can direct the vehicle to the closest vehicle service location in the vicinity.

II. LITERATURE SURVEY

Using GPS and CDMA we can track real time location and radio technology using microwave.Components use to track the location are expensive and coverage area might stutter because of 2G network [1].This system controlled by an RFID module to switch on and off,GSM module will be used to send the location and GPS will to owner mobile phone.We can extend the use of this system to the car dispatching management field [2]. GPRS/GSM modules are used to find the real time location of the vehicle which is monitored on smartphone using Android UI . Power supply to the system is not constant which could go wrong if the battery dies[3]. Functionality of OBD 2 port increase in this project as it shows vehicle status using from ECU showing DTC (Diagnostic trouble codes.).The function is to send the error codes to the nearest Service Station. User can only track the status of the vehicle using Bluetooth but can't track it remotely [4].

A.EXISTING SYSTEM

In all the existing systems, we can analyze that vehicle

tracking integrates the GPS tracking system9 with existing vehicle alarm or provide alarm features when someone is tampering with vehicle. Before the vehicle is driven away, security threat is detected and the vehicle is able to be tracked over the internet which is done by utilizing Global Positioning Satellites. Through which, data such as Global Position, including latitude and longitude are transmitted over the Cellular network10.This information transmitted from the tracking device is stored in your private confidential account or sent over the wireless network. The data is cross referred on a street level map for viewing.

B.DRAWBACKS OF EXISTING SYSTEM

The crucial disadvantage of the existing system is that the system does not provide street wise address and speed of the vehicle and engine is no way controlled by the existing systems, thus exposing the vulnerability of a system that provides only tracking.

C.FINDING

A detailed survey was conducted in this concern. The following are the key finding:

a)	Using GPS and CDMA we can track real time location and radio technology using microwave
b)	This system controlled by an RFID module to switch on and off, GSM module will be used to send the location and GPS will to owner mobile phone.
c)	GPRS/GSM modules are used to find the real time location of the vehicle which is monitored on smartphone using Android UI.
d)	Functionality of OBD 2 port increase in this project as it shows vehicle status using from ECU showing DTC (Diagnostic trouble codes.).The function is to send the error codes to the nearest Service Station
e)	The commissioned a OBD-II linked cellular information-based system driving assistance. The information is run through a web based basic variable fusion algorithm which chelates events related to safety whilst interpreting and presents the results.
f)	The information gained from the status of the road and vehicle is sent to the central cellular server and together with the GPS data which allows visual data to be acquired.
g)	As developed an application which integrates the OBD-II connection to an external network. The driver is informed of the relevant variants to the OBD-II via the ECU in real-time

III.PROPOSED SYSTEM

The network stutter faced in Ref [1] can be solved by replacing CDMA module with the GSM module to access the 4g network, RFID tag Ref [2] will make it complex as the car key sensors has to be changed hence the module will be connected to battery directly or OBD 2 port to power the whole device which also provides the solution in Ref [3].To track the location in Ref[4], a user friendly interface can be made on Android OS which could be accessed by any android running smartphones. The planning process regarding the project is one of the main obstacle as it deals with many different configurations grade of the hardware components and compatible software used for the project. Deciding to make it bulkier or small would create the main impact on the project. The IC components changes from time to time and the compatible issue would grow bigger as sometimes the software is not updated with the new gen hardware or vice versa. Selecting and designing the right compatible hardware. The options between OBD 2 ELC327 or OBD 2 adapter to USB 2.0 which can change the whole prospect of the project as ELC327 comes ship with Bluetooth 2.0 and interface to check the car error codes from ECU but it might not be compatible with the programmable ArduinoAtmega 238 UNO R3. The problem revolves around the software or programming languages used in the project. Arduino shied will used to program with GSM module but android based programs are written in java which might create a huge compatible issue. And the main issue might be faced while setting up the GPRS/GSM module with the server using http TCP/IP protocol. 7 Implementing on the cars before 1996 production would not be possible as the OBD 2 port was not introduced. The battery consumption for the Bluetooth module is higher which can drain the battery when car is in idle or off state.

III. MOTIVATION

This application makes use of the Bluetooth & GPS features of mobile phone as a solution for Vehicle tracking system. It is the technology used to determine the location of a vehicle using different methods like GPS and other radio navigation systems operating through satellites and ground based stations. Even data can be stored and downloaded to a computer from the GPS unit 3 at a base station and that can later be used for analysis. This system is an important tool for tracking each vehicle at a given period of time and now it is becoming increasingly popular for people having expensive cars and hence as a theft prevention and retrieval device. No doubt, Vehicle tracking system whether it is GPS based or any other wireless medium has brought one of the most important technological advances in today's communication field. Now one doesn't have to leave a place to know where a particular vehicle is at a given period of time. The automatic

vehicle locating system only with the help of a tinny electronic device and tracking software can detect the real-time location of a vehicle by using the conventional cell phone network and Internet.

V. PROPOSED METHODOLOGY

Phase 1 (Design): The main goal for the design is to make the project much compact including all the necessary hardware components. The UI for operating the app is user friendly and simple as it instructs.
Phase 2 (Hardware components testing): Compatible IC Components such as GPS/GSM module, OBD 2 bluetooth and OBD 2 adapter to USB type A. Operating the OBD using IS0-9141-2 protocol and TTL/CMOS based ICs.
Phase 3 (Software components development): Theuno r3 arduino which is programmable in python language which will interact with GPS/GSM module following sending the data to server which will display the output on the smartphone using Android UI. Android apps are specifically made for programming in java.
Phase 4(Implementing and testing): Using the OBD 2 port to power the system continuously which will trigger if the car is not started and changes its initial position by sending the notification to user's cell phone in the created app by using TCP/IP protocol and https port 80 send by the gsm module to database and server.

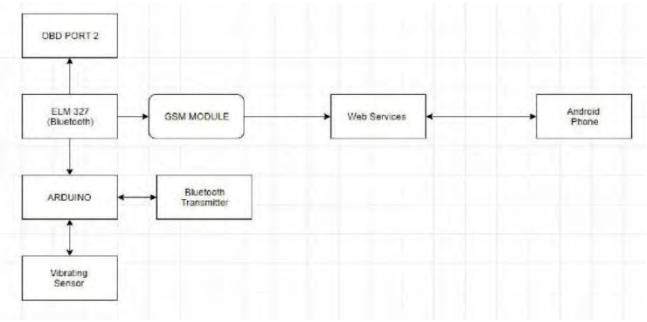


Fig.1 Proposed Framework for Connection from OBD-II to Mobile

Tools used

1. GSM module.
2. GPS module.
3. ELC327 OBD 2.
4. Arduino UNO R3.
5. SIM 808 Arduino shield (elements).

The features of the project are:

a) Real time monitoring of vehicle sensors
b) Rules concept for checking state of sensors
c) Digital Dashboard
d) Acceleration Test
e) Read Diagnostic Trouble Codes (DTCs)
f) Clear DTCs and turn off the MIL (engine check light)
g) Retrieval of car details such as OBD standard and Vehicle ID (VIN)

VI. RESULT



Fig.3 Output 2

An On Board Diagnostics (OBD) port interface using vehicle ECU interfacing unit STN1110 is designed through OBD II Data Link Connector (DLC). The GPS location of the vehicle is obtained through this interface. The remote server application diagnoses the fault by querying the data which is made through OBD commands to the vehicle. The current system worked using 2G based GPRS network but it can be updated by 3G network service. The interfacing module is developed to read OBD data from the vehicle can be reduced in size and mounted in vehicle. Remote server application can be made to run on mobile. The system can be made to correct the fault and guide the driver. By using the sim card for tracing the vehicle we can sent the message and live location message.

VII. CONCLUSION

We implemented a universal integrated system which is composed of a combination of a low-cost hardware unit

and user-friendly Android-based mobile application software utilized to create an on-board vehicle diagnostic system. The mobile application software will interact with the hardware interface unit wirelessly via Bluetooth to acquire desired vehicle parameters from the ECU of the vehicle. These readings will be displayed locally to the user then can be sent to a remote maintenance server as HTTP packets via a cellular internet connection. The packets received will be tabulated in the server, then made use of by the maintenance department which holds the server.

VII. ACKNOWLEDGMENT

The people we would to thanks are many for their help in presenting this project seminar. we specially want to thanks our Principal Dr.B.K.Mishra, our HOD-IT Dr. Rajesh Bansode and our project guide Dr. RadhikaKotecha whose advise have inspired and shaped much of the content of this project also guiding us on this project work and spending hours working to improve the content of this project. Also we want to thanks to all the staff members who help us to improve.

VII. REFERENCES

[1] Dhumal, A. Naikoji, Y. Patwa, M. Shilimkar, M. Nighot, "*Car Tracking System Using Android Application*", vol. 4,no. 4, UARCET, IEEE,April 2014.

[2] L. Zhingang,Z. Angi,L. Shaojun,."*Vehicle anti-theft Tracking system based on IoT*".No 48-51,Qinhuangdao,China,IEEE, April 2013.

[3] L. SeokJu, T. Girma, K. Jaerock."*Design and Implementation of Vehicle Tracking System Using GPS/GSM/GPRS Technology and Smartphone Application*.",Electrical and computer Engineering,(Flint,MI,USA), {iee7704,gtewolde,jkwon}@kettering.edu,IEEE ,March 2014.

[4] T. Ashraf, S. Ahmad, J. Fouad, Q. Waleed.*"Android-Based Universal Vehicle Diagnostic and Tracking System*".Communication Engineering Department, Princess sumaya University forTechnology,Amman,Jordan,aat@ieee.org,IEEE,June 2012.

[5] L. SeokJu, T. Girma, K. Jaerock."*Design and Implementation of Vehicle Tracking System Using GPS/GSM/GPRS Technology and Smartphone Application*." ,Electrical and computer Engineering,(Flint,MI,USA),{iee7704,gtewolde,jkwon}@kettering.edu,IEEE ,March 2014.

[6] Enriquez, D. J., et al. "*CANOPNR: CANOBD programmable-expandable network kenabled reader for real-time tracking of slippery road conditions using vehicular parameters.*" Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on. IEEE, 2012

[7] PankajVerma, J.S Bhatia "*DESIGN AND DEVELOPMENT OF GPS-GSM BASED Tracking System Paper 2013*". International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.3, No.3, June 2013

[8] Abid khan, Ravi Mishra "*GSM Based Tracking System Paper 2012*". International Journal of Engineering Trends and Technology- Volume3Issue2- 2012ISSN: 2231-5381.

[9] Mohammad A. Al-Khedher "*Hybrid GPS-GSM Localization of Automobile Tracking System 2011 Paper*". International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 6, Dec 2011.

[10]AshishShrivastava , Noopur Sharma, RadhikaArora, ShwetaDikshit"*GSM Enhanced GPS Based Vehicle Tracking System*" 2nd National Conference in Intelligent Computing & Communication organized by Dept. of IT, GCET Greater Noida, INDIA ISBN: 9788175157538

[11]BaburaoKodavati, V.K.Raju, S.SrinivasaRao, A.V.Prabu, T.AppaRao, .Y.V.Narayana "*GSM AND GPS BASED VEHICLE LOCATION AND TRACKING SYSTEM*" International Journal of Engineering Research and Applications(IJERA) ISSN: 2248-9622 Vol. 1, Issue 3, pp.616-625.

[12]Yang, Yalian, et al. "*Research and Development of Hybrid Electric Vehicles CAN-Bus Data Monitor and Diagnostic System through OBD-II and Android-Based Smartphones*." Advances in Mechanical Engineering 5 (2013): 741240.

[13]Zaldivar, J., et al. (2011). *Providing accident detection in vehicular networks through OBD-II devices and Android-based smartphones. Local Computer Networks (LCN*), 2011 IEEE 36th Conference on, IEEE.

[14]Ceuca, Emilian, et al. "*Embedded system for remote monitoring of OBD bus."*Electronics Technology (ISSE), 2013 36th International Spring Seminar on.IEEE, 2013.

[15]Jhou, Jheng-Syu, et al. "*The Implementation of OBD-II Vehicle Diagnosis System Integrated with Cloud Computation Technology.*" Robot, Vision and Signal Processing (RVSP), 2013 Second International Conference on.IEEE, 2013.

[16]Baek, Sung-hyun, and Jong-Wook Jang."*Implementation of integrated OBD-II connector with external network.*" Information Systems 50 (2015): 69-75

[17]Ruta, Michele, et al. "*A mobile knowledgebased system for on-board diagnostics and car driving assistance*." UBICOMM 2010, The Fourth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies. 2010

[18] Čabala, M. and J. Gamec (2012). "*Wireless real-time vehicle monitoring based on android mobile device.*" ActaElectrotechnicaetInformatica 12(4): 7-11.

[19] Chen, Y., et al. (2009). *Design and implementation of multi-source vehicular information monitoring system in real time.Automation and Logistics*, 2009.ICAL'09.IEEE International Conference on, IEEE.

[20] Tahat, A., et al. (2012). *Android-based universal vehicle diagnostic and tracking system.Consumer Electronics* (ISCE), 2012 IEEE 16th International Symposium on, IEEE.

Intelligent Tourist System

Rohan Shetty
Department of Information Technology, Thakur College of Engineering & Technology, Mumbai, Maharashtra shetty.rohan83@gmail.com

Priyanka Verma
Department of Information Technology, Thakur College of Engineering & Technology, Mumbai, Maharashtra priyanka96.tcet@gmail.com

Dharti Singh
Department of Information Technology, Thakur College of Engineering & Technology, Mumbai, Maharashtra dhartisingh13@gmail.com

Radhika Kotecha
Department of Information Technology, Thakur College of Engineering & Technology,Mumbai, Maharashtra kotecha.radhika7@gmail.com

ABSTRACT - Tourism is one of the biggest sources of revenue of any country. But people find it difficult to plan their trip in foreign countries or even other states of their own country. In this technological era, people prefer to plan their trips themselves online rather than paying a huge amount to tour operators. The paper proposes a solution using of machine learning techniques and recommendation systems to simplify the trip planning process for the end user. Various methods that utilize finding patterns in given user's past behavior or to find similarities and correlations between known points of interest are included. These smart systems could be used to give personally tailored experience and recommendations for user. These recommendations could be dependent on their schedule, preferences and budget. Machine learning would also automate the travel planning process and make the finding of the new interesting places easier for wider audiences.
Keywords: Intelligent information, Travel Guide, Decision Support, Bayesian algorithm, Information filtering

INTRODUCTION

Normally tourists have to search a lot of websites to get information and routes about the destination and this takes a lot of time, thus making it frustrating. Many a times tourist don't have adequate information of the places they're visiting, thus having a tough time finding appropriate destination as per their taste and budget. Existing system will take a lot of time. The traveller might have to limit themselves into only few places to visit if they don't have enough time.

The proposed system is helpful for people that want to visit a city without having much information about it. The main idea is to design a system that is able to find a route using user criteria. It should also be able to estimate time needed to travel from one object to the next and if it is possible, advise which bus line or other public means of transport may be used. It also tracks the user's location using GPS.To build an application, that would-be context aware and assistive for the user, the concept of machine learning and predictive algorithms are needed. The application would update its recommendations, when it detects that the user is in foreign country and connected to Wi-Fi-network. The current location and date would be used as a context, and automatically suggested activities would be given, based on the past interests. More the user travels, more attractive the recommendations should be for the user. The chosen machine-learning algorithm should recognize patterns in users' preferences and use these patterns, as base for the suggested points of interest.

LITERATURE SURVEY

In existing system, the traveller has to search a lot of websites in order to get the knowledge of the routes or information about the particular place. Existing system will take a lot of time. The traveller might have to limit themselves into only few places to visit if they don't have enough time. This system has no functionality to find a route using user criteria. We are unable to estimate the time needed to travel from one place to another and know which bus line or public transport to use in order to reach our destination. This system is not adapted to track the user's current location using Global positioning system to give better and accurate results. Features such as online forum which makes the system more user friendly and gives users a chance to communicate with each other for better feedback is absent. Also, personalization to cater the specific needs of the user is absent thus making it not very user friendly. Table I describes the existing systems.

Table I: Summary of Existing Systems

Ref. No.	Approach	Proposed Solution
[1]	An application, that would-be context aware and assistive for the user, the concept of machine learning and predictive	Categorical data found out to be difficult one to deal with Scientific Python

	algorithms are needed. The application would update its recommendations, when it detects that the user is in foreign country and connected to Wi-Fi-network.	libraries, and after pre-processing those values to numerical representatives, so instead of using numerical values we use discrete values which is much simpler to be dealt with.
[2]	Recommender systems (RS) are used to help users find new items or services, such as transportation or even people, based on information about the user, or the recommended item. These systems also play an important role in decision-making, helping users to maximize profits or minimize risks.	RS development lacks studies analysing early stages, such as requirements and design, and late stages, such as maintenance. Open questions in these stages must be investigated to improve the knowledge about the field.
[3]	An Intelligent Tourist AttractionsDecision Support System (ITAS) that has an intuitive and interactive user interface, combines CF-based data mining with the Google Maps API, and uses an ROC curve to evaluate the accuracy of the Bayesian network.	Continuation of this study can, with the same methods, extend to include information regarding accommodation to provide more comprehensive travel information.
[5]	Although there is numerous information provided on the web, the user gets puzzled in finding accurate information. In order to solve these web problems, the concept of semantic web comes into existence to have communication between human and computer. In this paper, we propose intelligent recommendation system	Since this paper takes the help of artificial intelligence, ratings given to places won't be recommended and so the same results will keep coming thus making the results cold.
[4]	The system provides information about the locations in the form of notifications and also provides the shortest path to reach the destination location. The application also provides information about the weather conditions of a location the tourist wants to visit and recommends him/her to take the appropriate accessories according to the weather	This system does not use machine learning. Thus because of this new data is not updated, leading to the same stale results coming up and filtering is the same. Errors are also called due to multipath fading, which occurs when a signal bounces off a building or terrain before reaching the receiver's antenna, also can reduce accuracy.
[6]	Simplified service architecture as well as the employed technologies used for the development of the different system components. The system is mainly composed by a representational state transfer (REST) service with a data access layer which exposes a set of endpoints in order to be accessed by the client applications. Client applications are available as Web-basedand Mobile	Cold start or first-rater problem - which means that a new item added to the system cannot be recommended to any user until that item is rated. Sparsity - users tend to rate only a few items over the entire set, resulting in a very sparse matrix, i.e., a matrix with a high percentage of empty cells. This compromises the ability of the systems at successfully locating neighbours, yielding poor recommendations . Good performances require that a very large number of users have rated a very large number of items. Grey sheep problem - individuals whose opinions are unusual, rarely receive accurate recommendations

PROBLEM DESCRIPTION

This system is helpful for people that want to visit a city without having much information about it. The main idea is to design a system that is able to find a route using user criteria. It should also be able to estimate time needed to travel from one object to the next and if it is possible, advise which bus line or other public means of transport may be used. It also tracks the user's location using GPS.

The system will work in 4 phases, as follows:

Phase 1(Analysis and Planning): -
We have taken reference of 6 literature survey papers that contain knowledge including substantive findings, as well as theoretical and methodological contributions similar to our topic. The reviews in these papers are secondary sources, and do not report new or original experimental work. It is associated with academic-oriented literature, and are found in academic journals. We have used the key findings of these papers that add value to our project and elaborated them.

Phase 2(Design and Implementation): -
The intelligent tree algorithm covers both classification and regression. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions. To make sure that the front end (login forms and the questionnaire) are connected to the databases and function properly. Python will be used for coding the frontend

Phase 3(Testing): -
To check if the information filtering process works properly and try to use it with the GPS system to make sure suggestions favouring the user's interests are given.

Phase 4(Deployment): -
Implementation of the filtering process and making sure the user's forum functions properly.

Following functionalities are included in the project:

Login and register: User enters his/her credentials and is able to use the services provided by our website.
Interactive questionnaire: Series of questions asked to find out where the user's interest lies.
Personalization: Caters to the specific needs of the user and makes it user friendly
Location tracking- When the user logs in to our site, his current location is automatically tracked using GPS enabled system, based on which nearby places are suggested to him.
Information filtering: With the help of machine learning, places liked and suggested by the previous users stored in the online database will be generated for the next user based on similar interestOnline forum: Not only do they help in gathering useful feedback but helps the user in shortlisting the places he would like visiting.

PROPOSED METHODOLOGY

The proposed methodology is illustrated in Fig. 1. It contains following steps:
Data Extraction
Data Preprocessing
Data Integration and Transformation
Feature Selection
Classification

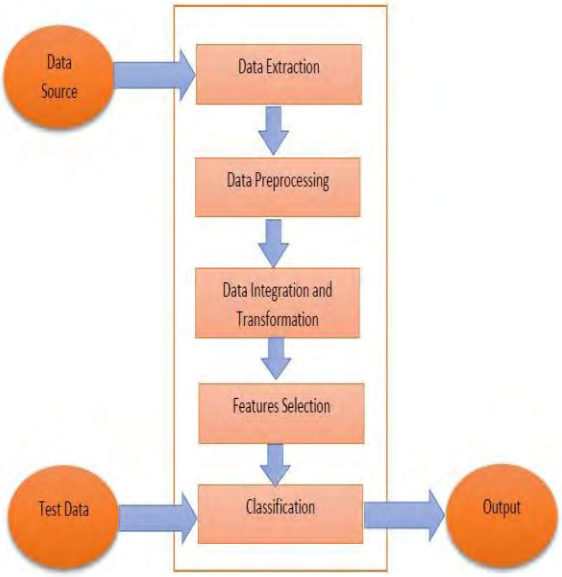


Fig. 1: Proposed Methodology

We need to extract important characteristics form user's data, pre-process them, integrate and transform all data into a unified format, select appropriate features and them perform classification for the user. Following classification techniques can be used:

Random Forest:
In bagging (Bootstrap Aggregation), numerous replicates of the original dataset are created to reduce the variance in prediction. Random Forest use multiple decision trees which are built on separate sets of examples drawn from the dataset. In each tree, we can use a subset of all the features we have. By using more decision trees and averaging the result, the variance of the model can be greatly lowered.

Decision Tree: decision tree is commonly used in classification in machine learning, it is a tree-like graph or model to make decisions and get consequent results. Decision tree implicitly

perform variable screening or feature selection and the best feature of using it is easy to interpret and explain!

kNN:
K nearest neighbors is a non-parametric method used for classification and regression. The input normally consists of k closest training instances in the space. An object is classified to a class because it is most common among its k nearest neighbors. It is chosen because knn don't need make any assumption on the underlying data distribution, it is also a lazy algorithm and don't use training data to do any generalization.

Naïve Bayes:
Naive Bayes classifier is a probabilistic classifier based on Bayes' theorem. It depends on the assumption that some features are conditionally independent. It is super simple and can converge quicker than some other models, so it has a chance to achieve fast and easy process.

Following feasibility study has been conducted for our system:

Technical Feasibility – The application will operate on Web Browser. The system will be implemented using python, django python framework, and web technology as required. Proven and tested mature technologies to deliver the proposed solution. It works on all platforms which supports python.

Economic Feasibility– This system has functionality to find a route using user criteria. We are unable to estimate the time needed to travel from one place to another and know which bus line or public transport to use in order to reach our destination. This reduces the cost of a tourist guide and the effort needed to look for places in a new city.

Legal Feasibility-The project is developing under legal license of windows, python and other software's required. The data will not be used for some illegal purpose and will be used for making the news relevant to user as well as for publisher for making under certain constraint. Data will not be misused or misguided under any circumstance.

Operational Feasibility -An overview of the process portrayed that the proposed Machine Learning for Online Tourist System be beneficial for news publisher to monitor the growth and improvement in the future article content. Also to user to view and read the relevant content.

Social Feasibility- General users have to go to the place where the system is held. Because of that, the system affects the existing culture of the user. And it is a time waste. It can also be a reason to avoid this system. People do not have a method to compare old and new products. The system provides this facility to users.

Fig. 2 shows the process of training the model using training data and implementing the model for getting the testing data. By training the model it will be used for input testing the input data.

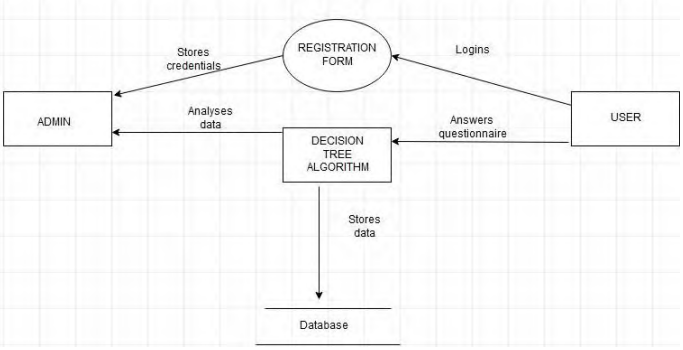


Fig. 2 Basic flow of information

Fig. 3 shows how the model in which the sample articles are being segregated into popular and unpopular articles using the model which is trained using the 49k sample.

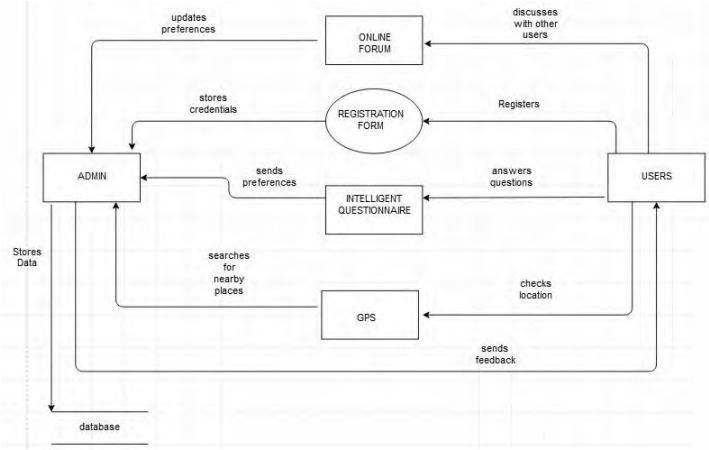
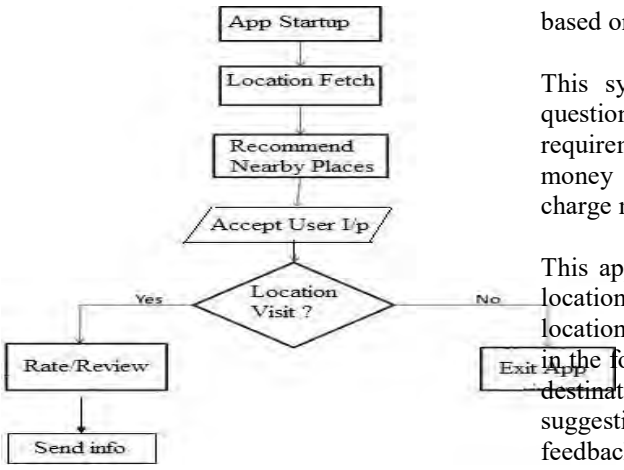


Fig. 3 Model of data flow

FLOW CHART
Fig. 4 shows the collection of raw data from online repository and training the data using different machine learning algorithm like random forest, decision tree, knn. etc.

In proposed system, the user will have to activate the application in their mobile or devices they are using. They can get every relevant information related to the place they wish to visit. The search mechanism and finding up of relevant path feature is the best, this intelligent tourist information application contains



Proposing statement of work, scope definition and scope boundary for planning the prototype from problem definition to decide what needs to be done and what not to be done
Finalizing each and every requirement with the customer
Phase 1.2: Analysis
Finalizing the synopsis, literature survey and feasibility study for proposed solution which will act as an input to the design phase
Phase 1.3: Design
Proposing network architecture for the website and it's modules to implement the surveyed case studies and synopsis
Creating an interactive questionnaire for analyzing user's query
Phase 1.4: Coding
Various modules will be created like
Creating the frontend using HTML,CSS and Python
Creating the Backend using SQL
Phase 2.1: Integration
Constructing the various modules of the proposed solution to integrate them into a prototype.
RESULTS
Machine Learning (ML) uses computers to simulate human learning and allows computers to identify and acquire knowledge from the real world, and improve performance on some tasks based on this new knowledge.

This system solves all problems of the client just by asking some questions. And it provides three best options that fit into users requirements along with the place details and facilities. It saves their money and time in finding and consulting a travel agency where they charge more.

This application provides recommendations to the tourists of the nearby locations worth visiting automatically without the need to search for the locations explicitly. The system provides information about the locations in the form of notifications and also provides the shortest path to reach the destination location. So basically, this application provides accurate suggestions according to user preferences and gives them other's users feedback and can even interact with them using the online forum. The results are shown in Fig. 5 to Fig. 8.

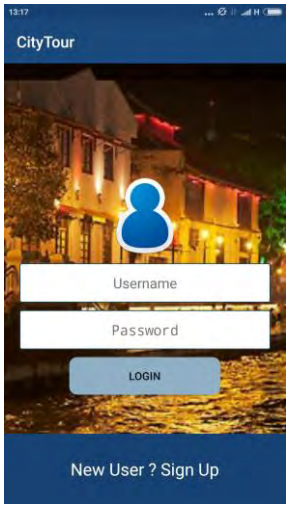


Fig. 5 : Login Page



Fig. 6: Categories to choose from



Fig.7: Questionnaire

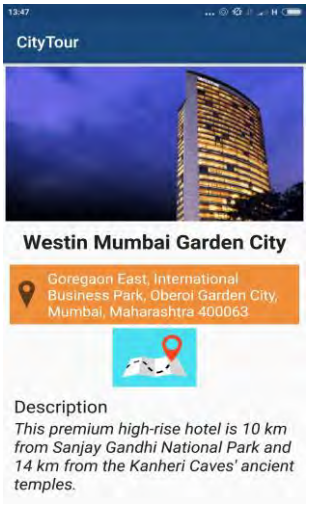


Fig. 8 Results

CONCLUSION

Machine Learning (ML) uses computers to simulate human learning and allows computers to identify and acquire knowledge from the real world, and improve performance on some tasks based on this new knowledge.

This system solves all problems of the client just by asking some questions. And it provides three best options that fit into users requirements along with the place details and facilities. It saves their money and time in finding and consulting a travel agency where they charge more.

This application provides recommendations to the tourists of the nearby locations worth visiting automatically without the need to search for the locations explicitly. The system provides information about the locations in the form of notifications and also provides the shortest path to reach the destination location. So basically, this application provides accurate suggestions according to user preferences and gives them other's users feedback and can even interact with them using the online forum

REFERENCES

[1] J. Kämppa, "Recommendation systems in the context of tourism", 2016.

[2] P. Portugal, D. Alencar, D. Cowan, "The use of machine learning algorithms in recommender systems: a systematic review", 2015.

[3] H. Fang-Ming, Y. Lin, T. Ho, "Design and implementation of an intelligent recommendation system for tourist attractions: The integration of EBM model, Bayesian network and Google Maps", Expert Systems with Applications, 2012.

[4] S. Walunj, P. Bhalere, Y. Bhaidkar: "Machine Learning Based Tourist Place Recommendation System", 2016.

[5] C. Choi, M. Cho, J. Choi, "Travel Ontology for Intelligent Recommender System", 2009.

A. Umanesta, A. Ferreria, N. Leita, "A Recommender system with social interaction", 2014.

Smart Warehouse – Reduction in postharvest losses

Aishwarya Nair
Department Of Information Technology
Thakur College of Engineering and Technology
Mumbai,India
aishun12@gmail.com

Anand Ranka
Department Of Information Technology
Thakur College of Engineering and Technology
Mumbai,India
anandranka25@gmail.com

Vikram Deshmukh
Department Of Information Technology
Thakur College Of Engineerin and Technology
Mumbai,India
vikramtcet97@gmail.com

Dr. Kamal Shah
Department Of Information Technology
Thakur College of Engineering and Technology Mumbai,India
kamal.shah@thakureducation.org

ABSTRACT

Current world population is expected to reach 10.5 billion by 2050 (UN March, 2013), which will further add to global food security concerns. Globally one-third of the food produced is wasted costing the world economy about $750 billion with the main reason being poor storage facilities of the grains. Food availability and accessibility can be increased by reduction of postharvest food losses which is a critical component of ensuring future global food security. Storage systems and smart warehouses are the need of the hour as various technologies have helped us increase the production but these goods tend to spoil due to lack of proper infrastructure, monitoring and management for their storage.

In the process of grain storage, temperature, moisture and gases have an influence directly on the quality of the Grain. Hence, there is a necessity to monitor the vital parameters continuously during storage and communicate the status to the manger in real time which becomes a challenge. The proposed solution is to make a traditional warehouse a smart warehouse with IOT enabled sensors and intelligent monitoring system to avoid quality losses. The project, ―Smart Warehouse – A probable solution for hunger‖, aims at implementing a system which will enable us to monitor the quality of the stored grains by reading the temperature, moisture and gas concentration values, recording the read sensor parameters and further take the necessary control actions when the threshold values are crossed such as notifying the manager and ringing the buzzer or in case of temperature rise turning on the fan.

The Experimental results shows that grain condition intelligent monitoring system designed in this paper has several good features such as real time online detection, easy acquisition and good site stability. This project will help cut loss of food grains due to spoilage by about 70% and help farmers in their development which will eventually lead to the development of the society as there is a dependency between agricultural growth and economic prosperity. A proper, secure storage management, retrieval management, data acquisition module and monitoring system provided by the proposed system will ultimately present a probable solution for hunger.

The temperature sensor LM35, soil moisture sensor and gas sensor are placed in the environment of the grain in close proximity. The sensors read the parameters and relay the analog signals to the Arduino Uno. The Arduino Uno is programmed to convert the analog signals into digital signals. The raspberry-pi is the main component which continuously pings the Arduino for the parameters and it is programmed to send out notifications via Pushbullet API (It is a messaging service which provides free services for a short period of time) using Wi-Fi.

Keywords –
Smart Warehouse, LIFO,Grain,Farmers,storage

I. Introduction

India is an agricultural country and agriculture contributes to about 16% of total GDP and 10% of total export. Thus, it is significant that proper advancements be made in this field but one phase that lags behind is storage of agricultural products

.Storage systems and smart warehouses are the need of the hour as various technologies have helped us increase the production but these goods tend to spoil due to lack of proper maintenance and management for their storage. Grain loss occurs by adverse environmental conditions and from the activities of insects and microorganisms. Temperature and moisture content of the stored grain environment are the most important factors that can influence stored-product. Maintaining optimum temperature, relative humidity and proper moisture content in the storage facility are the challenges faced in postharvest phaseThese activities release CO2 in Grain depot so CO2 concentration can be effectively used to monitor early detection of spoilage during storage. Hence, taking into consideration these factors, the proposed system makes use of temperature sensor , moisture sensor and gas sensor to read the grain conditions in real time and record the data and take the required control actions for the same, which helps us to improve the level of grains storage and reduce the grain losses during storage procedure and also reduce man power and labor intensity. The whole implementation is done by combining Embedded and IoT technology.

II. Overview

A new wave of connected devices, also known as The Internet of Things, is poised to grant warehouses a heightened level of visibility into every facet of the warehouse. IoT technologies offer the promise of facilitating real-time workforce interaction to elevate productivity, while adding a new level of precision. When it comes to the application of IoT technology, once cost prohibitive and now significantly more affordable, will play a critical role in modernizing warehouses.

III. Motivation

The real motivation of our project is to help the farmers and the Food Corporation of India (FCI) by providing them with a storage structure which fulfils their needs and demands on today's advanced Technologies. The goal is to minimize the loss of food grains and other agricultural products incurred by the farmer and provide them with a sense of security for their efforts.This system will play a vital role in the development of the storage of smart and automated warehouse.

IV. Literature Survey

A literature review surveys scholarly articles, books, dissertations, conference proceedings and other resources which are relevant to a particular issue, area of research, or theory and provides context for a dissertation by identifying past research. Research tells a story and the existing literature helps us identify where we are in the story currently. It is up to those writing a dissertation to continue that story with new research and new perspectives but they must first be familiar with the story before they can move forward.

1. The solution found is Distribution chain follows First-in-first-out and first-expired-first-out . The key findings are Distribution chain follows First-in-first-out and first-expired-first-out, Loss of quality taking into account the combined effects of temperature, O_2, CO_2 and relative humidity by equation :

— $\sum$ $_{def,i}K_{i.}$ And the Research gap found are High implementation complexity, therefore not feasible for implementation of bulk storage of grains.[1]

2. The solution found is Surface space required: 3 cars System can accommodate : more than 9 cars. Chain and sprocket mechanism for rotation to minimize vibration and noise. Operation : Key pressing. Thus, flexible and does not require care taker .
The key findings are Surface space required: 3 cars System can accommodate : more than 9 cars.
Chain and sprocket mechanism for rotation to minimize vibration and noise.
Operation : Key pressing. Thus, flexible and does not require care taker. and the Research Gap found is . It can be fully automated by integrating it with a panel Board..More secured by providing each platform a specific password, so that only whenever a particular password is typed the platform is retrieved.[2]

3. The solution found is .NODE 1: GPS mobile robot for weeding, spraying, keep vigilance
.NODE 2: Motion detector, light sensor, humidity sensor, temperature sensor, room heater, cooling fan altogether interfaced with AVR microcontroller. NODE 3 :Smart irrigation node

The Key findings are Setup for node1 consists of mobile robot with central server, GPS module, camera &other sensors interfaced with microcontroller which is interfaced with the raspberry pi. And the Research Gap found is Storage management , Retrieval management [3]

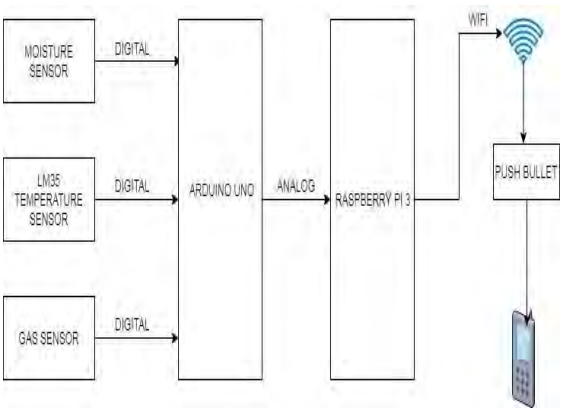
V. Problem definition (Phase wise)

The project focuses on making agricultural storage smart using automation and IOT technologies which can provide a probable solution for hunger. This we plan to achieve by accomplishing following phases:

PHASE 1: (Planning, Analysis, Design, Coding)
1. Planning – Conceptualizing and designing the architecture for proposed solution. Studying the feasibility and efficiency in the real market. Planning required to make use of soil moisture sensor, LM35 temperature sensor and MQ series CO concentration detector Sensor,Pushbullet. 2. Analysis – Critical study, analysis and review of feasibility for proposed solution. If any smoke/Gas detected OR Temperature goes beyond threshold OR humidity goes beyond threshold or CO2 concentration goes beyond Threshold activates controller to take control actions 3. Design – Hardware setup is designed using the components as described in planning phase and smart storage system with an integrated intelligent monitoring system will be developed. 4. Coding – The project was implemented using python code in raspberry pi OS.

PHASE 2: (Integration, Testing, Deployment)
1. Integration – The sensors were then integrated with the notification system and database 2.
Testing –Exhaustive testing using test cases to check the integration and fixing bugs for proposed solution. Perform alpha testing after completion of prototype. 3. Deployment – Give the completed prototype to a local grain warehouse

VI. Methodology
The temperature sensor LM35, soil moisture sensor and gas sensor are placed in the environment of the grain in close proximity. The sensors read the parameters and relay the analog signals to the Arduino Uno. The Arduino Uno is programmed to convert the analog signals into digital signals. The raspberry-pi is the main component which continuously pings the Arduino for the parameters and it is programmed to send out notifications via Pushbullet API (It is a messaging service which provides free services for a short period of time) using Wi-Fi.



Hardware:
1. Temperature sensor : Temperature sensors LM35 is chosen since LM35 provides reliable accurate surrounding temperature directly in °C according output voltage. These sensors are analog in nature which senses surrounding temperature produces output voltage in mV proportional to surrounding temperature, so it is easy to take accurate readings by connecting to 13 development boards. It can measure temperature ranges from -55°C to 150 °C.
The threshold value set for temperature sensor is 120

2. Soil moisture sensor : The soil moisture sensor is capacitive type. The sensor gives analog output of zero volts when there is 100% moisture and 5V for 0% moisture. The sensor measures the dielectric constant of the grain in order to find its Volumetric Water Content (VWC). It obtains VWC by measuring the dielectric constant of the media through the utilization of the frequency domain technology. The dielectric constant of water is much higher than that of air or soil minerals, the dielectric constant of soil is sensitive measure of volumetric water content. The sensor has a low power requirement and also very high resolution. Due to this the ability to make many measurements (i.e. hourly) over a long period of time with minimal battery usage.Threshold value for this sensor is 200 3.
MQ series gas sensor : Gas sensor MQ series gas sensors are used to detect the CO2 concentration level. These are Electrochemical in nature having gas detector that can measures the concentration of a targeted gas by oxidization or reduction of the targeted gas at an electrode and measuring the resulting current. The various gas sensor modules are used now a days to detect different types of gases on such MQ-series gas sensor detects the presence Propane (LPG), Carbon Dioxide (CO2), Methane (CH4). Threshold value set is 50.
4. Raspberry Pi 3
5. Arduino Uno

Software:
1. PushBullet API
2. Python 3
3. Arduino IDE
4. Raspberry Pi OS

VII. Results
a. Outputs
According to the task performed by us on the system, the real time environment around the grain could be gauged and if the parameters crossed the threshold values respective control actions could be undertaken and the respective authority would be notified of the same.



Figure 5.1(a) Push Bullet API Notifications

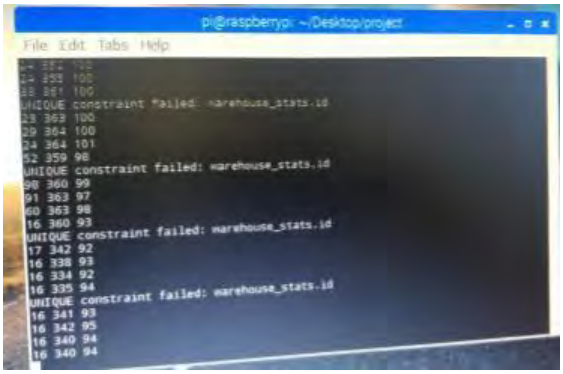


Figure 5.1(b) Current status per sec of sensors

b. Outcomes

Phase 1

i. Planning: Proposing statement of work , scope definition and scope boundary for planning the prototype from the definition to decide what needs ti be done and what not to be done.

ii. Analysis: Creating the synopsis, literature survey and feasibility study for proposed solution which will act as an input to the design phase.

iii. Design: Proposing Design of sensor nodes for wireless sensor network for implementing the surveyed case studies and synopsis. Creating a visually appealing and easy to use front end on web page deployed by warehouse.

iv. Coding: Used Python language to program Raspberry Pi 3.Used pre-configured Pushbullet API which provide us with a token which can be used to send messages directly to the mobile phone of the registered user.

Phase 2

i. Implementation: Various modules have been created:

- Connection of sensors to Raspberry Pi 3 through Arduino UNO
- Completion of coding using Python
- Collecting data from sensors
- Setting up notification system
- Displaying the collected data

ii. Testing:

- The Integrated prototype will be tested exhaustively within the test cases to validate and verify the prototype's functioning (unit /performance testing) and perform integration testing, system testing and stress testing.
- Sample data will be collected and tested for smooth process.

iii. Deployments:

- Real time Deployment of project
- Collecting data from sensors
- Displaying Data on monitor
- Sending notification to the authority

VIII. Conclusion

In this Project advanced intelligent system is designed for monitoring and controlling of the Grain condition in grain warehouse. The implementation is done through Raspberry Pi 3 and sensor technology which provides good performance, clear structure and good scalability. The grain system design is done to meet all the requirement and specification as mentioned in the objective. It is shown from the design implemented in the current work provides flexibility, scalability, portability and security/integrity of the data transmission over wireless network with lower power consumption. Smart warehouse monitoring system serves as a reliable and efficient system for monitoring various parameters that affects the quality of grain. The correct and appropriate actions can be taken to retain the quality of the grains.

Thus, in this work we proposed to design a project named Smart Warehouse Monitoring System which will help to monitor grains and take necessary actions to maintain the quality of grains and accordingly send alerts to the farmer and warehouse administrator.

IX. Acknowledgement

The people we would to thanks are many for their help in presenting this project seminar. we specially want to thanks our Principal Dr.B.K.Mishra, our HOD-IT Dr. Rajesh Bansode and our project guide Dr. Kamal Shah whose advise have inspired and shaped much of the content of this project also guiding us on this project work and spending hours working to improve the content of this project. Also we want to thanks to all the staff members who help us to improve. So majority of our credit goes to all staff members.

X.References

[1] Textbook references

1. Python Programming for Raspberry Pi by Richard Blumand Christine Bresnahan.
2. Getting Started with Raspberry Pi by Matt Richardson & Shawn
3. Internet of Things: A Hands-On Approach- By Arshdeep Bahga, Vijay Madisetti
4. Internet of Things: Principles and Paradigms -Amir Vahid Dastjerdi, RajkumarBuyy

[2] Journal references

1. “Vertical Car Parking – A Prototype”, Sawankumar G. Narone, Swapnil S. Chabukswar, Shriharh A. Valyal, Ravikant B. Hirapure, Prof. V. R. Solapure. International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com ,ISSN 22502459 Volume 5, Issue 4, April 2015. 2. “Conveyorised Systems for Smart Warehouse of Food Corporation of India”, TENDER NO. : TID/CN014/2018L/LT, SOCIETY FOR APPLIED MICROWAVE ELECTRONICS ENGINEERING & RESEARCH, WEB SITE : www.sameer.gov.in 3. “IOT based smart agriculture”,Nikesh Gondchawar, Prof. Dr. R. S. Kawitkar , International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 6, June 2016. 4. “IMPLEMENTATION OF SMART FARM MONITORING USING IOT”, M.Mahendran, G. Sivakannu, Sriraman Balaji3,Assistant professor, Dept. of ECE , Saranathan College Of Engineering, Venkateswaranagar, Panjappur, Thiruchirapalli, ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-4, ISSUE-6, 2017. 5. “Intelligent System for Monitoring and Controlling Grain Condition Based on ARM 7 Processor —,Vinayaka H, Roopa J PG Student in VLSI Design and Embedded System, Assistant Professor, Dept. Electronics and Communications, R.V College of engineering Bengaluru, India. [3] Web references

http://www.fao.org/docrep/015/i2433e/i2433e10.pdf http://www.sameer.gov.in https://www.zebra.com/ap/en/solutions/transportation-logistics-solutions/warehouse.html https://www.researchgate.net/file.PostFileLoader.html?id=58bc1455cbd5c2b9d72690a8&assetKey=AS%3A468543910223877%401488720981390 https://www.raspberrypi.org/documentation/usage/gpio/

Tracking System for Mentally Challenged People

Chirag Rai
Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India

Aditya rao Nuty
Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India

Mr. Namdeo Badhe
Assistant Professor
Thakur College of Engineering and Technology
Department of Information and Technology
Mumbai, India

Ayush Kedia
Department of Information Technology
Thakur College of Engineering and Technology
Mumbai, India

Abstract —**The treatment for Mentally challenged patients can be very troublesome and difficult for the doctors and the patient's family as they have to be attended to constantly which is quite tiresome for the patient's family and doctors. Also, the patients may be unable to express any type of discomfort that they may be feeling, therefore, we create a device and supporting device to help locate the patient and monitor the patient's heart rate. The device will monitor the patient's heart rate constantly so if there is any drastic variation in the patient's heart rate a list of emergency contacts can be notified. The device is connected to an android device that will upload all the patients data on to a Website so that the doctors and patient's family can easily track the patient.**

Keywords – analysis; GPS; Website; database; monitoring; device; smart band.

I. INTRODUCTION

Census 2001 has revealed that over 21 million people in India as suffering from one or the other kind of disability. This is equivalent to 2.1% of the population. Among the total disabled in the country, 12.6 million are males and 9.3 million are females. Although the number of disabled is more in rural and urban areas. Such proportion of the disabled by sex in rural and urban areas. Such proportion has been reported between 57-58 percent for males and 42-43 percent females. The disability rate (number of disabled per 100,000 populations) for the country as whole works out to 2130. This is 2,369in the case of males and 1,874 in the case of females.

Tracking system for mentally challenged people is an device which is primarily aimed to help mentally challenged people by continuously monitoring their location and notifying the emergency contacts in case some unusual behavior is seen. The proposed system is connected to a Website and all the database of the patients are stored on the Website and whenever required can be retrieved. Also the patient's heart rate will be monitored and recorded so that any other heart indication which is different from that record will be checked upon for an emergency.

To locate the patient, we use a portable GPS device in the smart band and a heart rate monitor to monitor the patient heart rate. In case of any emergency, the emergency contacts are notified. All the patient's database is stored on the Website. The smart band is connected to a smart phone using android programming which uploads all the data to the Website.

II. BACKGROUND

About 15 percent of the world's population -- some 785 million people -- has a significant physical or mental disability, including about 5 percent of children, according to a new report prepared jointly by the World Health Organization and the World Bank. Disability is further complicated simply by definition, as was evident in two surveys done in Ireland in the past decade. In 2005, the Central Statistics Office conducted a rigorous survey of the population that included two questions about physical functioning. It found that 9.3 percent of the population was disabled. Later that year, a separate survey interviewed households more extensively and added chronic pain and breathing problems to the menu of disabling symptoms. It came up with a national disability rate of 18.5 percent.

The **Internet of things** (**IoT**) is the network of physical devices, vehicles, and other items embedded with electronics, software, sensors, actuators, and network connectivity which enable these objects to collect and exchange data. Each thing is uniquely identifiable through its embedded computing system but

is able to interoperate within the existing Internet infrastructure. The IoT allows objects to be sensed or controlled remotely across existing network infrastructure,[5] creating opportunities for more direct integration of the physical world into computer-based systems.

.

III. MOTIVATION

Problems regarding monitoring of every single patient and his or her activities in the hospital are a major part for better functioning of the hospital as well as for the family of the patient. So the family of the patient face many problems such as if the patient's shows up with some unusual activities or is in some danger so for overcoming this problem we came with our project in which we will make a track band which will track all the patients activities and heart rate of the patient so that even if any unusual activity is observed so the emergency contacts such as patients family, hospital or the ambulance can be notified.

The main purpose of our project is to analyse the data objects which will be monitored timely and provide growth opportunities to the patient. Through the survey we found that there is need to develop such bands which will be used by the hospitals effectively to provide stability to the mentally ill patients and their families. And this motivated us to develop a smart band which will fulfil the needs of the patient as well as the family of the patient.

IV. SCOPE

The patient may have a smart band on his wrist that may continuously monitor patient's location and heart rate and may update it to the Website. The patient data is stored on the Website and the doctors may access all the patients' data through Website. In case of any unusual behaviour the emergency contacts are notified about the patient so that his/her parents or guardian can look after the patient or the hospital can have check over it.

V. PROBLEM DEFINITION(PHASE WISE)

Phase 1: Analysis and Planning

We have taken reference of 3 literature survey papers that contain knowledge including substantive findings, as well as theoretical and methodological contributions similar to our topic. The reviews in these papers re secondary sources, and do not report new or original experimental work. It is associated with academic-oriented literature, and are found in academic journals. We have used the key findings of these papers that add value to our project and elaborated them.

Phase 2: Designing and Implementation:

To locate the patient, we use a portable GPS device in the smart band and a heart rate monitor to monitor the patient heart rate. In case of any emergency, the emergency contacts are notified. All the patient's database is stored on the Website.

The smart band is connected to a smart phone using android programming which uploads all the data to the Website.

Phase 3: Testing:

To check if the GPS and heart rate monitor are working properly and sending all the data accurately to the smart phone.

To check if all the data is uploaded correctly to the Website.

Phase 4: Deployment:

If new key findings are found during testing phase we will add up to our project and deploy the same product to the proposed market.

Methodology

1. A device is to be created which will track the patient's location and his heart rate and notify the emergency contacts if any danger is seen.
2. An device is to be created that is linked to the device and provides all the functionalities of the device to the users.
3. The device has to be linked to a Website service sos as to make the data easily available to all.

Overview of project

VI. FEASIBILITY STUDY

Technical Feasibility: The system works on the app model and is the system implemented using Android and C/C++ for Arduino. as programming and scripting languages. Proven and tested mature technologies to deliver the proposed solution.

Economic Feasibility: The Website owner has a more than one Websites to operate then our security system will not be applicable for providing security, therefore in the future enhancement it can enhance our existing device to manage a Website environment which has more than one Website. Website computing is a future for organizations. The considerable benefits that provide will make eventually all the organizations totally move their processes and data to the Website. Even after the development, the organization will not be in a condition to invest more in the organization. Therefore, the system is economically feasible.

Legal Feasibility- The project is developed under legal license of Android, Arduino and other software's and databases. The data will not be used for some illegal purpose and will be encrypted using algorithms.

Operational Feasibility- An overview of the process portrayed that the proposed Tracing System and Android device would be beneficial by providing a unique concept to users hence being the only player in the market.

VII. FLOW DIAGRAM

Figure 1 is a flowchart diagram. Flowchart is a type of diagram that represents an algorithm, workflow or process, showing the steps as boxes of various kinds, and their order by connecting them with arrows. This diagrammatic representation illustrates a solution model to a given problem.

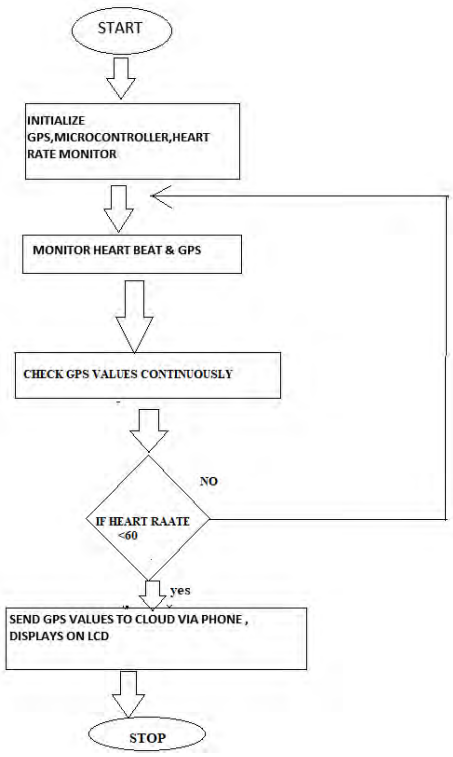


Figure 1: Flow Chart Diagram for registration and sign in.

`

VIII. RESULTS

Phase 1: Analysis and Planning:

Collected all the required information and tools regarding the building of the smart band which will help to track the patients via different set of activities.

Phase 2: Designing and Implementation:

Designed the basic GUI of the device which contains all the basic functionalities provided to the users, thereby simplifying the process of tracking the patients.

Phase 3: Testing:

With the developed devices different test cases were performed to find the efficiency of the product.

Phase 4: Deployment:

After performing different test cases the product is ready for deployment and now can be used by the patient's family to track their daily activities.

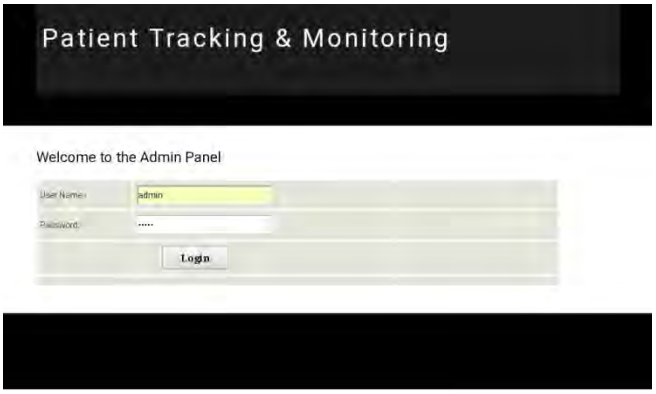


Figure 2: Login page of the device.

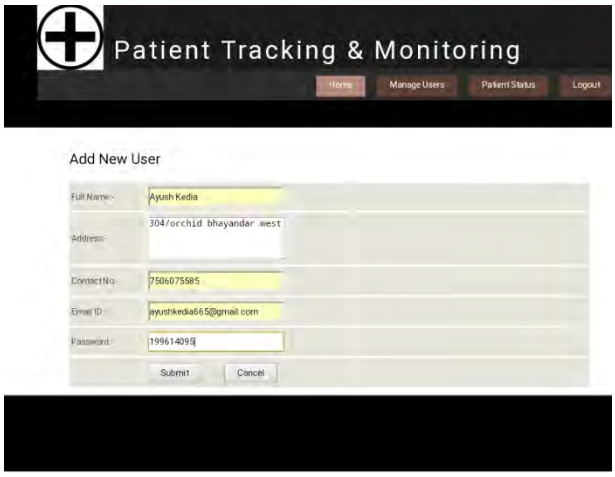


Figure 3: Sign Up page of the device.

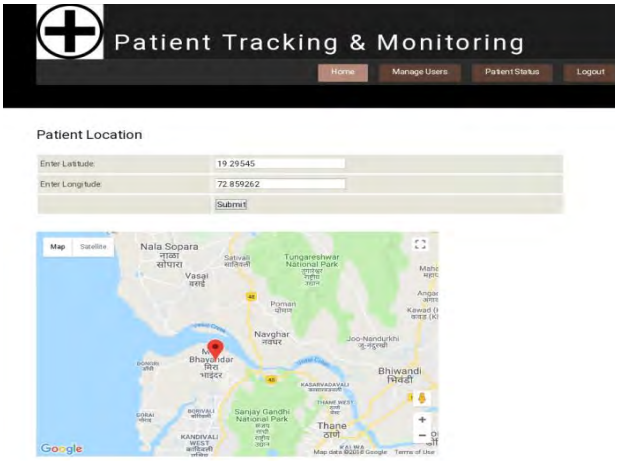


Figure 4:patients location address page

IX. CONCLUSION

The tracing system has been a work in progress in terms of the backend connectivity and the user interface. The final version of the app is expected to support the entire feature set,along with the smart band being able to efficiently track the patient. The tracking system is expected to strive based on the lack of any competition in the market currently.

X. REFERENCES

Research Papers:

[1]. KaYiu Sham, Philip Lim-Kong Wong"Fitness monitoring device having an electronic pedometer and a wireless heart rate monitor", Apr 6, 1999, In US 5891042 A

[2]. Erin-Ee-Lin Lau, Wan-Young Chung "Enhanced RSSI-based Real-Time user location tracking system for indoor and outdoor environments ",23 Nov. 2007,10.1109/ICCIT.2007.253.

[3]. R. Sumitha1, C. Vijaya2,S.Parimala" Location Tracking System for mentally challenged people", 1 March 2017, ICITET17002.

,

[4]. Tom OuYang, Farah FerdousIshita, Peter Le Ashton Novak-Louie, GiftyQuansah"– Intelligent wearable wristband for personal safety."February 12,2016, In ENSC 440.

Text Books:

1) Moving to the Website- SITARAM
2) Arduino: 101 Beginners Guide-ERIK SAVASGARD

Web References:

1) http://www.ijitet.com/ICITET17/ICITET17002.pdf
2) https://www.google.co.in/patents/US5891042
3) http://ieeexplore.ieee.org/document/4420422/

Machine Learning Methods for Online News Popularity Prediction

Ravi Sitapara
I.T Department TCET Mumbai, Maharashtra
ravisitapara007@gmail.com

Pradeep Kotian
I.T Department TCET Mumbai, Maharashtra
pradeepkotian1995@gmail.com

Prashant Chaudhary
I.T Department TCET Mumbai, Maharashtra
pchaudhary31@gmail.com

Shridhar Kamble
I.T Department TCET Mumbai, Maharashtra
shridhar.kamble@thakureducation.org

Abstract—Machine learning techniques are utilized to find the best model and set the features to predict the popularity of online news. Our data comes from Mashable, a well-known online news website. Using over 49000 sample articles from mashable.com different learning algorithms are implemented on the dataset, ranging from various regressions to SVM and Random Forest. Their performances are recorded and compared. Feature selection methods are used to improve performance and reduce features. Our work can help online news companies, editors, and bloggers to predict news popularity before publication. Accuracy & speed of different algorithms are being measured first. After running the algorithms on the dataset, it showed distinct results. The results were compared on the basis of their accuracy & speed to compute the articles.

Keywords— Machine Learning, News Prediction, Random Forest, KNN, Likes, Comments, Shares, Views, Regression.

I. INTRODUCTION

Online website has been the majority source for news to spread. Interesting news is going to be shared thousands of times through the Internet. With the expansion of the Internet, more and more people enjoy reading and sharing online news articles. Popularity prediction for news articles is a relatively novel problem and very few studies addressed this problem. However, a growing number of studies have been carried out on predicting the popularity of other types of online content. The number of shares, likes, comments under a news article indicates how popular the news is. In this project, we intend to find the best model and set of features to predict the popularity of online news, using machine learning classification techniques. Prediction will be based on the topic which will be the current trends in the online digital feed and depending upon this the news article will give a rank. As if the news article shares, likes and comments changes the ranking position will be changed. Depending upon the rich features the news article will predict the popularity and getting the top position in the news feed world.

II. OVERVIEW

The Online News Prediction of our project will be based machine learning algorithms & methods. Our model is based on random forest regression and learned on a set of features derived from the given time series data to capture the expected number of visits, rate of change. This project will be able to tackle the problem of ranking online news by using a two-phase procedure. First step consists understanding two underlying properties of news articles that are relevant to our ranking problem: articles' lifetime and distribution of popularity. Then, based on these observations, we recommend a more rigorous evaluation strategy adapted to the characteristics of news.

III. LITERATURE REVIEW

Using SVM, Naive Bayes, Trees, Bagging, Decision Tree & Regression it shows an accuracy of 84% in predicting the popularity range of a news article [1]. They have extracted 2.89 million tweets referring to 24 different movies released over a period of three months [3]. The Accuracy found for different algorithms were: i) Linear Regression -66% ii) Logistic Regression -66% iii) Random Forest (500 Trees)- 69%. Four classification methods were tested to predict three popularity classes (1 to 20 tweets, 20 to 100 tweets, more than 100) and results ranged from 77% to 84% accuracy, for Naive Bayes and Bagging, respectively and using over 39000 sample articles from Mashable.com [4]. K-Means clustering, Polynomial and Linear Regression was applied on 2510 movies released 1990 onwards to study and build a predictive model to get the expected revenue. They achieved accuracy of 36.9% [5]. They are able to achieve the least accuracy with neural network multilayer perceptron 79.07% and highest with simple logistic of 84.34% [6].

IV. PROBLEM DEFINITION

This project, intends to find the best model and set of features to predict the popularity of online news articles, using machine learning techniques. And which will be predicted based on

number of shares, likes, comments under a news articles.

Phase 1(Information Gathering):

Extraction of more than 49,000 sample of article from mashable.com (which is UCI machine learning repository). It extracts 59 attributes (as numeric value) describing different aspects of each article, published in the last two years from Mashable website.

A. Phase 2(Analysis & Design):

Prediction models Random Forest, Linear Regression, Logistic Regression & Naïve Bayes Classification Methodology are done on articles for Preprocessing using Python. Prediction model's accuracy will be measured & compared among all the Classification Methodology. It would be correct if its value and the actual result have the same sign (both + or -).

B. Phase 3(Implementation & Testing):

Development of Graphical user interfaces (GUI) using Django Framework. Django will dynamically serve web pages without using database store information. Django framework will provide inputs like comment, share, likes, dislikes.

C. Phase 4(Deployment): -

Pre-Processed data will be integrated onto the Front-End using Django framework. A ranking table will be provided which will rank the articles according to the popularity. Graphical representation of factors which provides top rankings.

V. FEATURES

The News Prediction of our project will be based machine learning algorithms & methods. Our model is based on random forest regression and learned on a set of features derived from the given time series data to capture the expected number of visits, rate of change.

Prediction will be based on the topic which will be the current trends in the online digital feed and depending upon this the news article will give a rank. When the news article shares, likes and comments changes the ranking position will be changed.

Depending upon the rich features the news article will predict the popularity and getting the top position in the news feed world.

It tackles the problem of ranking online news by using a two-phase procedure. The first step consists in understanding two underlying properties of news articles that are relevant to our ranking problem: articles' lifetime and distribution of popularity.

Then, based on these observations, we recommend a more rigorous evaluation strategy adapted to the characteristics of news.

VI. METHODOLOGY

The Proposed methodology contains following steps: -

- Data Extraction
- Data Preprocessing
- Data Integration and Transformation
- Feature Selection
- Classification

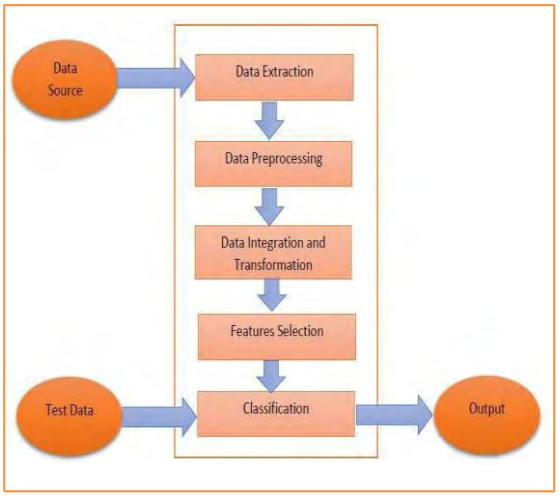


Figure 1. Methodology

Random Forest:

In bagging (Bootstrap Aggregation), numerous replicates of the original dataset are created to reduce the variance in prediction. Random Forest uses multiple decision trees which are built on separate sets of examples drawn from the dataset. In each tree, we can use a subset of all the features we have. By using more decision trees and averaging the result, the variance ofthe model can be greatly lowered [9].

DECISION TREE:

Decision tree is commonly used in classification in machine learning, it is a tree like graph or model to make decisions and get consequent results. Decision tree implicitly perform variable screening or feature selection and the best feature of using it is easy to interpret and explain [9].

KNN:

K nearest neighbors is a non- parametric method used for classification and regression. The input normally consists of k closest training instances in the space. An object is classified to a class because it is most common among its k nearest neighbors. It is chosen because KNN don't need make any assumption on the underlying data distribution, it is also lazy algorithm and don't use training data to do any generalization [9].

NAÏVE BAYES:

Naive Bayes classifier is a probabilistic classifier based on Bayes' theorem. It depends on the assumption that some features are conditionally independent. It is super simple and can converge quicker than some other models, so it has a chance to achieve fast and easy process [9].

VII. FESEABLITY STUDY

- **Technical Feasibility** – The application will operate on Web Browser. The system will be implemented using python, Django python framework, and web technology as required.
- **Economic Feasibility**- In this project, we tried different methods to predict the popularity before its publication. By predicting the news it will helpful for news publication to put the relevant topic for good revenue. Therefore, the system is economically feasible.
- **Legal Feasibility**-The project is developing under legal license of windows, python and other software's required. The data will not be used for some illegal purpose and will be used for making the news relevant to user as well as for publisher for making under certain constraint**.**
- **Operational Feasibility** -An overview of the process portrayed that the proposed Machine Learning for Online News Popularity Prediction be beneficial for news publisher to monitor the growth and improvement in the future article content.
- **Social Feasibility**-Easy to make allocation and accurately with this help online news publisher to predict the news popularity and also make user to read the current trending topic.

VIII. BLOCK DAIGRAM

This below block diagram shows the process of training the model using training data and implementing the model for getting the testing data.

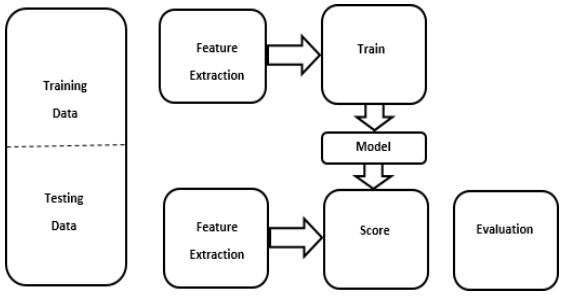


Figure 2. Training and Testing data

By training the model it will be used for input testing the input data. This below diagram shows how the model in which the sample articles are being segregated into popular and unpopular articles using the model which is trained using the 49k sample.

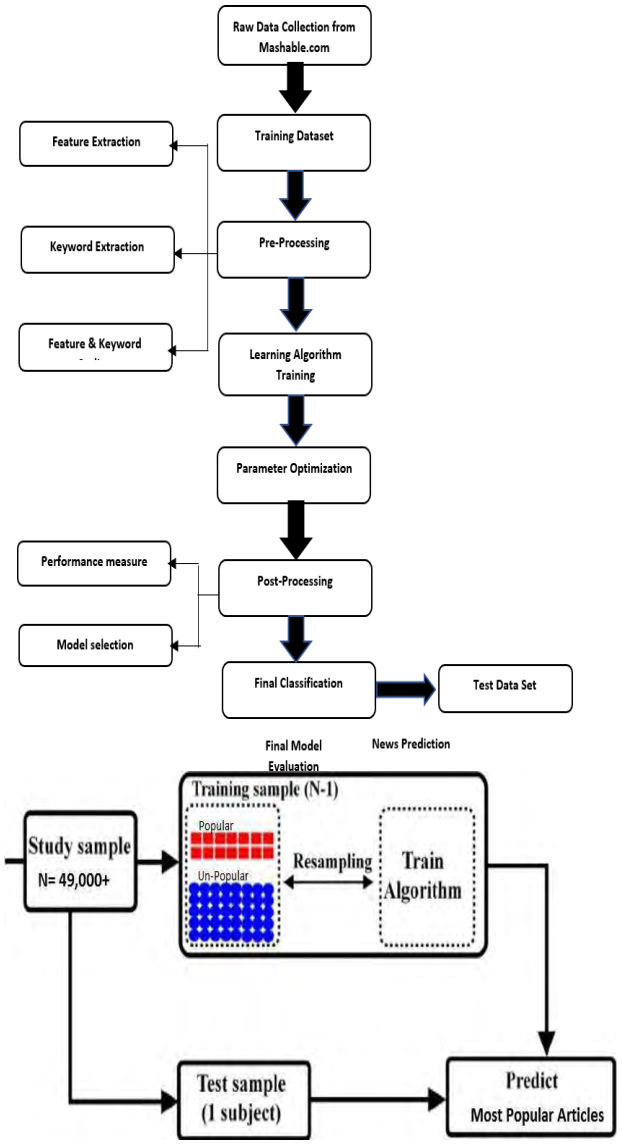


Figure 3. Predict the popular and unpopular articles [14]

IX. IMPLEMENTATION

Below flow chart shows the collection of raw data from online repository and training the data using different machine learning algorithm like random forest, decision tree, KNN. Once the model is implemented then the testing data will be provided for predicating the popularity and measuring the accuracy and time lapsed.

Figure 4. Flow Cha

RESULTS

After implementing the following algorithms like decision tree, KNN, Random Forest, Naïve Bayes, it is being overserved that random forest achieves better accuracy and Naive Bayes shows less time elapsed. Performance analysis carried out and its and results are as follows:

TABLE 2: ACCURACY

Algorithm	Decision Tree	KNN	Random	Naïve Bayes	SVM
Accuracy (Original)	57.9%	57%	66.4%	61.2%	53.1%
Accuracy (Cross Validation)	55.5%	54.9%	64.4%	60.8%	53.3%

TABLE 2: SPEED

Algorithm	Decision Tree	KNN	Random	Naïve Bayes	SVM
Time elapsed (Original)	1.2s	3.23s	6.33s	0.07s	366s
Time elapsed (Cross Validation)	8.44s	10.83s	37.71s	0.48s	2112s

From two tables above, it is observed Random Forest model achieves the highest accuracy (66.4%), but it consuming the most time (6.63s) among 4 models. Naive Bayes model achieves the least time elapsed (0.07s) and a properly medium accuracy. The other two models have bad accuracy. The K- fold cross validation has proved that the accuracy should be less than that gained from the original method. But considering its big time consuming, the difference can be ignored relatively under the purpose of comparing these 4 models.

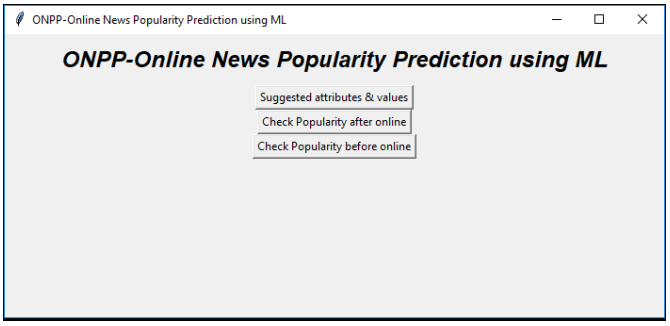


Figure 5: Front-End Using Python Interface Tkinter

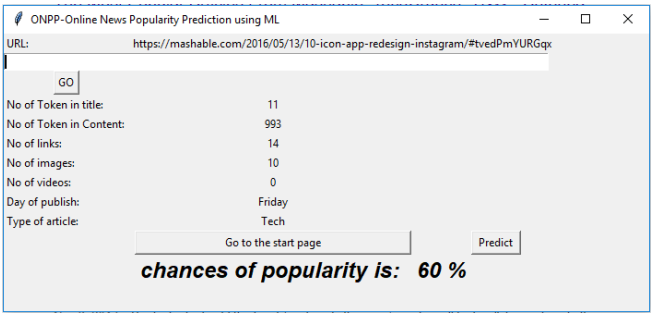


Figure 6: Online Article with 60% chances of popularity

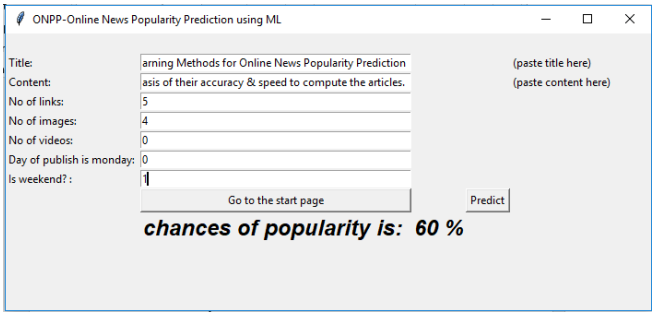


Figure 7: Article with 60% chances of popularity

X. CONCLUSION

Current state-of-the-art was reviewed on News popularity prediction methods. We presented the different prediction methods, reported their performance, and suggested several applications that can benefit from these findings.

Various regression and classification models were trained for predicting the popularity of an online news article before its publication. We use shares to indicate popularity. The models were trained using Mashable data from UCI.

Overall, this project pursues a faster process; Naive Bayes is the best choice for this problem without sacrificing too much accuracy. Otherwise, if accuracy is considered first, obviously Random Forest is a better one.

REFERENCES

[1]. Alexandru Tatar, Panayotis Antoniadis, Marcelo Dias de Amorim, Serge Fdida. From Popularity Prediction to Ranking Online News. Social Network Analysis and Mining", Springer, 2014, pp.4:174.

[2]. Babita M. Jangid, Chaitali K. Jadhav, Swati M. Dhokate, Girish M. Jadahv, S urvey on Movies Popularity Prediction System Using Social Media Feature, IJIRCCE, VOL. 4,2016

[3]. Muhammad Hassan Latif, Hammad Afzal, Prediction of Movies Popularity Using Machine Learning Techniques . IJCSNS, VOL.16 No.8.

[4]. He Ren, Quan Yang, Predicating and Evaluating the Popularity of Online News , CS -229, 2016

[5]. Xuandong Lei, Xiaoti Hu Hongsheng Fang, Is Your Story going to spread like a Virus? Machine Learning Methods for News Popularity Prediction. IEEE CS -299,2015

[6]. Prof. G.M.Bhandari Asur, Sitaram, and Bernardo Huberman. "Predicting the future with social media." Web Intelligence and Intelligent Agent Technology (WIIAT), 2010 IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, 2010.

[7]. Fernandes, Kelwin, Pedro Vinagre, and Paulo Cortez. "A Proactive Intelligent Decision Support System for Predicting the Popularity of Online News." Progress in Artificial Intelligence. Springer International Publishing, 2015. 535-546.

[8]. Statistical Learning and Data Sciences : Alexander Gammerman,Vladimir Vovk,Harris Papadopoulos(Springer).

[9]. Machine Learning in Python (Essential Techniques for Predictive Analysis): Michael Bowles. (Wiley)

[10]. T he Definitive Guide to Web Development Done Righ t drain Holovaty & Jacob Kalpan-Moss (Apress).

[11]. A Primer on Scientific Programming with Python: Hans Petter Langtangen.

[12]. UCI Machine Learning Repository
http://archive.ics.uci.edu/ml/datasets/OnlineNewsPopularity access ed on 17/7/2017

[13]. Github
http://inseaddataanalytics.github.io/INSEADAnalytics/groupprojects/Ja n uary2017/MashableNews.html accessed on 8/8/2017

[14] Researchgate
https://www.researchgate.net/figure/288993135_Flow- diagram-showing-algorithm-training-and-testing-process-The-class-imbalance- problem accessed on 18/9/2017

Web Application on Result Analysis

Kirtan Patel, Maulik Shah, Rishab Tiwari, Ashish Soni
Dept. of Information Technology Thakur College of Engineering and Technology
Mumbai, India

Abstract: **The goal of this project is to create web application for computer science and engineering department; purpose of this is to manage the "result analysis and time table generation, progress report, placement". It is helpful for students as well as the college authorities. In the current system all the activities are done manually. It is very time consuming and costly. This web application gives authorization to the staff to view their result details, academic details, allotted subject and Time Table. The project is more user-friendly & easily operated so that anyone have minimum computer knowledge can approach and handle the software without having any convoluted. The commonly used tools in the system are Data Flow Diagram, ER Diagram, Block Diagram, Flow chart. Training, experience and common sense are required for collection of relevant information needed to develop the system.**

Keywords- Data Mining, Natural language Processing, Cobweb Algorithm.

I. INTRODUCTION

This is a tracking application where at a time we can see all the year‟s result in a single sheet and a choice of viewing the candidates result separately is also possible. The proposed work is defined by the work to be done in the project and also the ways to complete the project successfully. It will help give clarity on what the project is going to achieve. The structure of the project is defined in this section. The flow of the processes that will be followed in order to produce successful results or outcomes is also listed in this section for better understanding of the project. Also states the probable solutions that could be achieved by the project. The main objective of the project is to provide the examination result to the student in a simple way.

This project is useful for students and institutions for getting the results in simple manner. By a result analyzer with subject status and marks is an application tool for displaying the results in secure way.
The system is intended for the student. And the privileges that are provided to student are to read and execute his/her result by providing user name and password for secure login and in case of new student the registration is available. And the guest user has the privilege only to read. The whole result analyzer will be under the control of the administrator and the admin as the full privileges to read, write and execute the result. And admin gives the privileges to the Teacher and student and the guest user to access the result.

II. IMPORTANCE

In this project students‟ performance is evaluated according to how they perform academically. The main importance is that it becomes easy for the faculties to list out the students list according to their performance. Using this web application, the students with the lower grade can be sorted easily so that more attention can be put on that particular list of the students.

III. OVERVIEW

Student report generator is a web based application designed and engineered for colleges that need to manage results across multiple branches students that need to track, manage and report results. This application can run on any kind of operating system. At a time we can see all the years result in a single sheet and we can see the individual candidate‟s results separately

IV. PROBLEM DEFINITION

The traditional approach of maintaining the data of student is done manually and this work is too tedious and time consuming. As a human the possibility of entering wrong information by a single person is also high. Hence the Analysis of student is not accurate as per his criteria such as marks, attendance, Term work etc. This website bridges the gaps of all the previous proposed software like Outlier analysis of data according to Student record[Ref 1] , Human Friendly results is obtained with the help of NLP [Ref 2] and K-means algorithm help to analysis the data in an efficient way [Ref 4].

Phase 1(Planning, Analysis, Design, Coding, Implementation): -

1. Perform literature survey and identify all the research gaps, learn about the marks obtained by individual in each semester.
2. Design the database and different modules of the project.
3. Integration of data from all the tables of the database into a data warehouse.
4. Create code for Analysis of data and generate a pie chart and bar Graph according to marks obtained by students
5. Code to create a software that fills all the research gaps such as Outlier analysis of data according to Student record, Human Friendly results is obtained With the of NLP and K-means algorithm help to analysis the data in an efficient way.

Phase 2(Testing, Deployment): -

1. Create various cases of test case data to check if the integrated data warehouse functions as desired by the client.
2. Conduct beta testing for identifying any further errors and the identify and improvements and that can be performed.
3. After testing and approval, deploy the proposed system to the college.

V. FEATURES OF PROJECT

"Progress report of student and, result analysis" this project is mainly designed to reduce manual work.
In the Progress report of student generates the report according to students roll number, semester, year and other details.
The result analysis is generating the graph according to semester university result of students. User friendly interface.
Fast access to database.
It consists of modules like student login, faculty login.
Providing the result in a pattern like internal marks, external marks, total marks.
Providing results in CGPI as well as Percentage.

VI. METHODOLOGY

Progress report analysis

I. Importing data from excel sheet-

The data is give as an input in the form of excel sheet which contains different number of fields. The fields like roll number, percentage of attendance, number of practical"s, lectures attended by all student.
We take 3 excel sheet as an input those are:
1) Student Details
2) Month wise Attendance Details
3) Month wise unit test marks Details

Student details are provide information like serial number, roll number, name of student, batch. Month wise attendance details contains information like number of taken lectures, practical"s, total attendance, total percentage, status about defaulter, etc. Month wise unit test marks details contain subject wise marks of each roll number.

II. Store data into database in required format-

After getting all data it is stored into database by cropping only required field of the imported data. A structure of imported sheet is totally different then database structure so it select only required field from imported data.

III. Generation of crystal report-

Progress report is created as per required format by using the crystal report tool. When proctor provides the inputs like class, year and proctor name, etc. the data will be fetched from the database and crystal report of student will be generated.

IV. Search data of particular student according to roll number and class-

When there is need to search details of particular student then this facility is useful. We must require to give input as roll number, class name, and year of studding to finding data within database.

V. Print the progress report on crystal report-Printing the report on print paper is one of the main features of our application. Printing only required data on paper like printing of bank passbook. It means that we are able to print the details of same student, on same paper having result of different semester.

VII. RESULT ANALYSIS

I. Importing data from excel sheet-

II. We take excel sheet as an input those are:

Subject wise Student marks Details are provide information like serial number, roll number, marks of each subject, and name of student.

III. Storing data into database in required format-

After the getting all data is stored into the database by cropping only required field of the imported data. Whatever fields present into imported data that all as it is not stored into database. A structure of imported sheet is totally different then database structure so it"s required to select only required field from imported data.

IV. Generation of semester report for overall performance-

Here we are generating the report of overall performance of student for all subjects by considering there mark. We are maintaining all details according to semester. When user needed to find any record then there is need to get input as semester, year, and class.

Time Table Generation:

I. Taking data from user-

a. Subject Details
b. Subject teacher
c. Batch
d. Class rooms/Lab
e. Time Details

All input is providing as per semester and year for all classes.

II. Storing data into database in required format-After getting all data it is stored into database by insert only required field of the inputted data. Whatever fields enters into input data that all as it is stored into database.

III. Generate time table-

After getting all the details an algorithm automatically generate time table. It does not contain any overlapping of lectures or practical"s.

IV. Make the changes in time table as per need-

If any faculty wants to changes in timetable then make changes into editable slots of timetable. It should not any overlapping of lectures or practical‟s.

VII. Feasibility Study

Result analysis is a web based application designed for colleges that need to manage results across multiple branches students that need to track, manage and report results.The project is analysed first by the project team and after the analysis phase, planning phase is designed. In this phase, the layout of the project is designed and also the working flow. Project planning includes Costing, budgeting and scheduling of the project.

- **Technical Feasibility** – To determine whether the proposed system is technically feasible, we should take into consideration the technical issues involved behind the system. If the system is not comprehensive which supports manyplatforms to work with that goes to proposed system is technically feasible.

- **Economic Feasibility** – To decide whether a project is economically feasible, we have to consider various factors as:
  I. Cost benefit analysis
  II. Long-term returns
  III. Maintenance costs.

- **Operational Feasibility** - To determine the operational feasibility of the system we should take into consideration the awareness level of the users. This system is operational feasible since the users are familiar with the technologies and hence there is no need to make changes to the system. Also the system is very friendly and to use.

Tables:

- Project Planning Table

Organizational Impact	Value	Metric	Time Frame
Strategic	Increases the amount of analyzing large amount of student data.	Increase the amount of data compiled by 40%	12 Month S
Customer	Reduce the result analyzing time .	Reduce the average wait time between evaluation of papers and display of results.	6 Month S
Financial	Reduce personnel & communicatio n cost of scheduling.	Reduce personnel & communicatio n cost of analyzing by 75%.	
Operational	Reduce the analyzing cycle time.	Reduce analyzing time from 1 min to 30 seconds.	12 Month S

Organizational Impact	Value	Metric	Time Frame
Social	Increase the number of institutes accessing this service.	Increase the number of institutes accessing this service from 10 to 50	24 Month S

Figures:

1. FLOW CHART

The flowchart contains the following modules:

ADMIN:

The admin is responsible for uploading the tsheets generated. He can add the subject marks and make any kind of modifications like storing, updating, deleting. Apart from the regular results, the supplementary results can also be uploaded by the admin.

STUDENT:

Student must be an authenticated user of the college to access this application. The student can view the semester marks, individual subject marks and aggregate till the current semester. The student can also make a comparative analysis with the results generated.

FACULTY:

Faculty must also be authorized users of the college. Faculty can view the class result, subject wise result. The percentages of students pass/fail in a subject can be viewed. The total marks (internal, external and both) can be calculated.

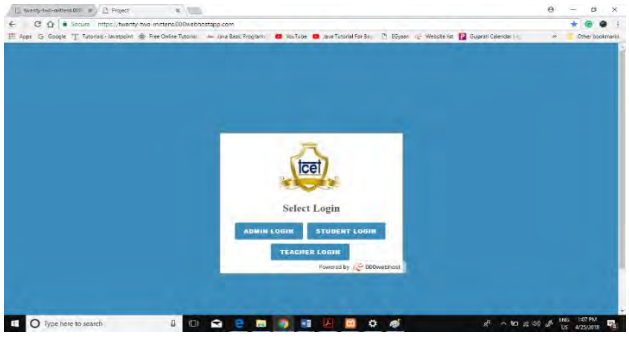
VIII. Results

The computer software application is required to be independent of any platform; It is designed to have three main sections, namely: the login window, the main menu and sub menu. The login window requests a valid user name and password from the dministrator to be able to gain access into the software. The Administrator is any staff that is authorize by the management of the school to be in charge of exams and records unit (e,g the Director of studies), hence he should have a valid user name and password created by him to be able to login to the software

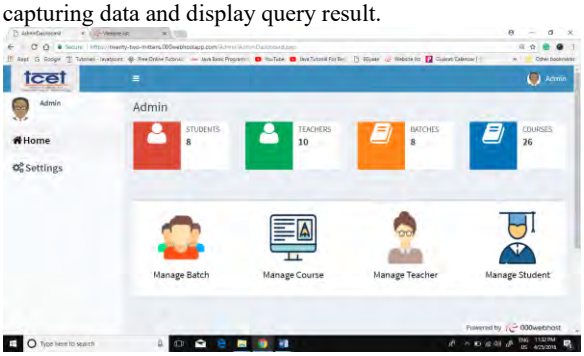
(a)The Login:

The login form/window enables the user to enter his \ her password. If the password entered is valid, the software will then

display the main menu. But if the password is not valid the user will be denied access to the program



.(b)The Main menu:
The main menu contains dashboard and panel, on the dashboard sub menu which serves as a link to all other modules in the program, the sub menu are; Home, Results, Register, profile, supports and contact us. While the panel holds the forms for capturing data and display query result.



(c) Register Sub Menu:
This sub-menu is use for capturing fresh students‟ data or record into the data base such as students surname, first name, other name, class, e-mail, date of birth, sex and admission year/number. It handles students Registration at the beginning of the first term especially because it‟s the term in which fresh students are registered into the school. The student‟s personal information are all captured in this menu.
(d) The Results Sub Menu:
This sub menu contains modules such as register subject, view subject, view students, compute results, and view results. Clicking on any of these modules, operation takes place.
(e) Register subject module:
All subjects offered by students in the school are captured and recorded into the data base in this module. (f) View subject module:
All subjects registered into the data base can be viewed in this module.
(g) View students module:
All registered students in the database are viewed in this module
(h) Compute results module:
This module allows the students‟ Scores to be captured for the purpose of results computation.

In fact this is the
(i) View results module:
This module displays each student result or report sheet as the case may be, the result sheet can be printed from this module as well.
(j) Edit module:
This module allows the already entered data about students to be modified in case of errors or updates.

VIII. CONCLUSION

Our main aim through this project is to reduce manual work, to provide efficient way of handling data. The ‘progress report ‘module our project reduces the work of staff of making entries for each student manually. Result analysis generate graph of students marks and timetable generation includes by taking only the input it will automatically generate timetable.

In future our project can be done as mobile application. So that student can see their progress report on their mobile.

The projects clearly depicts that the Student Result System is very efficient. It shows how the concept of File Handing can be used in database management system in absence of databases like Oracle, MySQL etc. Although, this software can be further modified to be used as multitasking and bigger software, but it effectively works under the condition of limited resources and time.

SCOPE OF THE PROJECT

The aim of this project is to create software for computer science and engineering department; purpose of this is to maintain the “PROGRESS REPORT, RESULT ANALYSIS. The system has an inbuilt feature for adding the details of the students by themselves. So the overhead of the college authorities and the teachers become less.

REFERENCES

[1].C. Romero S. Ventura P. G. Espejo and C. Hervás "Data mining algorithms to classify students ". In EDM 2008 pp. 8-17.
[2].Y. Zhang S. Oussena T. Clark and H. Kim "Use data mining to improve student retention in higher education-A case study=" In ICEIS (1) 2010 pp. 190-197.
[3].B.Namratha, Niteesha Sharma “Educational Data Mining – Applications and Techniques”. In IJLTET Vol 7 issue 2 July 2016.
[4].Z. Erdogan and M. Timor "A Data Mining Application in a Student Database " in Journal of Aeronautics and Space Technologies 2005 vol. 2 No. 2 2005 pp. 53-57

Design and develop algorithm to translate audio into multiple Indian languages scripts using ML

Ms.Akshata Hirve, Mr.Amey Patil, Mr. Anand Khandare

Department of Computer Engineering and Technology, Thakur College of Engineering and Technology, Mumbai, India

anand.khandare1983@gmail.com

***Abstract*— Transcription means a written or printed version of material originally presented in another medium. This service in the current world is used by Educational Institutions and stream based services. The current market trend is to upload your audio file on the service provider's server. This requires the a lot of server space and processing power as a service provider. Also the user is concerned about his data given to a 3rd party website. It leads to a state of trustlessness. The solution we are offering is to provide a software copy to the user on his local computer. It will ensure data privacy and save server space.**

Keywords—trustlessness , data privacy

I. INTRODUCTION

A transcription service is a business service basically based on Speech to Text[1] process which converts speech (either live or recorded) into a written or electronic text document. Transcription services are often provided for business, legal, or medical purposes. The most common type of transcription is from a spoken-language source into text such as a computer file suitable for printing as a document such as a report. Common examples are the proceedings of a court hearing such as a criminal trial (by a court reporter) or a physician's recorded voice notes (medical transcription). Some transcription businesses can send staff to events, speeches, or seminars, who then convert the spoken content into text. Some companies also accept recorded speech, either on cassette, CD, VHS, or as sound files. For a transcription service, various individuals and organizations have different rates and methods of pricing. That can be per line, per word, per minute, or per hour, which differs from individual to individual and industry to industry. Transcription companies primarily serve private law firms, local, state and federal government agencies and courts, trade associations, meeting planners and nonprofits.. The applications of transcripts in today's industry are

- Videos on websites.
- Subtitling.
- Translation.
- Legal and court hearings.
- Market research and focus groups.
- Organization conferences.
- Press conference for journalists.
- Transcribing HR meetings and grievance proceedings.

II. THEORY

Technology used: python, Machine learning

Technology used at Back end: PHP, Mysql

Resources Required: Manpower, Programmers, testers, debuggers

Software required: Testing Tools

(to perform black box and white box testing), Github

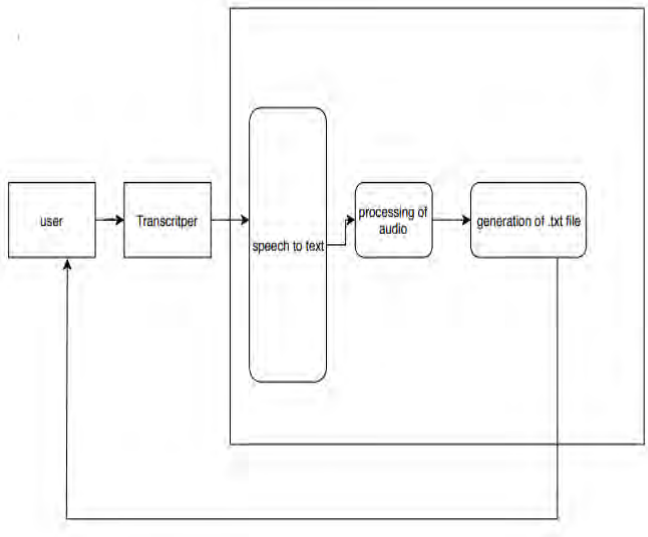
Editor: Sublime Text

Hardware required: PC for development, server for deployment

Infrastructure: For scaling up the project

III. RESEARCH/TECHNICAL/PROJECT WORK

A . System architecture

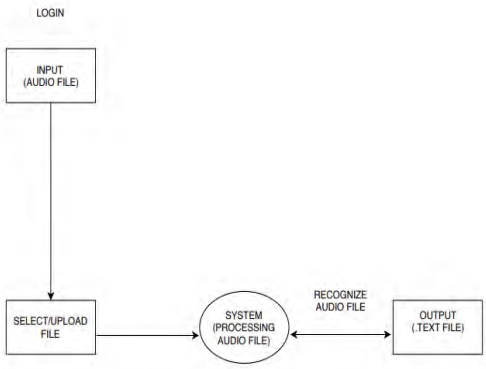


B. Algorithm

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications

that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

C. Flowcharts



D.Literature Survey

A transcription service is a business service basically based on Speech to Text process which converts speech (either live or recorded) into a written or electronic text document. Transcription services are often provided for business, legal, or medical purposes, educational purposes.

There was a time when most businesses used to have a dedicated staff member whose job was to take down notes in the meetings, conferences, interviews and other business events. This process was slow and time-consuming. Moreover, the chances of mistakes were high. Things have changed now. In today's fast-paced work environment, business transcription services have assumed a very important role. In meeting and conferences, the proceedings are audio recorded. These audio recordings are then converted into texts. This not just reduces the chances of errors but also ensures that nothing is missed out. It is impossible for team members to remember everything that was discussed in a business meeting. With transcription services, a business ensures that all business interactions are properly recorded for future references.

Traditionally the wpm of typists used to be 80 wpm ,whereas today's ML software are much efficient and can work for a long time unlike a human. Though the accuracy becomes a issue when using ML.

The software aims to help the transcription industry of India as such that it is stuck with problems of languages

and efficiency as English is the global language and our native languages are In transcript software, user will upload audio file which contains a spoken languages. That audiofile will be recognized by software. After that audio file will be converted into .text file. By using transcript software different process will be perform on audio file. The steps included in process are Log in, select an audio file, upload an audio file and then recognized using an (NLP)

algorithm. Finally, after performing all these operations we will get clear output of uploaded audio file.

IV. RESULT AND DISCUSSION

The software aims to help the transcription industry of India as such that it is stuck with problems of languages and efficiency as English is the global language and our native languages are not of interest to developers. The project also aims to spread knowledge and assistance in learning the generations. It will bridge the gap of transcription in our nation and server the population if 1.3 billion in its native languages.

V. CONCLUSION

Thus the coming generation will be able to conveniently communicate through this software. This will be more helpful to the disable person.

The probability of error and time consumed will be reduced by a considerable proportion with the implementation of the system.

This is completed with partial fulfillment of BE project and we have successfully completed the GUI phase of the project. We aim at successfully analyzed the audio file and generate a .text file by the next phase of the project i.e. in the next semester.

VI. FUTURE SCOPE OF THE PROJECT

Currently there are many available ways of doing this that are increasingly accurate but are designed for speech spoken into the device microphone (e.g. the Google Translate/corresponding API for web, Dragon app for iOS). I need a way to directly feed an audio file into the audio recognition engine. Don't want to play the audio through a speaker and capture it with a microphone takes considerable time for long audio files, and degrades audio quality and resulting transcription quality. The transcription System can be further extended to perform successfully on audio and video. The problem of current transcription is that it does not recognize difficult English words, pronunciation of English, Hindi, and Marathi words, takes too long time to transcript, can also be solved. It can be further expanded for long size of audio files.

Social Impact: This will encourage the Indian developers to focus more on our country languages and server the population of 1.3 billion.

Economic Impact: More competition and new ways and algorithms are always beneficial to the consumers as we know that competition always benefits the consumer and results in the dropping of prices and new features.

VII. ACKNOWLEDGEMENT

We sincerely thank our guide Mr.Anand Khandare for his guidance and constant support
and also for the stick to our backs. We also thank the project coordinators for arranging
the necessary facilities to carry out the project work.

We thank the HOD, Dr. Sheetal Rathi, Dean Academic, Dr. R.R Sedamkar, the Principal,

Dr. B. K. Mishra and the college management for their support.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, “Title of paper if known,” unpublished.

[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

Prediction of Rainfall in Pune City Region Using Deep Neural Networks

Mr.Sanat Moharir, Dr. Nitin Pise
MTech Data Science& Analytics
MIT WPU Pune, India
sanatmohrir@gmail.com, nitin.pise@mitpune.edu.in

Abstract—**Rainfall forecasting is crucial in the field of meteorology and hydrology. precise weather prediction is essential for planning our day-to-day actions. Farmers require the knowledge to help them plan for the farming and harvesting of their crops. The main cause of rain in India is southern monsoon rain which transforms agriculture and flora and fauna. Nevertheless, present solutions always deliver low prediction precision for medium-term rainfall forecasting. Numerical forecasting models operate unsatisfactorily in many situations. This paper proposes a fresh solution to enhance the overall forecasting accuracy for medium-term rainfall. This paper introduces a model which considers important 12 environmental constituents which influence the Southern Monsoon Rain in the Indian subcontinent using the deep learning technique like Multilayer Perceptron(MLP). First, Principal Component Analysis (PCA) is applied to minimize the dimension of twelve physical factors, which serves as the input of MLP. Second, a greedy algorithm is employed to determine the structure of MLP. Finally, the model will provide us with the precise forecasting of the next seven days.**

Keywords— Machine Learning, Rainfall Forecasting, Deep Learning, medium term.

I. INTRODUCTION

As the principal part of water resource, rainfall plays a significant role in the ground of hydrology and meteorology. Generally speaking, rainfall is the completion of multi-scale air system interaction, and it is influenced by many environmental factors, such as thermal power, flow field, and terrain [4]. These intricate physical mechanisms make forecasting rainfall very difficult. Moreover, rainfall forecasting is intimately associated with resident life. Especially, short-time heavy rainfall is the main weather element that effects urban water-logging, and can also prognosticate road water accumulation ahead of time [4]. Consequently, due to the intricate progressive changes inside the atmosphere and the real-time provision of short- term rainfall forecasting, a large-scale and high accuracy forecasting model is crucially needed, which postures a big challenge to the field of meteorology and hydrology. Currently, there are mainly two means to forecast rainfall: atmospheric models and machine learning procedures. Atmospheric models affect atmospheric operation. The atmospheric equation is an exclusive system illustrating atmospheric motion. Atmospheric equations can be utilized to predict atmospheric physical quantities and weather components, including rainfall. Atmospheric models can be farther divided into atmospheric circulation models, climate models and numerical models [4]. At present, numerical models are the principal mode of forecasting medium and short-term rainfall in India. Since atmospheric equalizations are made up of partial differential equations, the overall precision of numerical computation depends on rational approximation. Moreover, atmospheric models usually need a lot of computations and supplies. These stringent requirements for hardware make atmospheric forecasting tough to implement. Machine learning strategies are based on the principles of various machine learning algorithms [1], [2], [6]. Generally speaking, one kind of machine learning strategies are based on statistics, which simply learn the time series fluctuation characteristics from historical rainfall series. Other machine learning methods usually investigate the factors that may influence rainfall and the internal relationship between these factors and rainfall. Because machine learning can learn the intrinsic relation of data, an appropriate machine learning algorithm is the solution to accurately learn the relation between different circumstances and rainfall.

Artificial neural network (ANN) [13] is a set of machine learning algorithm which resembles the human brain, and it has strong learning ability. However, conventional neural networks usually comprise only one hidden layer with the simplistic structure and shallow layers. They tend to befall into the local minimum and over-fitting. Deep Neural Network (DNN) [10], [11], a multi-layer neural network, enhances the most prominent learning network in ANN. Due to the exceptional learning ability, DNN-based procedures have become the most effective machine learning approaches. In recent years, DNN has been applied to the domain of image recognition, speech recognition. and Natural Language Processing. Compared with conventional ANN, DNN has stronger learning ability.

It can learn deep links amongst data, which is confirmed to be an efficient approach for

classification and prediction. Usually, DNN comprises Multilayer Perceptron (MLP) [15], Deep Belief Network (DBN), Long Short-Term Memory (LSTM) [17], Conventional Neural Network (CNN) and so on. At an instant, DNN-based procedures usually take a particular area as the research object and study the association between factors and rainfall. How to select the suitable factors and the learning algorithm is the fundamental step for the successful application of this set of approaches. This paper practices the decision tree algorithm C4.5 and it will be implemented on the model and the

results will be used for the comparison purpose with the results of the multilayer perceptron(MPL)

II. LITERATURE REVIEW:

Mark Holmstrom, Dylan Liu, Christopher Vo [1], In this paper, we will consider the applicability to meteorology Forecast to produce a more accurate weather forecast for a conceivably big time. The range of the paper was restricted to predicting the 7-day highs and lows given meteorological data for the last two days. With a linear regression model alterations of the functional regression model were used, the latter being able to capture the following trends weather. Both of models have outperformed by expert weather forecast service, although the inconsistency between our models and the professional ones has been rapidly abbreviated due to the anticipation of the later days, and possibly because of the long time range, our models have been specialized.

Qing Yi Feng, Ruggero Vasile, Marc Segod, Avi Gozolchiani, Yang Wang, Markus Abel, Shilomo Havlin, Armin Bunde, and Henk A. Dijkstra [2], the author presents a toolbox climatology to tackle climate prediction problems applying machine learning techniques and Climate Network analysis. This package illustrates the basic procedures of data mining, i.e. data reading, merging, cleaning, and mapping of multi-layer artificial neural networks and genetic programming. Such data can be employed as input to machine learning algorithms because the transient information of climate change can be efficiently expressed by intricate network measures.

Timothy DelSole, Michael Tippett, Jyothi Nattala [3], based on this predicted SST, the regression prediction for all Indian monsoon rainfall produced a Gaussian distribution of 95% of the average of the long-term average in 1871-2010 and 10% of the standard deviation of the long-term average in 1871-2010. For the reference, the mean of 1999-2010 was 96% of the long-term average, so the mean of the probabilistic prediction is not common of the latest period. This is an experimental forecast for informational objectives only and is not a replacement for official recommendations, forecasts or warnings advertised by government agencies.

S.R. Kalsi [4], Satellite data is practiced in combination with conventional weather observations in synoptic analysis and traditional weather forecasting to extract information correlated to agriculture in India. The synoptic application of satellite imagery used by the Indian Meteorological Department is highlighted in this paper. In particular, importance is stored on the identification of large-scale convective precipitation methods such as monsoon depressions and tropical cyclones. This paper briefly examines national and international satellite systems, especially those associated with agricultural weather forecasting.

Pengcheng Zhang, Yangyang Jia, Jerry Gao, Wei Song, Hareton Leung[5], the author presented the procedure called Dynamic Regional Combined short-term rainfall Forecasting approach (DRCF) and use the multilayer perceptron deep neural network for calculating short term rainfall forecasting. They conducted the experiment in 57 real locations in China with the accuracy of 87%.

S. monira sumi, m. faisal zaman, hideo hirose [6], the author present, an effort was made to determine an optimal data- driven machine learning approach for predicting average daily and monthly rainfall in Fukuoka city in Japan. This comparative study is concentrated on three aspects: modeling inputs, modeling methods, and preprocessing techniques. The linear correlation analysis and the correlation of the average joint information were carried out, and the optimum input technique was achieved. In order to model rainfall, a unique hybrid multi-model method is introduced and compared with its constitutive model. The model incorporates artificial neural networks, multivariate adaptive regression splines, k-nearest neighbor, and radial basis Support Vector Regression. Each of these methods is implemented to model daily and monthly rainfall in succession with pre-processing procedures, including moving average analysis and Principal Component Analysis.

Main research gap of these paper was, due to the use of different machine learning techniques another then Deep neural network the prediction accuracy of Medium-term weather forecasting was limited.

The research papers related to the Indian context

S.NO	ATTRIBUTE	TYPE	DESCRIPTION
1.	STN	Integer	Station number Of the location
2.	DAY	Integer	Year, month, day
3.	TEMP	Numeric	Mean temperature In F
4.	DEWP	Numeric	Mean dew point In F
5.	SLP	Numeric	Mean sea level pressure In mb
6.	STP	Numeric	Mean station pressure In mb
7.	VISIB	Real	Mean visibility In miles
8.	WDSP	Numeric	Mean wind speed In knots
9.	MXSPD	Numeric	Maximum sustained wind speed In knots
10.	MAX	Numeric	Maximum temperature in F
11.	MIN	Numeric	Minimum temperature In F
12.	PRCP	Binominal	Total precipitation In inches.

did not consider the parameters of Southern Monsoon air for the model building, which is the major reason for monsoon in India.

III. PROPOSED SYSTEM:

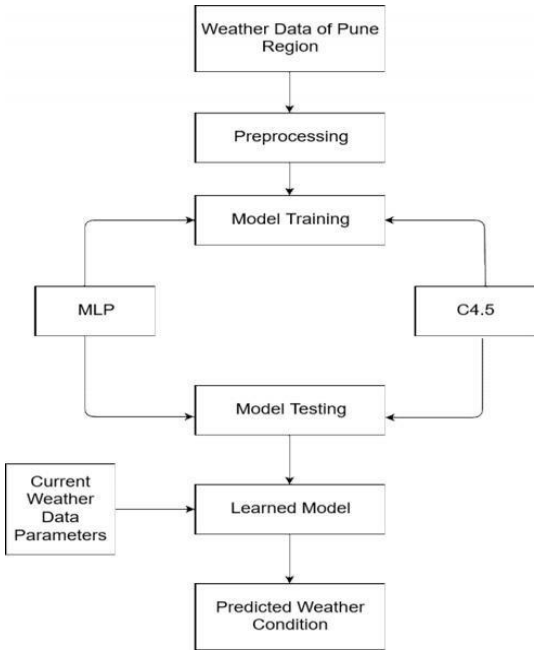


Fig 1: System Architecture

We propose a new model which considers the features of southern monsoon rain. For the model, we are going to consider the following twelve factors which profoundly affect the rainfall.

Table 1: Weather Data Parameters

The historical dataset will be collected from the Metrological Department Pune.
The prediction will be done by using the Multilayer-Perceptron(MLP) algorithm and we will use the C4.5 decision tree algorithm for the comparison of outcome of the proposed model between the two algorithms.

1. Multilayer-Perceptron(MLP):

ANN is an operational model which consists of a huge number of interconnected neurons. Each neuron comprises the specific output function which is called as an Activation Function. The relationship between the two values consists of the weights. Simply weight is the memory of the ANN. DNN is a section of an ANN but has advanced layers. The first layer is called the input layer, and the last one is an output layer and middle layers are called hidden layers and each layer consists of neurons. The neurons in the MLP have the special properties which are intralayer connectionless and interlayer full connection[18]. Fig 2 is a structural diagram of the MLP which consists of two hidden layers. The input layer consists of one neuron N1 its exhibits the dimensions of an input data. The hidden layer 1 consists of neuron N2 and hidden layer 2 consists of neuron N3. The output layer consists of the neuron N4 which shows the dimensions of an output data. Besides all the layers also consist of some bias. First of all Principle Component Analysis(PCA) will be practiced to reduce the dimensions and find out the correct input on an MLP. The greedy approach called Stochastic Gradient Descent(SGD) will be used to determine the feasible structure of a network and its parameters respectively.

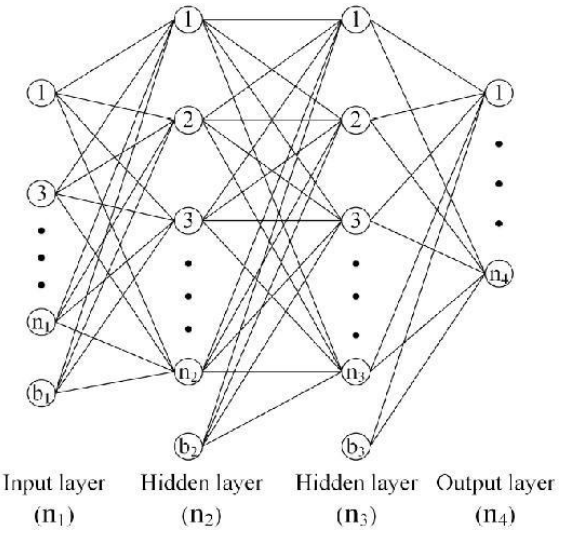


Fig 2: Multilayer Perceptron(MLP)

2. C4.5 decision tree algorithm:

The decision tree is based on conditional probabilities and practices a tree-like graph with

potential consequences, outcomes, costs, events, and utilities. Decision trees are broadly utilized in operations research, marketing analysis, financial drifts specifically in judgment making strategy that assists to reach the goal. C4.5 algorithm will be employed to produce a decision based on data which are both univariate or multivariate it is a statistical classifier based on the ID3 algorithm. It operates on the notion of Information Entropy The training data is a set of classified samples comprising p- dimensional vectors representing the attributes of the sample. C 4.5 creates a decision tree where each node divides the classes based on the accumulation of information. The attribute with the highest normalized knowledge gain is used for the split criteria.

IV. CONCLUSION:

For resolving the shortcomings of current weather forecasting models we introduced a model which considers the factors influencing the rainfall in the Pune region. As the model will use the parameters in the circumstances of southern monsoon rain which causes rain all over India. MLP is used for enhancing the accuracy of the model. We additionally use the C4.5 decision tree algorithm for the result comparison purpose.

The proposed model will outperform the current weather prediction system, providing us with immense accuracy. The model will predict the accurate rainfall of the upcoming 7 days.

V. FUTURE WORK :

Future work concentrates on the following aspects. First, using this methodology for short-term rainfall forecasting which is prediction of weather at the current instance of time. Using the model in the case of drought-affected areas in Maharashtra for assisting the farmers to plan their farming and harvesting.

REFERENCES :

[1] Mark Holmstrom, Dylan Liu, Christopher Vo, “Machine Learning Applied to Weather Forecasting”, Stanford University Press,pp.1-5,2016.

[2] Qing Yi Feng, Ruggero Vasile, Marc Segond, Avi Gozolchiani, Yang Wang, Markus Abel, Shilomo Havlin, Armin Bunde, and Henk A. Dijkstra, "ClimateLearn: A machine-learning approach for climate prediction using network measures", Geosci. Model Dev. Discuss,pp.273- 291,2016.

[3] Timothy DelSole, Michael Tippett, Jyothi Nattala, “ClimateLearn: A machine-learning approach for climate prediction using network measures”, pp.273-291,2013.

[4] S.R. Kalsi, "Satellite-based weather forecasting", IMD, pp.331-346,2014.

[5] Lily Ingsrisawang, Supawadee Ingsriswang, Saisuda Somchit, Prasert Aungsuratana, and Warawut Khantiyanan, “Machine Learning Techniques for Short-Term Rain Forecasting System in the Northeastern Part of Thailand”, World Academy of Science, Engineering and Technology, pp-248-253,2008.

[6] s.monira sumi, m. faisal zaman, hideo hirose, "A rainfall forecasting method using machine learning models and its application to the fukuoka city case", Int. J. Appl. Math. Comput. Sci,pp.841-854,2010.

[7] Fhira Nhita, Deni Saepudin, Adiwijaya, Untari Novia Wisesty, “Comparative Study of Moving Average on Rainfall Time Series Data For RainfallForecasting Based on Evolving Neural Network Classifier” 3rd International Symposium on Computational and Business Intelligence,pp.112-116,2015.

[8] T.R.V.ANANDHARAJAN, ABHISHEK HARIHARAN, “Weather monitoring using artificial intelligence” 2016 International Conference on Computational Intelligence and Networks,pp.106-111,2016.

[9] Yajnaseni Dash, S.K. Mishra, B.K. Panigrahi, “Rainfall Prediction of a Maritime State (Kerala), India using SLFN and Techniques” 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT),pp.1714-1718,2018.

[10] A.Geetha, Dr.G.M.Nasira, "Data Mining for Meteorological Applications: Decision Trees for Modeling Rainfall Prediction" 2014 IEEE International Conference on Computational Intelligence and Computing Research,pp.978-982,2014.

[11] James N. K. Liu, Bavy N. L. Li, and Tharam S. Dillon, “An Improved Naïve Bayesian Classifier Technique Coupled With a Novel Input Solution Method” IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS,pp.249-256,2001.

[12] Kesheng Lu, Lingzhi Wang, “A Novel Nonlinear Combination Model Based on Support Vector Machine for Rainfall Prediction” 2011 Fourth International Joint Conference on Computational Sciences and Optimization,pp.1343-1346,2011.

[13] Xiaoming Pan, Jiansheng Wu, “Bayesian Neural Network Ensemble Model based on Partial Least Squares Regression and Its Application in Rainfall Forecasting” 2009 International Joint Conference on Computational Sciences and Optimization,pp.49-52,2009.

[14] Chandrasegar Thirumalai, M Lakshmi Deepak, “Heuristic Prediction of Rainfall Using Machine Learning Techniques” International Conference on Trends in Electronics and Informatics,pp.1114-1117,2017.

[15] D. R. Sikka, "Two Decades of Medium-Range Weather Forecasting in India: National Centre for Medium-Range Weather Forecasting" COLA Technical Report,pp.276-279, 2009.

[16] Jitcha Shivang, 2S.S Sridhar, "Weather prediction for Indian location using machine learning " IJOPAM,pp.1945-1949,2016.

[17] Ali haidar, “Weather Forecasting using Machine Learning Algorithms ”AJC,pp.208-220,2017.

[18] Binghong Chen, Cong Luo1, Ke Zhang, Xing Shi, Xiaoguang Wang, Minghui Qiu, Wei Chu, Jingren Zhou, “Non-Linear Machine Learning Approach to Short-Term Precipitation Forecasting” World Academy of Science, Engineering and Technology. pp. 248-253, 2013.

[19] Pengcheng Zhang, Yangyang Jia, Jerry Gao, Wei Song. “Short-term Rainfall Forecasting Using Multi-layer Perceptron.” IEEE Transactions on Big Data, 2018.

Driverless system to enhance safety with voice assistance

Pooja B.Bhamare*1, Manjusha A. Adsare*2,, Bhushankumar N. Shinde*4.
Electronics & Telecommunication Department, Savitribai Phule Pune University.
Ganeshkhind, Pune – 411007, India.

ABSTRACT — In this project, a prototype of the small metro capable of showing the station on the LCD display with various other facilities is designed. This metro is run on the DC gear motor. Metro train is auto stop on 3 di erent stations and reverses automatically with Voice Announcement facility. The DC motor is attached with the gear inside to reduce the speed of the vehicle.This project is designed so understand the technology used in the now a days driver less metro train which is used in most of the developed countries like Germany, France, and Japan etc. These trains are equipped with the CPU, which control the train. The train is programmed for the speci c path. Every station on the path is de ned; stoppage timing of the train is prede ned.In this project we try to give the same prototype for this type of trains.We are using arduino micro- controller. The motion of the train is controlled by the DC Motor, for displaying message in the train we are using Intelligent LCD Display of two lines. The train is designed for three sta-tions, named as Aligarh, Ghaziabad New Delhi. RFID card is mounted at each station .When RFID reader mounted on train detects RFID card, train will be get stopped. The Stoppage time is of 3 Sec. There is a LCD display for showing various messages in the train for passengers. Before stopping at station the train blows the buzzer. It also includes an emergency brake sys-tem due to which the train stops as soon as the brakes are applied and resumes journey when the emergency situation is over.
When any obstacle is detected, train will get automatically

Key words — Zigbee and GPS module, RFID, Train, Station, Arduino and LCD.

I. INTRODUCTION

The main aim of this paper is to make an automated place announcement system for Train using voice module and the radio frequency wireless card for tracking the station data. The project consists of microcontroller with the RFID Tag-Reader and the voice recorder chip with speaker.The whole system is attached to the vehicle (Train). The encoded RFID tags are placed in the railway stations. The microcontroller in the TRAIN is programmed in such a way that every station name saved in the voice chip which is having a unique code. So whenever the bus or train reaches the station, the reader in the train receives the codes, which are transmitted from the tag and the microcontroller receives this code and checks in the look up table, saved in the chip. Whichever matches, the controller will send the command to the voice chip to play that particular voice. At the same time the train stops for about 10-15 seconds in the station and then before leaving the station, it will again start to announce PLEASE GET INTO THE TRAIN, THE TRAIN WILL LEAVE IN 6
SEC and the train starts to move to next station. The voice chip will play the voice and this will be heard in the speaker. This voice is repeated till the train leaves the station.

II. MOTIVATION

- City efficiency largely depends upon the effectiveness of its transport systems, that is, efficiency with which people and goods are moved throughout the city.
- This embedded application mainly focuses on overcoming loop holes in the existing system.
- It is optimized to meet the cost and power consumption

III . OBJECTIVES

- To save time
- To save manpower
- To save human life
- Avoid accident

IV . ANALYSIS AND DESIGN APPROACH

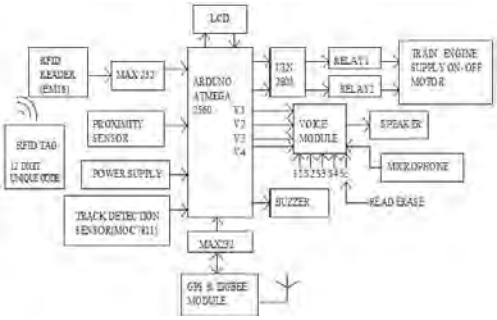


Figure 1: Block Diagram of System

RFID(Radio Frequency Identification)
RFID is Radio Frequency Identification used for discovering the objects. It uses wireless system or electromagnetic elds to transmit the data. It also used for automatically tracing and identifying the tag attach to the object. The tag contains general information about the object and it is the info which kept electrically. Tag can be read from numerous distances from the signal through RFID readers.
Radio-frequency identi cation (RFID) uses electromagnetic fields to automatically identify and

track tags attached to objects. The tags contain electronically stored information. Passive tags collect energy from a nearby RFID reader's interrogating radio waves. Active tags have a local power source such as a battery and may operate at hundreds of meters from the RFID reader. Unlike a barcode, the tag need not be within the line of sight of the reader, so it may be embedded in the tracked object. RFID is one method for Automatic Identi cation and Data Capture (AIDC).
Functioning of RFID

It comprises two parts, a tag and a reader. The RFID tag are implanted with transmitter

& receiver. RFID component on the tags have two parts. In which 1st one is microchip that stores object information and process objects information and 2nd part is antenna to receive and transmit a signal. A tag can superior serial number per each object. This number cannot be used for another RFID tag.

To read the information encoded in a tag two-way radio transmitterreceiver called as mediator that emits the signal to using the antenna. The tag answers with the information which stored in its memory and the mediator will then transmit the read results to an RFID program.

GPS technology is used for tracking vehicle. The Global Positioning System (GPS) is a space-based navigation system that provides location information of moving target can be determined and be transmitted with messaging platform of GSM network.
GPS has become an efficient tool in the field of scientific use, surveillance of objects and tracking. GPS uses Trilateration process to compute the position of an object. In our system, GPS is the key component in getting the vehicles coordinates for tracking.

Track detection sensor

MOC7811 is a slotted Opto isolator module, with an IR transmitter a photodiode mounted on it. Performs Non-Contact Object Sensing. This is normally used as positional sensor switch (limit switch) or as Position



Encoder sensors used to nd position of the wheel. It consists of IR LED and Photodiode mounted facing each other enclosed in plastic body.

V BUZZER
A buzzer is activated to alert the nearby people if the anyone tries to start the vehicle without key.

VI IR PROXIMITY SENSOR

An proximity sensor or IR receiver, is hardware that sends information from an infrared remote control to another device by receiving and decoding signals. In general, the receiver outputs a code to uniquely identify the infrared signal that it receives.

An IR sensor can measure the heat of an object as well as detects the motion. These types of sensors measure only infrared radiation, rather than emitting it that is called as a passive IR sensor. Usually in the infrared spectrum, all the objects radiate some form of thermal radiations.

An infrared sensor is an electronic instrument which is used to sense certain characteristics of its surroundings by either emitting and/or detecting infrared radiation. Infrared sensors are also capable of measuring the heat being emitted by an object and detecting motion.
C . GPS module

Figure 4.15: MOC7811

Microcontroller
A microcontroller is a computing device which usually consists of built-in peripherals. It included CPU, USART(Universal synchronous asynchronous Receiver Transmitter), timers, counters and other components. It is a programmable device which takes decision. It decodes commands received from GPS module and receives angle information from angle measurement sensor. And provides input to the GSM module buzzer, left and right indicator.
ATMega2560 :
The ATmega2560 is a low-power CMOS 8-bit microcontroller based on the AVR enhanced RISC architecture. By executing powerful instructions in a single clock cycle, the ATmega2560 achieves throughputs approaching 1 MIPS per MHz allowing the system designed to optimize power consumption versus processing speed.

Power Supply
The system is powered by battery present in two wheelers which is generally of 12 Volts. IC7805 is used to provide supply regulated voltage of 5 Volts to MICROCONTROLLER, and other units in the system.

Additional Components
Along with the major modules and components, a number of other components are also used in our project.
Some of them are listed below:

- Speaker,microphone
- LEDs, Buzzers, Indicators, Relays
- MAX 232 IC, Voltage Regulators & Other Electronic Components

F . Flow chart

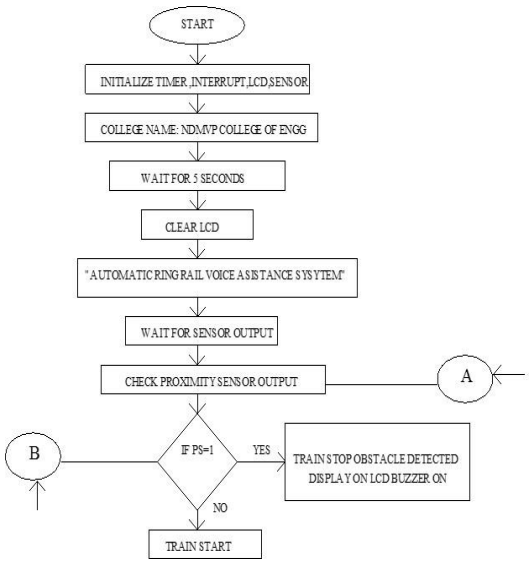
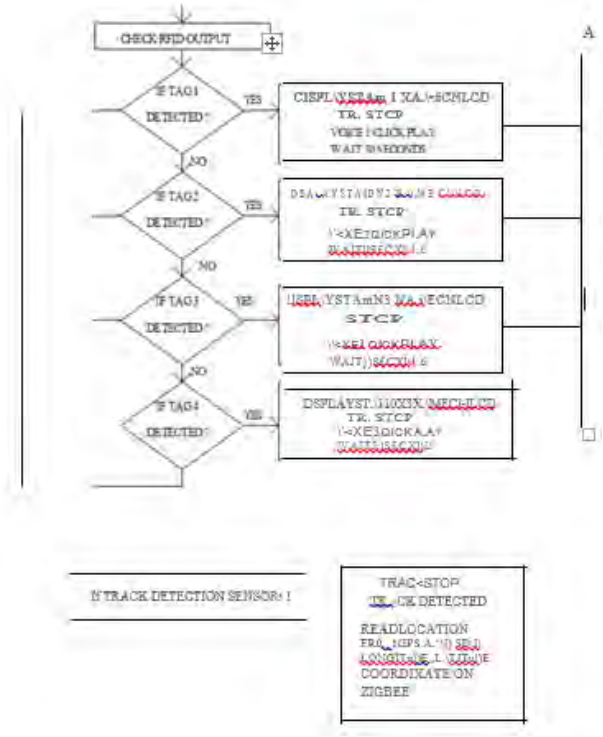


Fig A. Flow chart



VII CONCLUSION

We develop automatic railway system Which increases the efficiency of transport regulation. This automatic system with the help of sensor enhance safety by avoiding accidents and save human life. This system can be used in metro cities to make transport regulation efficient for people. This makes the transportation in city efficient for people.

REFERENCES

[1]Zhixiong Liu, Guiming He, “Research on Vehicle Anti-theft and Alarm System Using Facing Recognition” 2005 International Conference on Neural Networks and Brain.

[2]G.S.Prasanth,Gan esh,B.Balaji,,T.A.Sr inivasaVaradhan “Anti-Theft Tracking System for Automobiles (AutoGSM)”. 2011 IEEE International Conference on Anti- Counterfeiting, Security and Identification.

[3] Hu Jian-ming ,Li Jie,LiGuang-hui “Automobile Anti- theft System Based on GSM and GPS Module” Fifth International Conference on Intelligent Networks and Intelligent Systems 2012.

[4] K. A. Mamun1, Z. Ashraf “Anti-theft vehicle security system with preventive action”. in 2012.

[5] Shruthi.K , Ramaprasad.P , Ruschil Ray , Manjunath A. Naik , ShubhamPansari “Anti-theft vehicle Tracking System with a Smartphone Application” 2015 International Conference on Information Processing (ICIP) 2015.

[6] Md. MarufiRahman, JannatulRobaiatMou , Kusum Tara, Md. Ismail Sarkar “Real Time Google Map and Arduino Based Vehicle Tracking System”. 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE) 8-10 December 2016.

[7] MrinmoyDey*, Md. AkteruzzamanArif and Md. Asif Mahmud “Anti-theft Protection of Vehicle by GSM & GPS with Fingerprint Verification” . International Conference on Electrical, Computer and Communication Engineering (ECCE), February 16-18, 2017

[8] Muhammad Ali Mazidi,Shujen Chen, and EsharGhaemi’s “Atmel ARM programming for embedded systemVolume 5(Mazidi and Naimi ARM series)”.

[9] S.McNamaraY.B.G ianchandani’s “MEMS-Based Sensors”.

[10]ADXL335,smal lpower,3axis,3gravit yAccelerometer datasheet-https://www.sparkfu n.com/datasheets/Co mponents/SMD/adxl 335.pdf.

Implementation of Effective Learning for Adolescents by Digital Solution and Evaluation Process

Shriya Sundriyal, Vaibhav Singh, Amit Sharma, Mr. Shailesh Sangle
Computer Department, Thakur College of Engineering & Technology, Kandivali (East), Mumbai-400101, India
sundriyalshriya@gmail.com, singhvaibhav9815@gmail.com, amitsharma3409@gmail.com,
shailesh.sangle@thakureducation.org

***ABSTRACT* – Analysis of student behavior came to conclusion that student has various problems and it should be solved at early stage. The students of age group between 10-19 years have a lot of difficulty in dealing with problems. As this is a changing phase, there are lot of factors (peer pressure, depression, panic, tension, jealousy, behavior changes etc.) which could affect the mental as well as physical health of adolescents. There are various website which helps to deal with the problems but they helped it temporarily, they lack in lifelong learning. Moreover, at times the adolescents keep on playing games which leads to addiction or they get pissed off playing the games. Our application aims to help the adolescents to deal with problems and to make their life better with help of puzzles, stories, Yoga Sutras etc. It includes various types of activities. In a nutshell, it's a learning platform for adolescents to overcome their problems and provide them lifelong learning.**

Keywords- Rubrics assessment, Yoga Sutras, adolescents, decision- making, peer pressure, NodeJS.

I. INTRODUCTION

The proposed work is is help adolescent over with their problems and help them to improve mental health by Yoga Sutras. As we know, adolescence is the age between adulthood and childhood. Teenagers are often confused about their role and are torn between their responsibilities as growing adults and their desires as children. Key objective is to increase the ability to face situations with full potentiality, which is already there but hidden due to chaos in mind. The main aim is to reduce anxiety levels and scope up with difficult situation easily an make them confident enough to take their own decisions, to know what is right and wrong for them. Understand their roles, responsibilities and lead happy life. They should have a tool, to adjust their physiology, their posture, their breathing, to access a relaxed and focused state for test-taking, anxiety, or problems at home. That tool is nothing but Yoga Sutras. The application will also help teens be able to take a step back before reacting to difficult situations. Students should be able to respond not out of anger, but to develop space, through Yoga for processing trapped energy.

II. LITERATURE SURVEY

Sr.No	Title	Year	Author name	Review
1.	Effect of Yoga on academic performance in relation to stress.	2009	Kauts A. Sharma.	The finding of the study that with the intervention of Yoga, academic performance improves by optimizing the stress levels.
2.	'Risk Factors for Adolescent Academic Achievement.	2001	Magdol L .	Peer factors are identified and objective is to make adolescent to handle all factors in positive ways.
3.	Adolescent Yoga: Goals and settings in Therapeutic Practice.	2017	Roberts Alison	Statically analysis has been done and founded that adolescents who performed yoga shows good performance in studies and improvement in their behaviour.
4.	The Effects of Yoga Practice on Classroom Management in an Elementary School Setting.	2015	Willits, Gail	Quantitative inquiries measured behaviours before and after implementation of the Yoga program.

5.	A Comparative Analysis of Node.js	2015	Nimesh Chhetri	Nodejs is better than others as it has many npm packages which makes the work of developer easy.
6.	Performance evaluation of SQL and NoSQL	2016	Seyyed Aboutorabi , Mehdi Rezapo	MongoDB performs better for most operations excluding some aggregate functions.

III. Proposed Work

Step 1: Identify the Student on which Experiment to be performed:
The Study is delimited to English medium students having age group of 14-15 years The study Delimited to the students of 10 standards only. And the samples are delimited to the particular area.
Judgment Sampling: Non Probability Method
Informal consent:Selection of the students by screening was done by teacher of the school based on their poor academic result.
Inclusion Criteria: Those with poor academic record especially in Science are included from all the divisions of Xth Standard.
Exclusion Criteria**:** Those with any form of physical and mental deformity. Those not willing to practice and study, below 14 and 15 years, irregular attendance for daily session. Sickness during the course.

On the first day of the class, Questionnaire was given to the students to collect following
things:
1) Perception about Yoga i.e. Exercises or something else.
2) About their academic performance
3) Views on Yoga to improve academic performance.
Following Questionnaire is the partial list, was prepared and given to the students.
Actual Questionnaire was prepared in English language.
1) How you self-assess your education? (Average/ Poor /Bright Student)
2) Are you thinking of Yoga can be a used to improve the lifestyle?
(Yes/No/ absolutely not)
3) Is with the help of Yoga, physical, mental and spiritual progress is
Possible?

Step 2: Design the Pre-experiment Questionnaire and test (for students) based on aim of the experiment.

5) Benefits of Āsana only limited to physical Exercise (Yes/No)
6) Teaching Yoga in school is must (Yes /No/ absolutely not)
7) Can u believe that Yoga will help to improve mental power, Concentration?
(Yes/No/Absolutely not)
8) Will you like to practice Yoga along with normal subjects
(Yes/No/Absolutely not)

i] GAME
Adolescents love to play a memory game. So a memory game was conducted where pre-experiment results were noted down. They are able to remember 20 to 50 % of given English spelling. Important of Subjects they study? When this question asked to them, 50 % of students were in an opinion that all subjects are not worth studying it.

ii] SUMMARY OF PRE-EXPERIMENTAL STUDY

(Yes/No/Absolutely Not)
4) Yoga means only Physical Exercise (Yes /No)

Based on the answer got from all students, the facts about their misconception about the Yoga are realized. Following is the summary. Myth about Yoga – Another form of physical exercise that they do as part of PT. One of the social stigmas in this school system is the student's perception of themselves in relation to their academic section. This can affect the developing self-concept of an adolescent, positively or adversely. Students with low self-esteem, they did not see themselves as learners, struggled to pay attention, and gave up before attempting completing a difficult task Negative feeling about the subjectsNegative emotional Response towards Test, task etc.

Methodology:
Methodology to be used for building the project will be Agile Software Methodology as our priority is to satisfy the customer through early and continuous delivery of software. Users feedback is very important for our project. Agile would

help us to break down project into smaller tasks and prioritizing them.

IV. SYSTEM ARCHITECTURE

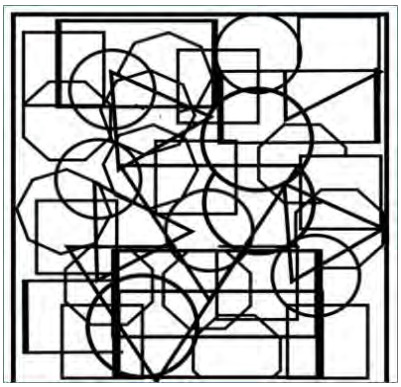


Fig 1 : activity sample

How many rectangles are there in diagram?
a. >10
b. 6-10
c. 1-5
d. None

As depicted from above question, four criteria have been provided. First three criteria represent range of quantities whereas last criterion indicates a single value none i.e. zero. Based on the alternatives selected by student, a quantitative assessment of Yoga can be done using Rubric as to how many students have opted for which alternative and what does it indicate. Each alternative is assigned a point and grade score for students can be calculated using the formula for same Now, the students are explained the logic for this game. That one or the other edge of the rectangle is formed by edge of the adjacent shape. Students are made aware that they need to be attentive.

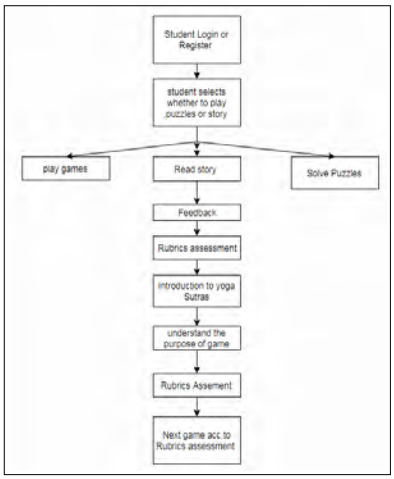


Fig 2: System architecture

Feasibility study:

1. Technical feasibility: Our aim is to achieve student friendly application using Node JS, MongoDB and Rubrics assessment.
2. Hardware required: PC for development, server for deployment.
3. Economical Feasibility: The EFS is composed of two required forms The Business Case provides an analysis of the business environment including- Expected customer: Students
4. The nature of the business: Social
5. Legal feasibility: Only open source data/ images with no copyright issues will be used.
6. Operational feasibility: This will be used in schools to help students overcome their problems.
7. Scheduling feasibility: Fully working Effective learning application for adolescents by 29th February 2019.

Requirements:
Functional
Requirements

Software Requirement:-

3.1.1.1 Front-end:HTML, CSS, Javascript , Ajax
3.1.1.2 Back-end:NodeJS, ExpressJS , MongoDB
3.1.1.3 Report
Generation
Hardware
Requirement:-
3.1.1.7 Intel® Core ™ i5-6500 CPU @3.2 Ghz
4.00 GB RAM 64-bit OS.
3.1.1.8.A Monitor, Keyboard ,
mouse Non Functional Requirements
3.1.2.1 Better Usability by providing simple instructions and user friendly interface
3.1.2.2 More efficient than existing system.
3.1.2.3 High performance

Database Structure:
There will be Users table, user report and session table.
Users table consists of id, name, email_id, password, number, gender, birthday and school name.
Users report consists of id problem, score, quality parameters, feedback.



Fig 3: Users



Fig 4. User Report

V. EXPERIMENT RESULTS

● Graphs before experiments:

This graph shows the resistance of students before conducting the activities. The graph is plotted based on the answers from the students

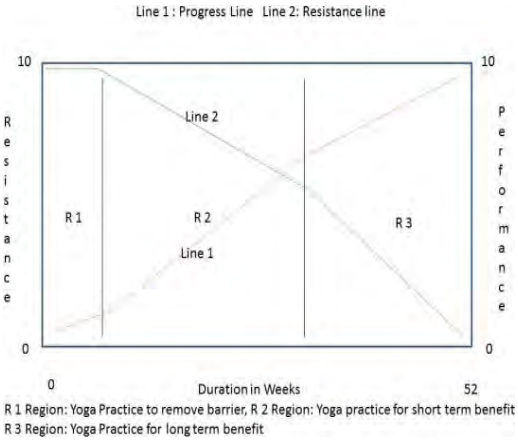


Fig 5 : Graphs before experiments

Comparison Study:

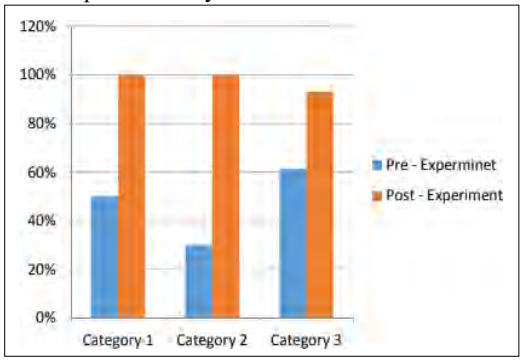


Fig 6: Comparison Study

Category 1: Number of Students wants to practice Yoga along with their study Pre – Experiment: 50 % students don't want to practice Yoga along with the study. Post – Experiment: 100

% Students wants to practice Yoga along with the study. Category 2: Myth about Yoga: Yoga is only physical exercise, no effect in Memory and mind development. Pre Experiment: 30 % students agree that Yoga is usefulfor mind and memory development. Post –Experiment: 100 % students agree that Yoga is useful for mine and memory development. These results only came after their personal experience of yoga and also improved academic performance.

Category 3: Memory Game: Pre- Experiment: An average 61.36 % English spelling (simple – three character) were remember by students Post-Experiment: An average 93.18 % English spelling were remembered by students. Most significant achievement of this experiment is an improved academic performance of experimental group.

VI. RESULTS AND DISCUSSION

The application main aim is to reduce anxiety levels and scope up with difficult situation easily an make them confident enough to take their own decisions, to know what is right and wrong for them. Understand their roles, responsibilities and lead happy life. They should have a tool, to adjust their physiology, their posture, their breathing, to access a relaxed and focused state for test-taking, anxiety, or problems at home. That tool is nothing but Yoga Sutras. The application will also help teens be able to take a step back before reacting to difficult situations. Students should be able to respond not out of anger, but to develop space, through Yoga for processing trapped energy.

VII. CONCLUSION

On daily basis, teenagers face many problems during the most awkward growth stages of their lives (i.e. 13-19 years) adolescence. During this phase, teenagers are exposed to some overwhelming external and internal struggles and are expected to cope with hormonal changes, puberty, social and parental forces etc. However, there are various websites which helps them to deal with their problems. Websites like happify, teen mental_ health.org etc. There are certain limitations in these websites. At times, adolescents keep on playing games which leads to game addiction and lack in lifelong learning. Some websites are not interactive to keep the teenagers busy. Our project aims to help the teenager to deal with their problems and make their life better with help of Yoga Sutras and various activities. Our aim is that after completing all the stages of our application teenager should be self-sufficient enough to take his/her decisions and able to solve their problems easily. The application would make adolescents mentally strong to cope up with stress, improve their decision-making skills. Our first objective is to identify key problems faced during adolescence and the need for removing negativity in their mind. We aim to design and develop a methodology to implement framework and evaluate the efficiency of yoga as well as website. The application work not only on short term benefit for adolescent but to create lifelong learning capabilities. We will develop a phase implementation plan comprising of games, asanas, breathing techniques, stories and puzzles on digital platform.

VIII. FUTURE SCOPE

This idea can be further extended for adults too. We can further keep the track of users typing speed, voice tone, word choice etc to improve the efficiency of our project. As with

help of changes in typing speed we can identify their mood swings and various aspects.

IX. ACKNOWLEDGEMENT

We sincerely thank our guide Mr. Shailesh Sangle for his guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out project work. We thank Dr. Deven Shah for support and advice. We thank the HOD, Dr. Sheetal Rathi, Dean Academic, Dr. R.R Sedamkar, the Principal, Dr. B. K. Mishra and the college management for their support.

REFERENCES

[1] Mending our future: starting a conversation about the support of student mental health. Paul Jenkins, May 2016

[2] The Effects of Yoga Practice on Classroom Management in and Elementary School Setting Willits, Gail, 2015

[3] Risk Factors for Adolescent Academic Achievement, Madgul L ,2001

[4] A Comparative Analysis of Node.js, Nimish Chhetri,2016

[5] Performance evaluation of SQL and NoSQL, Sayed Hamid Abortoria , Mehdi Rezapour,2015

[6] Adele Faber and Elaine Mazlish. How to talk so Teens will listen and listen so Teens will Talk. Happer Collins.

[7] Patanjali Yoga Sutra, Swami Vivekananda.

[8] Rubrics: https://www.nvcc.edu/assessment/_ docs/ FTW5.usingrubricsmeasurestuperfspr13.pdf

[9] Mental-health-and-wellbeing: http://www.learning.schools.nsw.ed u.au/crossroads/supportmaterials/teaching-andlearningresources/mental-healthand-wellbeing- resources/.

[10] Yoga Education for Children Volume 1, Swami Satyananda Saraswati. (2009). [12] Yoga for the Youth. Abstract-16 Kaivalyadhama, Seventh International Conference. (2012).

[11] The Adolescent:Development, Relationships, And Culture, Pearson.

[12] Developing_Games_for_Mental_He a lth_A_Primer:https://www.researchg ate.n et/publication/30 1940610_Developing_Games_for_ Menta l_Health_A_Primer

[13] Rubrics: https://www.nvcc.edu/assessment/_ docs/ FTW5.usingrubricsmeasurestuperfspr13.pdf

[14] Mental-health-and-wellbeing: http://www.learning.schools.nsw.ed u.au/crossroads/supportmaterials/teaching-andlearningresources/mental-healthand-wellbeing- resources/.

[15] Yoga Education for Children Volume 1, Swami Satyananda Saraswati. (2009). [12] Yoga for the Youth. Abstract-16 Kaivalyadhama, Seventh International Conference. (2012).

[16] The Adolescent:Development, Relationships, And Culture, Pearson.

[17] Developing_Games_for_Mental_He alth_A_Primer:https://www.research gate.net/publication/30 1940610_Developing_Games_for_ Menta l_Health_A_Primer

Design and Development of Language Translation and Detection System using Machine Learning

Aishwarya Rajeshkumar Verma
Computer Engineering
Thakur College of Engineering and Technology
Mumbai, India
aishwaryaverma.2015@gmail.com

Abstract— **Words can be more expressive in case of explanation and study. In this multilingual world, these concepts can take place through language translation, which plays a significant role in text processing application like information extraction, machine learning, natural language understanding, information retrieval andmachine translation.**

There are many challenges and issues in translation between languages, such as ambiguities, lexical divergence, syntactical and lexical mistakes and semantic issues, etc.; which are in the grammar, spellings, punctuations, spaces, etc. There are many algorithms which are used for the language translations. However, using various natural language processing algorithms does not provide the proper language translations. Therefore, in this project we are going to implement an algorithm which uses the concept of neural networks along with Natural processing algorithms which will be the combination of Phrase-Based Translation(PBT) and Hierarchy Phrase-Based Translation(HPBT) with N-gram model, so that it makes the language translation more effective and it makes the translated language more qualitative than any other algorithms translated languages. This quality can be determined in terms of grammar, spellings, abbre viations, punctuations and the most important, one must able to read and understand the translatedlanguage.

We are going to compare and analyze our new algorithm with the other algorithms through the quality of translated text, Principal Component Analysis (PCA) and Analysis of Variance (ANOVA).

Keywords—multilingual, information extraction, machine learning, information retrieval, machine translation, ambiguities, lexical divergence, syntactical and lexical mismatches, Natural processing algorithm, Phrase-Based Translation (PBT), Hierarchy Phrase-Based Translation (HPBT), N-gram model, Principal Component Analysis (PCA), Analysis of Variance (ANOVA)

I. INTRODUCTION

There are many cases where we need to find the exact meaning for the word and/or sentence for which we try to use translator for each time. But, at times it may happen that the translator does not translate it correctly due to which it leads to improper knowledge transfer by the translator. This includes incorrect abbreviations, grammar with tenses, meanings, punctuations etc.

There are some situations in the rural area where the people are unable to understand while filling the important form. This can demotivate people for finding the 3rd person who know that particular language very well for each time for filling the form.

We are going to work on implementing a new algorithm, which will be the combination of neural networks and natural processing language algorithm. This combination will not only help in improving the translation, but it will take care of the quality of the language translated such that the main meaning of the word and/or sentences does not get changed.

We are also going to compare how this algorithm is better than the other translator algorithms using only natural processing algorithms and how the quality get better while translating from one language to another.

II. LITERATURE SURVEY

There are various works and study conducted for the language translation to make the translation more effective.

A very few researches have done on English to Bengali language translation both in Bangladesh and West Bengal of India. They have considered total 12 tenses such as - present indefinite, continuous, perfect, perfect continuous; past indefinite, continuous, perfect, perfect continuous; future indefinite, continuous, perfect and perfect continuous are used for the purpose of translating English sentence into Bengali. This helps to find the meaning of the sentence/word from the dictionary, preloaded from the database. Using AI, The NLP algorithm are in the form of Cockey-Younger-Kasami (CYK) along with Chomsky Normal Form (CNF). Morphological analysis is done where morphemes means minimal unit of meaning of grammatical analysis. A phrasal Example Based Machine Translation (EBMT) along with Adaptive rule based machine translation between English to Bengali helps in the actual design of a phonetics lexicon for transliteration. This includes word to word translation, transfer approach, direct approach, Corpus-based approach, Interlingua approach and Statistical Machine Translation. The translated sentences from the approaches mentioned are then compared and analysis are made, which were compared with Google translator and it was found to be the better translator than Google translator[4]. Using SMT with transfer approach, the translation from English to Malayalam. The input English sentence is tokenized, and parsed with the help of Stanford parser, and a parse tree is generated. This is used for translation in Cricket domain[9].

There is some translation algorithm which uses Hybrid based approach. The machine translation is getting challenged day by day; therefore, the methods can be

implemented in a new way.

In one translator, there is acombination approach of phrase based Statistical Machine Translation (STM), Exa mple-Based MT (EBTM), Rule- Based MT (RBMT). This proposed Hybrid Machine Translation (HTM) works better, when compared with Google, BING and Babylonian[6]. In another translator, transfer approach and Exa mple-Based Machine Translation (EBTM) technique are used as hybrid approach, which have delivered greater quality and functionality while translating Malayalam to English sentence[7].

The quality of translation depends on the meaning of the target text. The one should be able to understand the meaning of the bilingual te xts on source and target selected. As per Hübner[12], there were tests conducted in which the students were checked whether they would able to translate German to English. To check the quality of the translation, Contrastive analysis and translation (CAT) were used in which, with the help of the concept of hypothesis, the assumptions were made. On the basis of assumption, the translation and the quality of the grammar were checked[12].

On proper studies, the following are the difference between phrase-based translation (PBT) and hierarchy phrased based translation (HPBT):

TABLE I. Difference between PBTand HPBT

PBT	***HPBT***
The basic unit of translation is becomes heavily on training data	The unit of translation is sub-phrase, which removes the problem of PBT
It fails to handle sparseness of data and provide translation for longer phrases	It can handle sparseness ofdata and provide translation for longer phrases
It is required to capture translations that are relevant beyond the standard three word phrase, but cannot translate longer phrases and need special structure	For longer and lengthy translation, it does not need a special structure as it creates the sub-phrases and get divided within it.

III. METHODOLOGY

Language translation from the source text to the target text are really important with respect to meaning. The meaning of the target text can also get affected if the grammar and tenses are not considered properly.

Therefore, in order to take care of every perspective of word/sentences translation, it is really important that the meaning of the target text should not get changed.

The following are the methodologies, which will help the exact working of the project:

1. In this project, it is a kind of hybrid approach as the algorithm will be the combination of neural networks and n-gram model. There may be possibility that the other approaches of the NLP can be used to enhance our algorithm.

2. In this algorithm, it will help to translate from the source text language to the target text, in the form of sentences, interrogation, exc lamation, etc; taking care of the grammar, tenses and the exact meaning of the conversion via 3rd language, which can be English/Hindi, using the concept of neural networks.

3. The algorithm used for the translation of source text into intermediate te xt is Phrase-based translation (PBT) system.

4. For the conversion of intermediate language text to the target text, the algorithm used here is Hierarchy phrase-based translation (HPBT) system.

5. To make it more efficient, we are going to take care of the special characters such as punctuation marks, brackets, decimals, spaces etc., while translating from one language to another, also considering the 3rd language for better understanding. Here, the ambiguities will be removed if the source language contains the unwanted/incomplete brackets, spaces etc., using the concept of n-gram model.

6. Abbreviations and their full form will also be taken care, such that it can translate the important abbreviations included in the source text into the target text for better understanding, if possible, the full form will also be provided the same, using neural networks and n-gram model.

7. For phrases and idioms evaluation, the exact meaning of the phrases and idioms which will be used as a source text and the target text, which will be the normal sentence that will describe the meaning of the source text given. Note that, the target text will be displayed on the language selected by the user, including the 3rd language (English/Hindi).

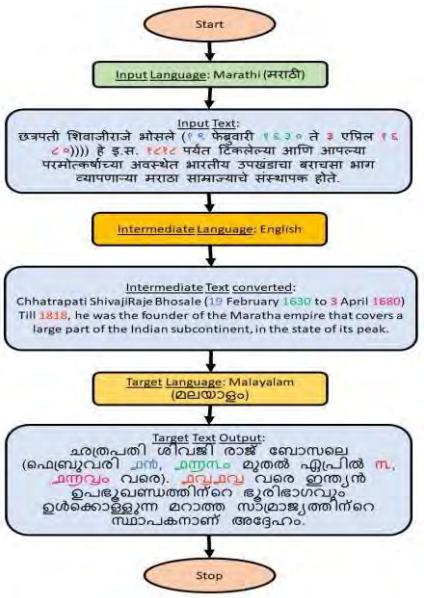
8. The most important thing is the evaluation of the quality of the target text language that are translated from the required source language, along with the 3 rd language translation check. From this quality check, our new proposed algorithm will be compared with other proposed algorithms used as translator. Here, we will use Principle Component Analysis (PCA) and Analysis of Variance (ANOVA).

IV. EXPECTED OUTCOMES

A. Working of the language translation

- We need to enter the required input language along with the paragraph for translation.
- We also need to select the intermediate language: either English/Hindi.
- We need to enter the required output language that we need for translation.

Fig. I: Actual working ofthe language translation



B. Quality check and comparison of the algorithm

To check the quality of the translated language text in the target text, we are going to consider the following parameters:

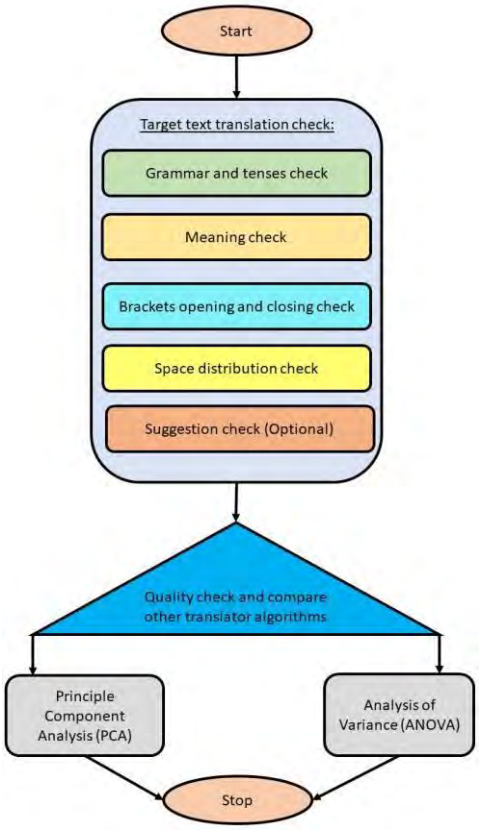
1. Grammar
2. Special characters (Punctuations, complete single/double quotations, brackets, space, etc.)
3. Meaning and tense of the statement/word
4. Abbreviation along with the full form (if possible/required)
5. Phrases and idioms only meaning in the target text along with the meaning in 3rd language.

This quality check will be done using the following models:

1. Principle Component Analysis (PCA)
2. Analysis of Variance (ANOVA)

The other algorithms are going to be compared based on the time complexity along with their proper accuracy for the translation.

Fig. II: Q uality check forthe language translation



V. FUTURE SCOPE

After the successful language translation, this language translation can be expected in the form speech-to-text and text-to-speech, in which we will be speaking on the translator, the translator will process and detect the language spoken and it will convert into the target language, and the machine will speak the target language of the translated text, within less time than other translator.

REFERENCES

[1] Kehai Chen, Tiejun Zhao, Muyun Yang, Lemao Liu, Akihiro Tamura, Rui Wang, Masao Utiyama, and Eiichiro Sumita, "A Neural Approach to Source Dependence Based Context Model for Statistical Machine Translation", IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 26, NO. 2, FEBRUARY 2018

[2] Richard Zens, Franz Josef Och, and Hermann Ney, "Phrase-Based Statistical Machine Translation", M. Jarke et al. (Eds.): KI 2002, LNAI 2479, pp. 18–32, Springer-Verlag Berlin Heidelberg 2002, https://www-i6.informatik.rwth-aachen.de/publications/download/434/ZensR.OchF.J.NeyH.--Phrase-BasedStatisticalMachineTranslation--2002.pdf

[3] David Chiang, "A Hierarchical Phrase-Based Model for Statistical Machine Translation", Proceedings ofthe 43rd Annual Meeting of the ACL, pages 263–270, Ann Arbor, June 2005, Association for Computational Linguistics, http://www.aclweb.org/anthology/P05 - 1033

[4] Hierarchical phrase based Machine Translation: Literature survey, http://www.cfilt.iitb.ac.in/resources/surveys/Hierarchical_MT_Bibek.pdf

[5] Chandramma, Dr. Piyush kumar Pareek, Swathi K, Puneet Shetteppanavar, " An Efficient Machine Translation Model forDravidian Language", 2nd IEEE International Conference On Recent Trends in Electronics Information & Communication Technology (RTEICT), May 19-20, 2017, India

[6] Shamsun Nahar, Mohammad Nurul Huda, Md. Nur-E-Arefin, Mohammad, "Evaluation of Machine Translation Approaches to Translate English to Bengali", 20th International Conference of Computer and Information Technology (ICCIT), 22-24 December, 2017

[7] Sindhu D.V, Sagar B M, " Study on Machine translation approaches for Indian languages and their challenges", International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), 2016

[8] Omkar Dhariya, Shrikant Malviya, Uma Shanker Tiwary, " A Hybrid Approach For Hindi-English Machine Translation", 31st International Conference on Information Networking (ICOIN-2017), 6 Feb 2017

[9] Rosna P Haroon, Shaharban T A, "MALAYALAM MACHINE T RANSLATION USING HYBRID APPROACH", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016

[10] Xinlian Hu, " Statistical Machine Translation Approach Based on Dependency-to-String Model", International Conference on Smart Grid and Electrical Automation, 2016

[11] Aasha V C, Amal Ganesh, "Machine Translation from English to Malayalam Using Transfer Approach", International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2015

[12] Brenda Reyes Ayala, Jiangping Chen, "A Machine Learning Approach to Evaluating Translation Quality", ACM/IEEE Joint Conference on Digital Libraries (JCDL), 2017

[13] Stephen Hampshire, Carmen Porta Salvia, "Translation and the Internet: Evaluating the Quality of Free Online Machine Translators", Quaderns. Rev. trad. 17, 2010, https://core.ac.uk/download/pdf/13292249.pdf

[14] Carol Ebbert-Hübner, Clare Maas, "Can Translation Improve EFL Students' Grammatical Accuracy?", International Journal of English Language & Translation Studies, ISSN: 2308-5460, 5(4). 191-202, 2017

[15] Marziyeh Taleghani, Ehsan Pazouki, " Free Online Translators: A Comparative Assessment in Terms of Idioms and Phrasal Verbs", 6(1). 15-19, 2018

[16] Jorge Leiva Rojo, "Phraseology as indicator for translation quality assessment of museum texts: A corpus-based analysis", Cogent Arts & Humanities (2018), 5: 1442116, https://doi.org/10.1080/23311983.2018.1442116

[17] Gert De Sutter, Bert Cappelle, Orphée De Clercq, Rudy Loock, Koen Plevoets, "Towards a corpus-based, statistical approach to translation quality: Measuring and visualizing linguistic deviance in student translations", Linguistica Antverpiensia, New Series: Themes in Translation Studies, 16, 25–39, 2017

Improved System For Appointment Scheduling In Hospitals

Rohit Ragde , Disha Yadav , Sudeet Mhatre , Mrs. Rashmi Thakur(Guide)
Department of Computer Engineering TCET

***Abstract* – This system aims to develop a computer-based appointment system, which would integrate with the existing clinical information system and help improve the patient access to the services offered by both General Practioners (GPs) and Allied Health Professional (AHPs) in the primary health care centres. To achieve this requires the development of an understanding of the features and processes of the patient booking in the primary health care environment. The aim of this system is to provide patient full access to manage there hospital appointment which facilate with there online service for appointment reservation,updating and canceling the appointment and meeting all the customer requirements.This system will provide an efficient and effective communication between patient –doctor and providing the feasible solution to there problems.**

I. INTRODUCTION

Health care is a fast-growing industry in India as well as abroad. Appointment scheduling is one of the key processes in this industry. This system focuses on online appointment scheduling for all primary healthcare centres. Hospitals can come on board this platform and provide their appointment slots for online booking of patients. The system facilitates

hospitals to easily manage their registration and appointment process and monitor the flow of patients. Scheduling an appointment online is thus an effective and a

lot less costly way for providers to communicate with their patients. Thus, Online Scheduling helps both the parties. This system majorly focuses on patient, Doctor and Hospitals who will use the system to make and manage the appointment online. This system will be helpful for patients, Doctors and the Hospitals.

II. SYSTEM ARCHITECTURE ALGORITHM

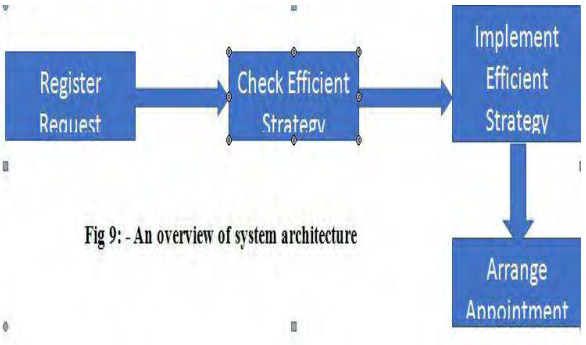
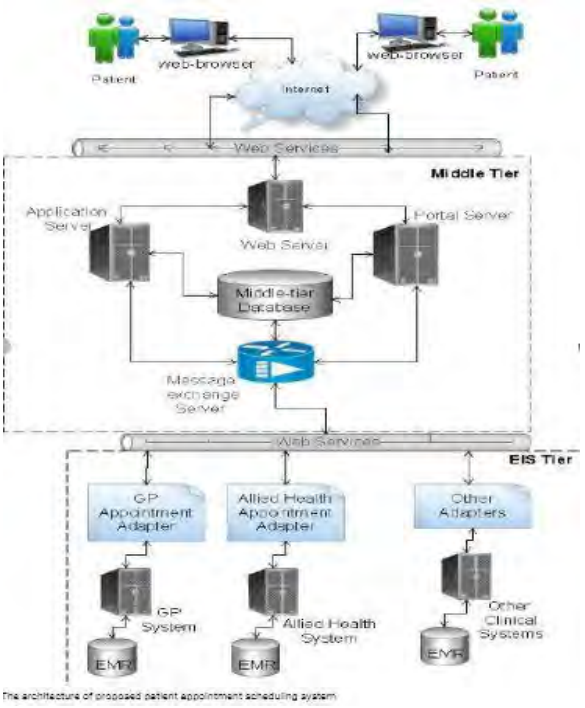


Fig 9: - An overview of system architecture



The architecture of proposed patient appointment scheduling system

Flowchart

Literature survey

III. FEASIBILITY STUDY (FAD) ECONOMIC FEASIBILITY

The EFS is composed of two required forms:

IV. BUSINESS CASE

The Business Case provides an analysis of the business environment including: -

Expected customers: Patients and doctors.

The nature of the business: Social

It also depicts the benefits of the proposed project.

V. COST BENEFIT ANALYSIS

The cost benefit analysis summarizes the revenues and costs involved with the proposed project. As the proposed system will be used for the benefits of the hospital, no additional cost will be paid by them. No hardware system is included in our project, so the hardware cost gets minimized. Only minimal amount will be required by certain softwires.
Hence, our system is cost efficient.

VI. TECHNICAL FEASIBILITY

Technology at the front end: HTML, Bootstrap, CSS.

Technology used at Backend: PHP, MySQL

Resources Required: Manpower, programmers, testers, debuggers.Software required: Xampp server, software for web portal, Adobe Dreamweaver, testing tools.

VII. MANAGERIAL FEASIBILITY

Managerial support, staff involvement and commitment are the key elements required to gauge the managerial feasibility in the proposed project. The success and the profitability of the project partly depend on managerial competence of the major ingredients of the proposed system which are the users i.e. doctors, patients and benefiters i.e. both and entire hospital.

VIII. OPERATIONAL FEASIBILITY

The proposed system is concerned with the zero-income people that is especially for the patients. At initial phase the scope of the system will be at local level which will be helpful for the locals. The portal will be handled by the central manager who is responsible for the registering and maintaining data of those people.

The end users of the system will be patients and the doctors. The important stakeholder of the system will be the admin. They would be responsible for managing the entire system who has all the rights with no restrictions. Providing the access to the patients and the doctors is the key process of the system which will be easily handled by an automated part of the system.

If the usage of the system is very beneficial and useful to locals. Then, the system can be expanded at the global level also.

IX. CONCLUSION AND FUTURE SCOPE

One of the prime reason that online patient scheduling is gaining popularity in recent days is that the system provide an easier for receiving healthcare facilities to the general user.The web service architecture will provide an appropriate paradigm for developing this integrated healthcare system.The system is design to achieve maximum user satisfaction and providing and efficient and effective appointment of patient in hospital by considering all critical solution and providing an feasibl solution.

REFERENCES

[1] Fatma Poni Mariah, Murshid Hasan Basra, “The Analysis of Appointment System to Reduce Outpatient Waiting Time at Indonesia’s Public Hospital”, Human Resource Management Research, 3(1) (2013), 27-33.

[2] Wenjun Cao et.al, “A web-based appointment system to reduce waiting for outpatients: A retrospective study”, BMC Health Services Research, vol 11, 2011

[3] Lucille Keir, Barbara A. Wise, Connie Krebs, Medical Assisting: Essentials of Administrative and Clinical Competencies, Cenage Learning pp.131-133

[4] Xiaojun Zhang “Developing an Online Patient Appointment Scheduling System Based on Web Services Architecture”, APAMI Conference Proceedings, 2012.

Smart Dustbin for Segregation of Wet and Dry Waste

Sagar Pathare, Annanya Sah, Saketsingh Sisodiya, Shiwani Gupta
Computer Department, Thakur College of Engineering & Technology, Kandivali (East), Mumbai-400101, India
sspathare97@gmail.com, annanya777@gmail.com, saketsinghb8@gmail.com, shiwanigupta@thakureducation.org

***ABSTRACT*- The idea is to design and develop a smart dustbin that can segregate wet and dry waste into separate compartments. Smart dustbin can capture the image of the waste, send it for processing and classification. Based on the classification result, it can segregate the waste into respective compartments. The main aim of smart dustbin is segregation of wet and dry waste. This system is primarily based on Convolutional Neural Network (CNN) in which there are several hidden convolutional layers and it is trained on the wet and dry waste image dataset. Currently, segregation is manual for the most part, thereby slowing down the process and also affecting the health of the labour segregating the waste. Random dumping of the waste creates toxic components and pollutes not only lands but the water bodies too. Therefore, segregation of waste is crucial for a healthy and pollution free environment. The proposed system solves this problem. Manual efforts of classification of wet and dry waste are eliminated. Smart dustbin automatically classifies and segregates the waste into separate compartments. The first prototype image classification model showed an accuracy of 80%.**

Keywords - Waste segregation, Wet waste, Dry waste, Image classification, Convolutional Neural Network (CNN), Raspberry Pi

I. INTRODUCTION

The proposed work is to develop smart dustbin for segregation of wet and dry waste. This system is primarily based on Convolutional Neural Network (CNN) in which there are several hidden convolutional layers and it is trained on the wet and dry waste image dataset. This is to create an automatic system which will be able to classify and segregate the wet and dry waste. The classification of waste will be entirely dependent on image processing using CNN. The CNN picks the waste patterns and eventually becomes sophisticated enough to differentiate the wet and dry waste. Since this project is entirely based on image recognition and detection, each convolution layer consists of several filters that help make the pattern detection more potent. This CNN is deployed on the Raspberry Pi. As soon as the waste is kept on the platform of the smart dustbin, the image of the waste is captured using a camera and the classification begins. Based on the classification results, Raspberry Pi sends rotation signals to the rotor mechanism and the waste kept on the platform is dropped into the respective compartment.

II. THEORY

There are a number of key factors to consider when capturing image of waste kept on the dustbin platform using camera, classifying it into wet or dry waste and then dropping it to respective compartment. Some of the main ones are listed below- The Cost Benefit Analysis summarizes the revenues and costs involved with the proposed project. As the proposed system will be used for the benefits of the government and small to large industries, no additional cost will be paid by them. Minimum hardware system is included in our project, so the hardware cost gets minimized. Minimal amount will be required by certain software. Hence, our system is Cost Efficient. Technology used at front end is Raspberry Pi board and other hardware components. Technology used at Back end is Convolutional Neural Networks. For scaling up the project, larger hardware components will be required. The key elements required for managerial feasibility in the proposed project are management support, employee involvement and commitment. Even if the project is found to be economically, technically and financially feasible it can still fail if the users, i.e. the companies and benefiters don't find the management capable. It is important that all ingredients of efficient management are taken into consideration in the proposed project as the success and the profitability of the projects partly depend on the managerial competences. The proposed system is concerned with small to large industries dumping huge amount of non-segregated waste. Initially the system will only be used for small amount of waste segregating at a time. The team will handle the management responsibilities, including the maintaining of the customers/users logs. The end user of the system will be industries (initially local level) or the government (global level). Spreading awareness regarding the fast, easy and automated waste segregation is the key process for the success of Smart Dustbin. Once the system is being used and recognized, the usage can expand up to global level. The end system will help in abolishment of manual waste segregation and decreasing the dumping fields.

In Classification of Trash for Recyclability Status, Mindy Yand and Gary Thung [1] have developed models based on computer vision to classify garbage into recycling categories. This project makes an attempt to take input images of recycling garbage and classify it into one of the following classes- glass, paper, metal, plastic, cardboard, and trash. They have also created an image dataset containing around 500 hand collected images per recycling category class. This project used following models- support vector machines (SVM) with scale-invariant feature transform (SIFT) features and a convolutional neural network (CNN). In Waste sorting - A look at the separation and sorting techniques in today's European market, WMW [2] takes a look at the wide range of separation and sorting techniques and also looks at the issues responsible for driving the new technological development. The common waste sorting methods are: Trommel separators/drum

screens, Eddy current separator, Induction sorting, Near infrared sensors (NIR), X-ray

technology. But manual sorting of waste is still the most common technique used worldwide. Comparing deep learning and support vector machines for autonomous waste sorting, this research [3] tries to automate the waste sorting process using machine learning techniques for waste type recognition from waste images. It uses the two popular learning algorithms: deep learning (DL) with convolution neural networks (CNN) and support vector machines (SVM). Each of algorithms classify the waste into 3 different categories like plastic, paper and metal using coloured waste image. In ImageNet Classification with Deep Convolutional Neural Networks, Krizhevsky, Sutskever and Hinton, [4] have trained a deep large convolutional neural network to classify the one million high- resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. The neural network with 60 million parameters and 650,000 neurons and consisted of five convolutional layers.

III. PROJECT WORK

For training the CNN, it is necessary to create an image dataset containing good number and variety of waste images. The waste is to be classified into two types- wet and dry, so we need at least 500-1000 images of each of these two classes. The CNN works by extracting the hidden features in the images, so it is necessary to gather variety of images. Following are some samples of collected images.



Fig.1: Wet Waste

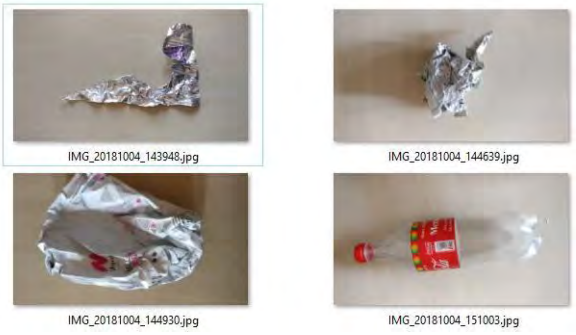


Fig.2: Dry Waste

Convolutional Neural Network (CNN) is a type of Neural Network (NN) mainly used for image classification, objection recognition, face recognition, etc. CNN image classification takes image as input and classifies it into one of the output categories. The input image is considered as an array of RGB matrix with dimensions h * w * d, where h and w are height and width of the image and d is dimensions in RGB, i.e. 3.

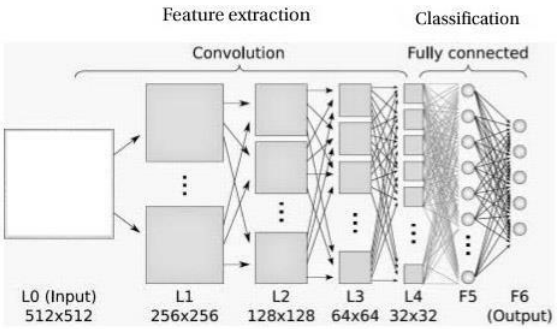


Fig.3: CNN Architecture

Depending on the number of classes for classification, there are two types of image classification: 1. Binary image classification- As the name suggests, there are only two output classes. The proposed model classifies the waste into wet or dry. Hence proposed model is based on Binary image classification.

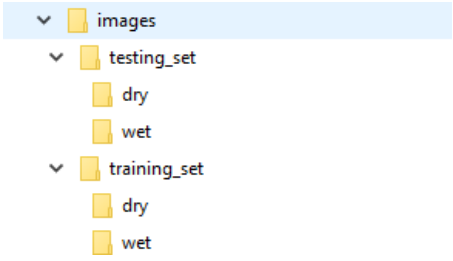


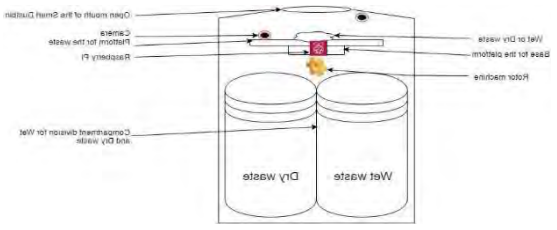
Fig.4: Dataset structure

2. Multi-class image classification- There are multiple classes of output (more than 2). A good example for this would be, further classifying the dry waste according to recycling categories like- paper, metal, cardboard, glass, etc. Following are the steps involved in image classification- 1. Training the CNN model using image dataset- The CNN model is constructed by specifying the number and types of hidden layers. The weights are initialized randomly. The training begins in iterations or epochs. With each epoch, the model leans more and more different hidden features in the input image dataset. 2. Testing the model- The model is evaluated for its performance based on the output produced for test dataset. The most common performance metrics that are used are- accuracy, confusion matrix, precision, recall and f1 score. 3. Prediction- After the model has been successfully trained and tested, the weights of the models are saved in the hard drive for later use (they can be directly loaded next time). For prediction, a single image is passed through the CNN model and the output is produced as one of the classes. A Raspberry Pi is a low cost small size computer board. It provides all the features of a Linux computer at a low-power consumption level. Due to its good performance at small size, it is used for small to large scale projects with medium hardware requirements. It is used for the

computational part of the system. The CNN prediction phase takes place on the Raspberry Pi. The input and output of CNN is connected to the Raspberry Pi. The other hardware components are connected to the Raspberry Pi board. It is a Raspberry Pi hardware component that is used for capturing the images of the waste placed on the smart dustbin. As it can be seen in Fig, two of such cameras are required for capturing

images from top view and side view. The output of camera, i.e. the waste image is connected to the input of the CNN, i.e. input waste image. It is a Raspberry Pi hardware component that is used for rotating the platform of the smart dustbin where the waste is placed. Based on the classification output from CNN as wet or dry, the left or right rotation signal is sent from the Raspberry Pi to the rotor mechanism. The rotor rotates and the waste placed on the platform falls into the respective compartment. The smart dustbin's main physical body consists of a dustbin with two compartments and the platform. The waste is placed on the platform and after segregation, it is

dropped to one of the dustbin compartments. The CNN model is deployed on the Raspberry Pi. The weights of the neurons in model saved in the step 2.2.2 are loaded into the model. The camera input image is connected to the CNN model input and CNN model output, the result of classification (wet or dry) is connected to the rotor mechanism. Finally, all the system components are connected to each other. The image of the waste kept on the dustbin platform is captured by the camera, sent to CNN, which sends classification output to rotor mechanism and the waste is dropped into respective compartment. Hence the input waste is successfully segregated.



RESULTS AND DISCUSSION

Fig.5: Block Diagram

another feature that can be added and using IoT technology, a

In Smart Dustbin, a user will add a waste which might be dry or wet. The image of the waste will be captured by the camera placed inside the Smart Dustbin. By using Raspberry pi the image of waste will be segregated with the help of CNN algorithm in Machine Learning. The algorithm detects and classifies the waste using CNN model trained by dataset. After classification a motor mechanism is use to drop the waste into respective compartments. The first prototype CNN model was built for image classification of wet and dry waste images. It showed accuracy of 80%. The accuracy can be further improved by collecting larger dataset and fine tuning the hyper parameters used in the CNN model.

IV. CONCLUSION

Segregation of wet and dry waste is very important in waste management. Despite having two separate dustbins, people often get confused and they put the waste in the wrong dustbin. The proposed system solves this problem. Manual efforts of classification of wet and dry waste are eliminated, thereby making the process faster and efficient. Smart dustbin automatically classifies and segregates the waste into separate compartments. Additional features can make the system more compelling as for general usage.

V. FUTURE SCOPE

This idea can be further extended to improve accuracy by using various sensors like metal detector, moisture sensor, etc. Indicating the amount of empty space left in dustbin can be network of the dustbins can be created, and the user will be informed about the nearest location of smart dustbin and its remaining waste collection capacity. Further the IoT can be used for indicating to the nearest waste collectors to empty the dustbin when filled.

VI. ACKNOWLEDGEMENT

We sincerely thank our guide Mrs. Shiwani Gupta for her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work. We thank the HOD, Dr. Sheetal Rathi, Dean Academic Dr. R.R. Sedamkar, the Principal, Dr. B. K. Mishra and the college management for their support.

REFERENCES

Mindy Yang , Gary Thung "Classification of Trash for Recyclability Status "Stanford university. Available: http://cs229.stanford.edu/proj2016/poster/ThungYan gClassificationOfTrashForRecyclabilityStatus- poster.pdf

C. Capel, "Waste sorting - a look at the separation and sorting techniques in today's European market," Waste Management World, 2008.

Comparing Deep Learning and Support Vector Machines for Autonomous Waste Sorting by George
E. Sakr, Maria Mokbel, Ahmad Darwich, Mia Nasr

Heart Disease Classification Using Optimized Machine Learning Algorithm

Ms. Aishwarya Manjalkar PG Student , Mrs. Shiwani Gupta Assistant Professor
Assistant Professor Department Of Computer Engineering Thakur College of Engineering and Technology

***Abstract*— Heart Disease is a highly significant contributor to loss of quality and quantity of life all over the world. Early detection and risk prediction is very important from the point of view of patient as well as doctors' perspective. The diagnosis is found to be a challenging issue which can offer a computerized estimate about the level of heart disease so that supplementary action can be made easy. The major risk factors include obesity, diabetes, family history, smoking and high cholesterol. Fatigue and chest pain are the most common symptoms for getting the heart disease. In general, treatment for these heart patients is given on the basis of lab test results, patient history and patient's answers to questions. Nowadays, a number of machine learning algorithms are developed for predicting the early stage of heart disease. In many cases, machine learning algorithms are used over a huge size of heart disease data. This would create high computation complexity and less efficiency. In order to overcome this issue, a range of feature selection algorithms is developed for identifying more significant features to predict the heart disease. The proposed work is to design Wrapper based Genetic Algorithm (GA) to select the significant features from the heart disease data where exhaustive search is infeasible. A vector of feature weights is optimized through fitness function based on no. of features and classification accuracy using Sequential Minimal Optimization (SMO) algorithm. SMO solves the quadratic programming (QP) problem which arises while processing large dataset by dividing into a series of smaller QP problems which can be solved analytically. Hence this combination will certainly work to provide good classification accuracy.**

Keywords: **Genetic Algorithm (GA),Sequential Minimal Optimization (SMO),Heart Disease.**

I. INTRODUCTION

Heart disease is result of alteration of functionality and structure of heart. This leads to inadequate pumping due to which organs and tissues receives insufficient amount of oxygen for their metabolic needs. Tissues could become necrotized due to lack of oxygen and may cause death. Diagnosis is complicated process that should be precisely executed. On basis of doctor's knowledge and experience, analysis is often made.

This may bring about undesirable results along with excessive medical costs of treatment given to patients. This work focuses on predicting heart disease on basis of features.

Causes of heart disease are; Age (For men, the risk of heart disease goes up after age 55; for women, the risk rises sharply after menopause.),Being inactive, Family history of coronary heart disease, Genetics, High blood pressure, High levels of LDL "bad" cholesterol or low levels of HDL "good" cholesterol, Obesity, Smoking, Stress, Diet of junk food, Diabetes, etc.

Heart disease describes a range of conditions that affect your heart. Diseases under the heart disease umbrella include blood vessel diseases, such as coronary artery disease; heart rhythm problems (arrhythmias); and heart defects you're born with (congenital heart defects), among others.

The term "heart disease" is often used interchangeably with the term "cardiovascular disease." Cardiovascular disease generally refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that affect your heart's muscle, valves or rhythm, also are considered forms of heart disease.

Cardiovascular disease symptoms may be different for men and women. For instance, men are more likely to have chest pain; women are more likely to have other symptoms along with chest discomfort, such as shortness of breath, nausea and extreme fatigue.

Symptoms can include:

- Chest pain, chest tightness, chest pressure and chest discomfort (angina)
- Shortness of breath
- Pain, numbness, weakness or coldness in your legs or arms if the blood vessels in those parts of your body are narrowed
- Pain in the neck, jaw, throat, upper abdomen or back

The main objective is to introduce a heart disease prediction system by selecting the essential feature and then using classification to predict heart disease. There are various feature reduction algorithms from which GA is the most efficient to optimize the features from the dataset.

The aim is to use first the Genetic Algorithm will be applied to select the significant features from dataset. After that heart disease will be predicted based on those features using Sequential Minimal Optimization (SMO).

Heart disease diagnosis is found to be a challenging issue which can offer a computerized estimate about the level of heart disease so that supplementary action can be made easy. Thus, heart disease diagnosis has expected massive attention worldwide among the healthcare environment. Heart disease is one of the most rampant causes of death around the world and was deemed as a major illness in Middle and Old ages. Coronary artery disease, in particular, is a widespread cardiovascular malady entailing high mortality rates. Heart disease is a deadly disease that large population of people around the world suffers from. When considering death rates and large number of people who suffers from heart disease, it is revealed how important early diagnosis of heart disease. world. Early detection and risk prediction is very important for patients' treatment and doctors' diagnose. Traditional way of diagnosis is not sufficient for such an illness.

In order to be practically used for collecting and analysing

patients' data in healthcare industries, the system consists of four parts: data interface, data preparation, feature selection and classification. Data interface response to obtain hospitals' raw data from hospital; data pre-processing is needed for data integration, data cleaning and rating mapping etc. Key features were then selected by various dimensionality reduction algorithms such as PCA, LDA, KNN, K-means, SVD, etc. to reduce dimensionality. Classification is where the different features are used to determine the heart disease.

II. LITERATURE REVIEW

		when evaluated using a collected database.	
1	Prediction of Heart Disease Using Neural Network [1]	In this paper, a heart diseas e prediction system which uses artificial neural network backpropagatio n algorithm is proposed.	The propos ed methodology can be enhanced as a hybrid model with other classification algorithms in order to obtain more accurate diagnosis for heart disease.
2	Heart Disease Diagnosis System Using Fuzzy Logic [2]	In this paper heart diseas e detection system is implemented using fuzzy logic. The proposed system has 98% accuracy	The propos ed system is not accurate with large dataset.
3.	Cardiovascular Risk Prediction Method Based on CFS Subset Evaluation and Random Forest Classification Framework[3]	In this paper key features were selected by CFS Subset Evaluation combined with Best-First-Search method to reduce imensionality. Random forest was inducted as basic classifier to identify risk level, which is a prior trial in CVD risk prediction field.	The system has a significantly higher accuracy of 91.6%
4.	Prediction of Heart Diseas e Using Hybrid Technique For Selecting Features [4]	This paper proposed a prediction of heart disease using random forest and naive bayes.	Naïve Bayes achieved best performance with 10 highly valued features at 84.16 % ,while Random Forest gains 84.16% with 12 features.
5.	An optimized feature selection based on genetic approach and support vector machine for heart disease[5]	This paper proposes an optimization function on the basis of SVM. This objective function is used in the GA for selecting the more significant features to get heart disease.	The SVM classifier produces 88.34% of accuracy when classifying the heart disease with the selected features.
6.	Efficient Heart Disease Prediction System using K-Nearest Neighbor Classification Technique	This paper proposes a K-nearest neighbor classifier to achieve accuracy of around 80%by using 14 attributes from dataset.	The accuracy can be increased by reducing the number of attributes.
7	Computer aided decision making for heart disease detection using hybrid neural network- Genetic algorithm	This paper proposes a hybrid method for the diagnosis of coronary artery disease using GA for feature selection and NN to train the data.	To enhance the erformance of the proposed method swarm intelligence methods can be used.

Table No. 1. Literature Review

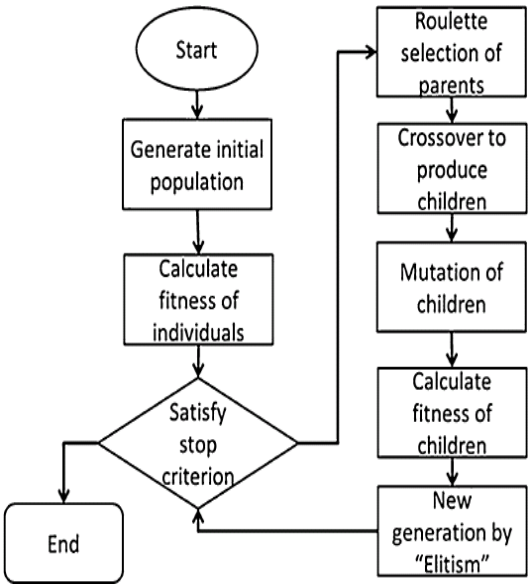
III. METHODOLOGY

1. Genetic algorithm based feature selection

A wrapper based GA is used to select the more significant features. The vital role of the wrappers used in the GA is to find the space and evaluate each subset by executing a model on the subset. A search algorithm used in the GA is to find a population of candidate solution(individual) and an optimization problem is established for an improved solution. A candidate solution is referred to an element of potential solutions to a given problem. The total population comprises a number of candidate solutions being developed by GA. A candidate solution consists of a set of parameters which are also called as genotypes or chromosomes. These set of parameters are represented as genomes. Genomes are represented as binary strings whereas the candidate solutions are represented as a numerical form. In addition, the evolution procedure of GAs is started from an initial population of randomly created genomes. The fitness function is used in all the iterations for calculating the fitness value. The overall best

fitness value is computed onthe basis of comparison between the current populations. Genotypes are used to represents the feature vector of the original dataset. The phenotypes of the chromosomes are used representing them asks of the feature vector. Hence, the phenotypes labeled as '0' represent the removed features while phenotypes labeled as '1' represent the selected features. On the basis of phenotypes, a group of subsets is created by each genotype and are used as training sets.

Figure below represents the working principle of a GA.



1.1 Roulette wheel selection

In roulette wheel selection, the expected value of an individual's fitness is divided by the actual fitness of the population. The size of the roulette wheel is fixed on the basis of size being proportional to the individual's fitness. Hence, each individual is allocated a part of the roulette wheel. The roulette wheel is rotated on the basis of the number of individuals in the population. On each rotation, an individual is selected as a pool of parents for the next generation.

Roulette Wheel Selection Algorithm:
Step 1: Addition of total expected value 't' of the individuals in the population.
Step 2: Up to N times:
i. Select a random number 'r' between o and 't'.
ii. For each individual in the population
Calculate the sum of the expected values, until the sum is greater than or equal to 'r'.
The individuals (features) are selected whose expected value greater than the sum of the expected values.

1.2 Crossover

In Crossover, two-parent solutions are taken to produce a child. Afterwards, the selection (reproduction) process is done followed by the development of better individuals. The reproduction process is used for developing replicas of significant strings but does not produce new ones. Here, a single point crossover is used to perform mating of two chromosomes. Here, two chromosomes are cut once and the cuts are exchanged between two chromosomes.

Crossover Algorithm:
Step 1: A pair of two individual strings is selected for the mating with the help of reproduction operator
Step 2: A cross-site is selected on the basis of random manner beside the string length.
Step 3: Swapping of position values between the two strings

1.3 Mutation

Once the crossover process is finished, the strings are subjected to perform mutation process. Mutation of a bit encompasses flipping a bit from 0 to 1 and 1 to 0. The vital role of mutation process is to recover the randomly troubling genetic information and missing genetic materials. The working principle of the mutation process is more similar to a traditional search operator. The mutation process is used for the examination of the entire search space when crossover abuses the current solution to find better ones. In addition, the objective of the mutation process is to preserve genetic diversity in the population. Hence, the mutation process

[At End of Each Step, Update *b*]

4. Compute *b1* and *b2* as follows

$$b_1 = E_1 + y_1(\alpha_1^{\text{new}} - \alpha_1)K(\vec{x}_1,\vec{x}_1) + y_2(\alpha_2^{\text{new,clipped}} - \alpha_2)K(\vec{x}_1,\vec{x}_2) + b$$
$$b_2 = E_2 + y_1(\alpha_1^{\text{new}} - \alpha_1)K(\vec{x}_1,\vec{x}_2) + y_2(\alpha_2^{\text{new,clipped}} - \alpha_2)K(\vec{x}_2,\vec{x}_2) + b$$

5. Choose $b = (b1 + b2)/2$

6. When $\alpha 1$ and $\alpha 2$ are not at bound, it is in fact guaranteed that $b = b1 = b2$

[Derivation of SMO Updated Rule]

7. Recall objective function that are to be optimize:

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j K(\vec{x}_i,\vec{x}_j)\alpha_i\alpha_j - \sum_{i=1}^{N}\alpha_i,$$
$$0 \le \alpha_i \le C, \forall i,$$
$$\sum_{i=1}^{N} y_i\alpha_i = 0.$$

8. Can write objective as:

$$\Psi = \frac{1}{2}K_{11}\alpha_1^2 + \frac{1}{2}K_{22}\alpha_2^2 + sK_{12}\alpha_1\alpha_2 + y_1\alpha_1 v_1 + y_2\alpha_2 v_2 - \alpha_1 - \alpha_2 + \Psi_{\text{constant}}$$

where:

$$K_{ij} = K(\vec{x}_i,\vec{x}_j),$$
$$v_i = \sum_{j=3}^{N} y_j\alpha_j^* K_{ij} = u_i + b^* - y_1\alpha_1^* K_{1i} - y_2\alpha_2^* K_{2i}$$

Starred values are the old values

produces new genetic structures in the population with the help of changing its original structure.

1.4 New Generation

Finally, new generation is created for improving the

9. Because of the constraint $\sum^{N}$

has to obey:

$$\alpha_1 + s\alpha_2 = \alpha_1^* + s\alpha_2^* = w$$

$(y_i\alpha_i) = 0$, each update

performance of the GA. In this procedure, the few best chromosomes are copied to the new population. The other chromosomes can be lost if they are not nominated to reproduce or if mutation or crossover terminates them.

2 SMO Algorithm as classifier

1. Input: C, kernel, kernel parameters, epsilon
2. Initialize b and all □'s to 0
3. Repeat until KKT satisfied (to within epsilon):

– Find an example *e1* that violates KKT (prefer unbound examples here, choose randomly among those)

– Choose a second example *e2*. Prefer one to maximize step size (in practice, faster to just maximize |E1 – E2|). If that fails to result in change, randomly choose unbound example. If that fails, randomly choose example. If that fails, re-choose e1.

– Update α1 and α2 in one step

– Compute new threshold b

where *w* is a constant and *s* is *y1y2*

2 11 22 12 11 12 2 1 2 1
*u +y α *(K -K) + y α *(K -K)) +y y –y y*

3 1 1
12 11
2 2 22 21

15. Based on values of *w* and *v*, we rewrite

$$\alpha_2(K_{11} + K_{22} - 2K_{12}) = s(K_{11} - K_{12})w + y_2(v_1 - v_2) + 1 - s$$

as:

α2(*K11 + K22 + 2K12*) + *y2(u1-u2+y2 + y1)*

16. From this, we get our update rule:

$$\alpha_2^{new} = \alpha_2 + \frac{y_2(E_1 - E_2)}{\eta}$$

Where,

E1 = u1 – y1, E2 = u2 – y2, η = K11 + K22 – 2K12

10. This lets us express the objective function in terms of α2 alone:

$$\Psi = \tfrac{1}{2}K_{11}(w - s\alpha_2)^2 + \tfrac{1}{2}K_{22}\alpha_2^2 + sK_{12}(w - s\alpha_2)\alpha_2 + y_1(w - s\alpha_2)v_1 - w + s\alpha_2 + y_2\alpha_2 v_2 - \alpha_2 + \Psi_{constant}$$

11. Find minimum by first derivative with respect to α2:

$$\frac{d\Psi}{d\alpha_2} = -sK_{11}(w - s\alpha_2) + K_{22}\alpha_2 - K_{12}\alpha_2 + sK_{12}(w - s\alpha_2) - y_2v_1 + s + y_2v_2 - 1 = 0$$

12. If second derivative positive (usual case), then the minimum is where (by simple Algebra, remembering *ss* = 1

$$\alpha_2(K_{11} + K_{22} - 2K_{12}) = s(K_{11} - K_{12})w + y_2(v_1 - v_2) + 1 - s$$

13. Recall that w = α1 + s α2 = α1 * + s α2 *

14. So we can rewrite

$$\alpha_2(K_{11} + K_{22} - 2K_{12}) = s(K_{11} - K_{12})w + y_2(v_1 - v_2) + 1 - s$$

as: α (*K + K + 2K)= s(K -K)(sα *+α *) + y (u –*

CFS Subset Evaluation and Random Forest Classification Framework'', IEEE, 2017

4] Kanika Pahwa, Ravinder Kumar, "Prediction of Heart Disease Using Hybrid Technique For Selecting Features", IEEE, 2017

5] Chandra Babu Gokulnath, S. P. Shantharajah," An

optimized feature selection based on genetic approach and

2 2 1 2
support vector machine for heart disease", Springer, 2018

IV. CONCLUSION

The system is divided in two parts:

1. The dataset will be optimized using genetic algorithm. In this the dimensionality of the dataset will be reduced and significant features will be selected.
2. The SMO classifier will be used to predict heart disease based on the optimized dataset.

The system will increase the accuracy of heart disease prediction system using GA for optimization and SMO for classification.

REFRENCE

1] Tülay Karay, Özkan Kilic, “Prediction of Heart Disease Using Neural Network “, IEEE, 2017
2] Hasan Kahtan, Kamal Z. Zamli, Azma Abdullah,” Heart Disease Diagnosis System Using Fuzzy Logic ”, ACM, 2018
6] Nida Khateeb, Muhammad Usman ,“Efficient Heart Disease Prediction System using K- Nearest Neighbor Classification Technique”, ACM, 2018
7] Zeinab Arabasadi , Roohallah Alizadehsani , Mohamad Roshanzamir , Hossein Moosaei , Ali Asghar Yarifard, “Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm”, Elsevier, 2017
8] Padma, T., Mir, S.A., Shantharajah, S.P., “Intelligent decision support system for an integrated pest management in apple orchard. In: Intelligent Decision Support Systems for Sustainable Computing”, Springer, (2017)
3] Shan Xu, Zhen Zhang, Daoxian Wang, Junfeng Hu, Xiaohui Duan,’’ Cardiovascular Risk Prediction Method Based on
9] Manogaran, G., Vijayakumar, V., Varatharajan, R., Kumar, P.M., Sundarasekar, R., Hsu, C.H.: Machine learning based big data processing framework for cancer diagnosis using hidden markov model and gm clustering. , Springer, 2017
10] Shouman M, Turner T, Stocker R., “Using data mining techniques in heart disease diagnosis and treatment”, IEEE, 2012
11] 8] S. U. Amin, K. Agarwal and R. Beg, "Genetic neural network based data mining in prediction of heart disease using risk factors", IEEE, 2013
12] A. H. Chen, S. Y. Huang, P. S. Hong, C. H. Cheng and E. J. Lin, "HDPS: Heart disease prediction system," ACM, 2011

Classification of Image Segmentation for Brain Abnormality

Mr Shreyash Sharma PG Student, Mrs. Megharani Patil Assistant Professor
Computer Engineering Department, Thakur College of Engineering and Technology, Mumbai, India.
shreyashssharma@gmail.com, megharanitpatil@gmail.com

***Abstract*— Brain tumour is one of the most dangerous disease occurring commonly among human beings. Minute structural difference of brain results in major disorders such as schizophrenia, Epilepsy, inherited speech and language disorder, Alzheimer's dementia etc. The MRI brain abnormality segmentation is an important technique in medical diagnosis. Due to large variance and complexity of abnormal characteristics such as size, location, intensity and shape in MRI images, prediction of abnormal region is very complex. So currently manual tracing and delineating of segmentation of brain abnormality is in practice. For segmentation of the Brain MRI images used different techniques like K-means, edge detection and Morphological operations. Further all these techniques are combined and observed for the segmentation results.**

Accuracy is measured for the classification process and the optimal classifier can be found by comparing the accuracies obtained. It will be described the network building algorithm, chosen practical field for proposed method application and showed the results of its programming implementation.

Keywords—CT, MEG, EEG, PET, MRI, K-means, FCM, SVM, ANN, SPECT, SBRG.

I. INTRODUCTION

Raymond Vahan Damadian (born March 16, 1936) is an American physician, medical practitioner, and inventor of the first MR (Magnetic Resonance) Scanning Machine. Image segmentation is one of the primary steps in image analysis for object identification. The main aim is to recognize homogeneous regions within an image as distinct and belonging to different objects[12]. The human brain is the center of the human nervous system and is the most complex organ in any creature on earth. Any abnormality in brain leads to the total collapse of entire vital functions of the body.

Brain abnormality is one a player in the abnormal in the sensory system, notwithstanding spinal abnormal and fringe nerve abnormal[7]. This brain abnormality can be an essential tumor or a metastasis from a abnormal in another organ. Abnormality segmentation from MRI data is an important but time-consuming manual task performed by medical experts. Brain abnormal analysis is done by doctors but its grading gives different conclusions which may vary from one doctor to another[9]. The MRI T1-weighted and T2- weighted images are the most commonly used techniques MRI imaging. The MRI brain tumor segmentation provides useful information for surgical planning and further medical diagnosis. The large variance and complexity of tumor characteristics such as size, location, intensity and shape in MRI images is a difficult task. So the manual tracing and delineating of segmentation of brain tumor is in practice[10][16]. The segmentation is a significant tool in medical image dispensation and it has many applications, in discovery of abnormal, measuring tumor volume, surgical planning and preparation, discovering coronary edges, categorization of blood cell. Along with the advance of medical imaging, imaging modalities play an important role in the evaluation of patients with brain abnormal and have a significant impact on patient care. The emerging new imaging modalities, such as X-ray, Ultrasonography, Computed Tomography (CT), Magneto Encephalo Graphy (MEG), Electro Encephalo Graphy (EEG), Positron Emission Tomography (PET), Single-Photon Emission Computed Tomography (SPECT), and Magnetic Resonance Imaging (MRI), not only show the detailed and complete aspects of brain abnormal, but also improve clinical doctors to study the mechanism of brain abnormal at the aim of better treatment[5]. Brain tumor segmentation is to segment abnormal tissues such as active cells, necrotic core, and edema from normal brain tissues. Intensity normalization is a very critical step for the pre- processing of MRI, especially when classification and clustering methods are used for the segmentation[13].

II. PERFORMANCE MEASURES

[1] To provide an algorithm that guarantees the presence of abnormal by combining several procedures to provide a fool proof method of abnormal detection in CT brain images.

[2] To observe the methods utilized are filtering, contrast adjustment, negation of an image, image subtraction, erosion, dilation, threshold, and outlining of the brain abnormal.

[3] To provide information like size, dimension and position of the tumor, plotting contour of the abnormal and its boundary provides with information related to the tumor that can prove useful for various cases which will provide a better base for the staff to decide the curing procedure.

[4] To bring some useful information in simpler form in front of the users, especially for the medical staff treating the patient.

[5] To detect the size and location of brain abnormal from the Magnetic Resonance Images.

[6] To demonstrate knowledge including signs and symptoms, methods of detection and patterns of

spread using the image segmentation.

III. PROBLEM DEFINITION

Brain abnormal detection is a serious issue in imaging science. Generally, the severity of disease decides by size and type of tumor. An important step in analysis of brain MRI scan image is to extract the boundary and region of tumor[7]. To solve the problem, the proposed work describes the strategy for detection, segmentation and feature extraction of brain tumor part inexpensive format using MATLAB software. This software based approach aims to introduce an algorithm for detecting and segmenting the brain tumor from normal brain using basic image processing operations, segmentation, feature extraction in MATLAB. The abnormality detection in MR brain images is the main objective of this research work[4]. The abnormalities are identified by the human observer. The probability of error for such identification is very high due to the presence of psycho-visual effect. An extensive literature survey has revealed the availability of semi-automated and fully automated techniques for this application[14]. Most of them are computer based approaches and involves the concept of soft computing approaches such as fuzzy logic techniques. But the major drawback is that if the techniques are accurate, the time requirement is high and vice-versa.

IV. DESIGN AND ANALYSIS

A. Methodology

This method is classified into supervised learning, semi-supervised learning, and unsupervised learning algorithms. Most of brain abnormality segmentation algorithms are based on classification or clustering methods in the literature such as Fuzzy C-Means (FCM), k-means, Markov Random Fields (MRF), Bayes, Artificial Neural Networks (ANN), Support Vector Machines (SVM), Atlas-based, etc.

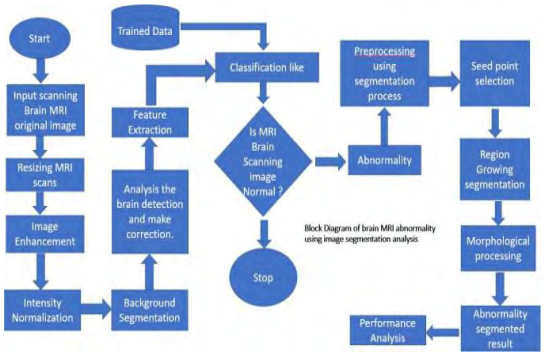


Figure 1: Block diagram

Step for this algorithm:

Input scanning Brain MRI original image: - Collect the brain MRI image from patient and read it.
Resizing MRI scans: - It is resizing the volume intensity of brain MRI image. It returns the volume that has the number of rows, columns, and planes specified by the three-element vector.
Image Enhancement: - Image enhancement is the process of adjusting digital images so that the results are more suitable for display or further image analysis.
Intensity Normalization: - It suggests a method where a set of standard histogram landmarks are learned from a set of MRI images. These landmarks are then used to equalize the histograms of the images to normalize[18].
Background Segmentation: - This method relies on the identification of a good threshold. Failing to find such a threshold may lead to poor segmentation. Thresholding is often used as an initial step in a sequence of image processing operations.
Analysis the brain detection and make correction: - It analysis the brain abnormal detection and correct the images after background segmentation[15].
Feature extraction: - It is a type of dimensionality reduction that efficiently represents interesting parts of an image as a compact feature vector. This approach is useful when image sizes are large and a reduced feature representation is required to quickly complete tasks such as image matching and retrieval.
Abnormality: - It find the small abnormal inside the brain images after classifying the feature extraction.
Seed point selection: - Seed-based region growing (SBRG) performs a segmentation of an image with respect to a point known as seed. Starting with a seed point the region will grow by appending to each seed those neighboring pixels that have properties similar to the seed[19].
Region Growing segmentation: - Region growing is a technique for extracting a region of the image that is connected based on some predefined criteria. This criterion can be based on intensity information or edges in the image. Morphological processing: - It is a broad set of image processing operations that process images based on shapes. Morphological operations apply a structuring element to an input image, creating an output image of the same size[20]. Abnormality segmented result: - It analysis and detects the brain abnormal images to give the result using image segmentation.
Performance Analysis: - It performs the brain abnormal image to find out the result of region growing, edge detection and morphological processing like dilation and erosion.

B. Technology Required

Feature Extraction Application	OpenCV 2.49, MS VS Community 2015, MATLAB 2017
Classifiers	rMATLABa 2017
Operating System	Windows 10, 64-bit
Processors	Intel® Core™ i7-4700MQ (2.4 GHz)
RAM	16GB

Table 1: Technology Require

V. IMPLEMENTATION

A. Thresholding algorithm

Select initial threshold value, typically the mean 8-bit value of the original image. Divide the original image into two portions - Pixel values that are less than or equal to the threshold background, Pixel values greater than the

threshold foreground. Find the average mean values of the two new images[23]. Calculate the new threshold by averaging the two means. If the difference between the previous threshold value and the new threshold value are below a specified limit, you are finished. Otherwise apply the new threshold to the original image keep trying.

B. Thresholding algorithm

Apply the derivative operator to detect edges of the image. Find the magnitude at the edges. Retain all edge having magnitude greater than threshold value. Find the position of crack edge, the crack edge is either retained or rejected. Check again in different values of threshold so as to find out the closed boundaries, segmentation of an image is obtained.

C. Region based segmentation

Segmentation of images is crucial to our understanding of them. Since the sixties a variety of techniques have been proposed and tried for segmenting images by identifying regions of some common property. These can be classified into two main classes. Merging algorithms which neighbouring regions are compared and merged if they are close enough in some property[25]. Splitting Algorithms in which large non-uniform regions are broken up into smaller areas which may be uniform. In all cases some uniformity criterion must be applied to decide if a region should be split, or two regions merged. This criterion is based on some region property which will be defined by the application, and could be one of many measurable image attributes such as mean intensity, colour etc. Uniformity criteria can be defined by setting limits on the measured property, or by using statistical measures, such as standard deviation or variance.

D. K nearest neighbors

The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, k is a user-defined constant, and an unlabeled vector is classified by assigning the label which is most frequent among the k training samples nearest to that query point. A commonly used distance metric for continuous variables is Euclidean distance. For discrete variables, such as for text classification, another metric can be used, such as the overlap metric.

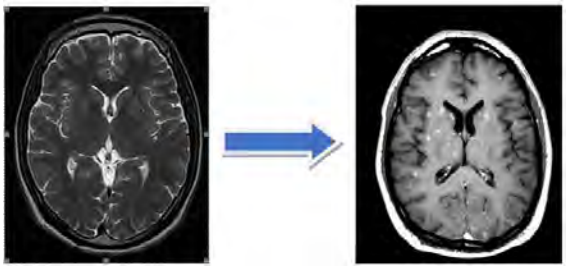


Figure 2: Convert normal brain into abnormal brain

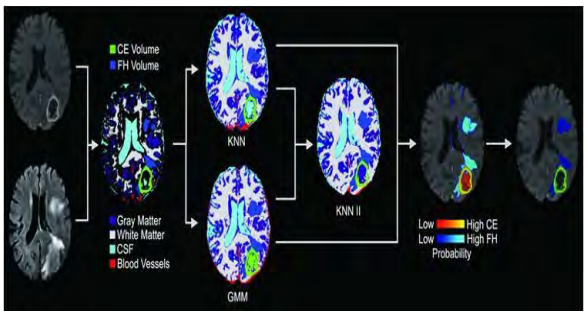


Figure 3: Classification of Image Segmentation

E. Fuzzy C Mean

The FCM based methods are modified by incorporating spatial constraint in the objective function of FCM. These methods are used for intensity inhomogeneity correction and partial volume segmentation of brain MR images.

VI. CONCLUSION

The MRI brain abnormality segmentation is an important technique in medical diagnosis. Recently, availability of public datasets provided a common medium to develop and objectively evaluate their methods with the existing techniques. MRI-based brain abnormality segmentation techniques have already shown great potential in detecting and analysing abnormal in clinical practice and will

undoubtedly continue to be improved in the future. Many of the current brain abnormality segmentation methods operate MRI images due to the non-invasive and good soft tissue contrast of MRI and employ classification and clustering methods by using different features and taking spatial information in a local neighbourhood into account.

REFERENCES

[1] Jin Liu, Min Li, Jianxin Wang, Fangxiang Wu, Tianming Liu, and Yi Pan," A survey of MRI-Based Brain Tumor Segmentation Methods", TSINGHUASCIENCEANDTECHNOLOGY ISSNll1007- 0214ll04/10llpp578-595 Volume 19, Number 6, December 2014.

[2] S. Mahalakshmia and T. Velmurugan," Segmentation of MRI Brain Images: A Comparative Analysis", International Journal of Control Theory and Applications ISSN : 0974–5572© International Science Press Volume 10 Number 23 2017.

[3] Hapsari Peni Agustin Tjahyaningtijas", Brain Tumor Image Segmentation in MRI Image", IOP Conf. Series: Materials Science and Engineering 336 (2018) 012012 doi:10.1088/1757-899X/336/1/012012.

[4] Natarajan P, Krishnan.N, Natasha S, Shraiya N, Bhuvanesh P, "Tumor Detection using threshold operation in MRI Brain Images", International Conference on Computational Intelligence and Computing Research,IEEE, 2012.

[5] Manoj K and Sourabh Y, "Brain Tumor Detection and Segmentation Using Histogram Thresholding", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-4, April 2012.

[6] M. Sinha, K. Mathur," IMPROVED BRAIN TUMOR DETECTION WITH ONTOLOGY", IJCER, Mar-Apr 2012 Vol. 2, Issue No.2, pp.584- 588

[7] M. Rakesh, T. Ravi, "Image Segmentation and Detection

of Tumor Objects in MR Brain Images Using FUZZY C-MEANS (FCM) Algorithm", IJERA,Vol. 2, Issue 3, MayJun2012,pp.2088-2094.

[8] Neeti Chadha, CSE Department, AKGEC Gzb, Digital Image Processing Notes, http://www.uptu.ac.in/pdf/sub_ecs_702_30sep14.pdf.

[9] Smita Pradhan, Dipti Patra, Development of Unsupervised Image Segmentation Schemes for Brain MRI using HMRF model, NIT, Rourkela, 2010.http://ethesis.nitrkl.ac.in/2870/1/final.pdf.

[10] Li-Hong Juan and Ming-Ni Wu, "MRI Brain Lesion Image Detection based on Color-Converted K-means Clustering Segmentation", Measurement, Elsevier, Vol. 43, pp. 941–949, April 2010.

[11] Dr. A.J.Patil, Dr.Prerana Jain, Ashwini Pachpande, Automatic Brain Tumor Detection Using KMeans And RFLICM, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 3, Issue 12, pp. 13896-13903, December 2014.

[12] R. Shalini, V. Muralidharan and M. Varatharaj, MRI Brain Tumor Segmentation using Kernel Weighted Fuzzy Clustering, International

[13] Journal of Engineering Research & Technology (IJERT), Vol. 3 Issue 4, pp. 121-125, April – 2014.

[14] Purcell, E. M., Torrey, H. C. and Pound, R. V. Resonance Absorption by Nuclear Magnetic Moments in a Solid. Phys. Rev. 69, 37-38.

[15] Kshitij bhagwat, Dhanshri more, sayali shinde, Akshay Daga, (2013): "Comparative study of brain tumour detection using k means, Fuzzy c means and Hierarchical clustrering algorithms," International journal of scientific and engineering research, vol. 4, issue 6. Pp 626-632.

[16] Nagalkar,V.J., Asole, S.S (2012): "Brain tumour detection using digital image processing based on soft computing," Journal of signal and image processing, vol.3, Issue 3, pp.102-105.

[17] Roy S, Carass A, Prince JL. A Compressed Sensing Approach for MR Tissue Contrast Synthesis. 22nd Conf. on Info. Proc. in Medical Image. (IPMI 2011); 2011. pp. 371–383.

[18] D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens, "A viscous fluid model for multimodal non-rigid image registration using mutual information," in Proceedings of the Medical Image Computing and Computer Assisted Intervention (MICCAI '02), pp. 541–548, 2002.

[19] J.Ashburner and K.J.Friston, "Nonlinear spatial normalization using basis functions," Human Brain Mapping, vol. 7, no. 4, pp. 254–266, 1999.

[20] Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D.

[21] J. Hawkes, "Nonrigid registration using free-form deformations: application to breast mr images," IEEE Transactions on Medical Imaging, vol. 18, no. 8, pp. 712–721, 1999.

[22] J. M. Fitzpatrick, D. L. G. Hill, and C. R. Maurer, "Image registration," in Handbook of Medical Imaging: Medical Image Processing and Analysis, M. Sonka and J. M. Fitzpatrick, Eds., vol. 2, SPIE Publications, 2004.

[23] Nagesh Subbanna, Doina Precup, Tal Arbel, "Iterative Multilevel MRF Leveraging Context and Voxel Information for Brain Tumour Segmentation in MRI," Conference on Computer vision and Pattern Recognition, IEEE, 2014.

[24] Poonguzhali, S., & Ravindran, G. (2006). A Complete Automatic Region Growing Method for Segmentation of Masses on Ultrasound Images. International Conference on Biomedical and Pharmaceutical Engineering (ICBPE).

[25] Kamalam Balasubramani, "Artificial Bee Colony Algorithm to improve brain MR Image Segmentation", International Journal on Computer Science and Engineering (IJCSE), ISSN: 0975-3397 Vol. 5 No. 01 Jan 2013.

[26] 20 Held K, Kops ER, Krause BJ, "Markov random field segmentation of brain MR images", IEEE Trans Med Imaging; 16:878–886, 1997.

[27] Rash Bihari Dubey, M.Tech., Madasu Hanmandlu, "The Brain MR Image Segmentation Techniques and use of Diagnostic Packages", AUR, 2010 doi:10.1016/j. acra.12.017, 2009.

Crop Prediction by Soil Analysis using Enhanced Random Forest Algorithm for Vidarbha Regions

Priyankar Ravindra Tiwari, Prof. Anand Khandare

Department of Computer Engineering,
Thakur College of Engineering and Technology,
Mumbai,India
priyankartiwarimy@gmail.com, anand.khandare1983@gmail.com

***Abstract*—In the years since its independence, India has made immense progress towards food security. Indian population has tripled, and food-grain production more than quadrupled. India ranked in the world's five largest producers of over 80% of agricultural produce items. India exported $39 billion worth of agricultural products in 2013, making it the seventh largest agricultural exporter worldwide, and the sixth largest net exporter. This represents explosive growth, as in 2004 net export were about $5 billion. India is the fastest growing exporter of agricultural products over a 10-year period, its $39 billion of net exports is more than double the combined exports of the European Union. So, yield prediction is very popular among farmers these days. Earlier yield prediction was performed by considering the farmer's experience on a particular field, weather and crop. Since farmers don't have knowledge about the presence of the nutrients and they don't have the idea about the crop to plough and pesticides to be used due to which performance of agriculture is degrading in economy and farmers getting into loss which they have to pay from their own pocket, which is also a root cause of farmers suicide. This makes the problem of predicting the yielding of crops an interesting challenge. Proposed system uses machine learning techniques in order to predict the category of the analyzed soil datasets. The category, thus predicted will indicate the yielding of crops. The problem of predicting the crop yield is formalized as a classification rule, where Boosted Tree and Random Forest algorithms will be used and for feature extraction K-Means will be used.**

Keywords—Random Forest, K-Means, Boosted Tree, Crop, Prediction, Vidarbha.

I. INTRODUCTION (*HEADING 1*)

Agriculture is the backbone of the Indian. The agriculture data increases day by day. Since a large population lives in rural areas and is directly or indirectly dependent on agriculture for a living. Outlay from farming forms the main source for the farming community[1]. The essential requirements for harvesting are water resources and ability to buy seeds, fertilizers, pesticides, labour etc. Most farmers raise the required capital by compromising on other essential expenditures, and when it is still insufficient they resort to credit from sources like banks and private commercial institutions. In such a situation, the repayment is dependent on the success of the harvest. If the harvest fails even once due to several factors, like bad weather pattern; soil type; improper, excessive, and ill-timed application of both fertilizers and pesticides; adulterated seeds and pesticides etc. Most power of soil in nature comes from soil survey efforts. Soil survey, or soil mapping, is the process of determining the soil types or other holding of the soil cover over a landscape, and mapping them for others to understand and use. Primary data for the soil survey is acquired by area sampling and supported by remote sensing. As the volume of data increase, it requires involuntary way for these data to be extracted when needed. Machine Learning can be used for pretend the next trends of agricultural processes. Every soil is a mixture of three main components: sand, clay and silt. Based on these factors we want to predict the soil for a particular cultivation.

Machine Learning task can be classified into two categories: Descriptive Machine Learning and Predictive Machine Learning. Descriptive Machine Learning tasks qualify the general properties of the database while predictive Machine Learning is used to predict expressed values based on patterns determined from known results.

Classification and prediction are two forms of data analysis that can be used to solution models describing important data classes or to predict future data trends. It is a process in which models learn to pretend a class label from a set of training data which can then be used to predict discrete class labels on new sample. Different subsets of these parameters are used in different prediction models for different crops.

II. LITERATURE SURVEY

The amount of data doubles almost every year. Hence, there is an urgent need for a new generation of computationally intelligent techniques and tools to assist human beings in extracting useful information from the rapidly growing volume of data. It is realized that extracting knowledge from large amount of data is typically ill-defined and it is difficult to model with large scale solution spaces. In such cases, precise models are impractical and are too expensive. Furthermore, the relevant available information is usually in the form of empirical prior knowledge and input and output data representing instances of the systems behavior. Therefore, one needs an approximate reasoning system which is capable of handling such imperfect information.

While Bezdek defines such approaches within a frame called Computational Intelligence, Zadeh explains the same using the Soft Computing paradigm. According to Zadeh, in contrast to traditional Hard Computing, Soft Computing is tolerant of imprecision, uncertainty, and partial truth.

From the research article, the researcher expresses that large amount of data which is collected and stored for analysis. Making appropriate use of these data often leads to considerable gains in efficiency and therefore economic advantages. There are several applications of Machine Learning techniques in the field of agriculture. The researchers implemented K-Means algorithm to forecast the pollution in the atmosphere, the K Nearest Neighbor is applied for simulating daily precipitations and other weather variables and different possible changes of the weather scenarios are analyzed using Support Vector Machines. Soil profile descriptions were proposed by the researcher for

classifying soils in combination with GPS based technologies. They were applied K-Means approach for the soil classification.

In a similar approach, crop classifications using hyper spectral data was carried out by adopting one of the Machine Learning approach i.e. Support Vector Machines. One of the researchers used an intensified fuzzy cluster analysis for classifying plants, soil and residue regions of interest from GPS based color images. In the agricultural science, clustering techniques are found in grading apples before marketing. Weeds were detected on precision agriculture. The researchers worked on rainfall variability analysis and its impact on crop productivity. The effect of observed seasonal climatic conditions such as rainfall and temperature variability on crop yield prediction was considered through an empirical crop model. Furthermore, there are two approaches to investigate the impact of climate change on crop production which include the crop suitability approach and the production function approach. Machine learning in Agriculture is a Novel field still a lot of work has been done in field of Agriculture using Machine learning.

In a proposed yield prediction model which used Machine Learning techniques for classification and Prediction. This model worked on input parameters crop name, land area, soil type, soil pH, pest details, weather, water level, seed type and this model predicted the plant growth and plant diseases and thus enabled to select the best crop based on weather information and required parameters. She proposed an approach which used unsupervised learning technique K Means Clustering technique to classify the soils into clusters based on the salinity factors. This work classified the soils as Sodic, Saline -Sodic and Acidic. This model enabled the analysts to select the best soil for crop productivity She proposed a crop yield prediction model that implanted two Machine Learning techniques namely Multiple Linear Regression and Density Based Clustering techniques.

The predict answers were Year', 'Rainfall', 'Area of Sowing', 'Yield', 'Fertilizers' (Nitrogen, Phosphorous and Potassium) and Response variable was 'Production'. In Kg/Hectares. A. conducted a study on the different Machine Learning techniques used in Agriculture. The techniques like K Means.KNN, ANN, SVM were studied related to Agriculture field and concluded that these techniques in combination with GPS and Remote sensing techniques can be used to study the characteristics of soil, classify soils, classify crops and for prediction too.

III. METHODOLOGY

1) Problem Study: A brief study of problems related to maximization of the productivity and prediction of crop yield will be done by going through the related literature review, and with the brief discussions with soil analysts and farmers and broader view of research problem will be gained.

2) Data Collection: After completing the problem study and gaining insight of research problem the related data will be collected from respective government department database and respective research center. The dataset will consist of Soil Nutrient status individual field wise. data will be divided for training and testing purposes.

Data will be preprocessed and will be transformed into two excel sheets one for training and one for testing purposes.

3) Parameter Study: The Dataset collected will consists of soil composition parameters and is one of subsets for the prediction of yield. The data will consists of 12 parameters out of Sample no, Ph, EC, OC, N, P, K, S, Cu, Fe, Zn, Mn out of which 7 parameters(Ph,EC,OC,N,P,K,S) are classified as Macro-Nutrients and remaining 4 parameters (Cu,Fe,Zn,Mn)are Micro-Nutrients.

4) Training Data Categorization: The individual tuple values of each parameter will be classified into LOW,HIGH and MEDIUM category based on critical limits defined by soil chemists for soils in Vidarbha Region.

5) Applying Machine Learning Approaches:

a) Random Forest: Let assume that the user knows about the construction of single classification trees. Random Forests grows many classification trees. To classify a new object from an input vector, put the input vector down each of the trees in the forest. Each tree gives a classification, and we say the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest).

Each tree is grown as follows:

1. If the number of cases in the training set is N, sample N cases at random - but *with replacement*, from the original data. This sample will be the training set for growing the tree.
2. If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.
3. Each tree is grown to the largest extent possible. There is no pruning

When the training set for the current tree is drawn by sampling with replacement, about one-third of the cases are left out of the sample. This oob (out-of-bag) data is used to get a running unbiased estimate of the classification error as trees are added to the forest. It is also used to get estimates of variable importance.

After each tree is built, all of the data are run down the tree, and proximities are computed for each pair of cases. If two cases occupy the same terminal node, their proximity is increased by one. At the end of the run, the proximities are normalized by dividing by the number of trees. Proximities are used in replacing missing data, locating outliers, and producing illuminating low-dimensional views of the data.

b) ID3 Algorithm: In decision tree learning, ID3 (Iterative Dichotomiser 3) is an algorithm invented by Ross Quinlan used to generate a decision tree from a dataset. ID3 is the precursor to the C4.5 algorithm, and is typically used in the machine learning and natural language processing domains.

The ID3 algorithm begins with the original set S as the root node. On each iteration of the algorithm, it iterates through every unused attribute of the set S and calculates the entropy H(S) (or information gain IG(S)) of that attribute. It then selects the attribute which has the smallest entropy (or largest information gain) value. The set S is then split or partitioned by the selected attribute

to produce subsets of the data. (For example, a node can be split into child nodes based upon the subsets of the population whose ages are less than 50, between 50 and 100, and greater than 100.) The algorithm continues to recur on each subset, considering only attributes never selected before.

Recursion on a subset may stop in one of these cases:

- Every element in the subset belongs to the same class; in which case the node is turned into a leaf node and labelled with the class of the examples.
- there are no more attributes to be selected, but the examples still do not belong to the same class. In this case, the node is made a leaf node and labelled with the most common class of the examples in the subset.
- there are no examples in the subset, which happens when no example in the parent set was found to match a specific value of the selected attribute. An example could be the absence of a person among the population with age over 100 years. Then a leaf node is created and labelled with the most common class of the examples in the parent node's set.

Throughout the algorithm, the decision tree is constructed with each non-terminal node (internal node) representing the selected attribute on which the data was split, and terminal nodes (leaf nodes) representing the class label of the final subset of this branch.

The ID3 algorithm is used by training on a data set S to produce a decision tree which is stored in memory. At runtime, this decision tree is used to classify new test cases (feature vectors) by traversing the decision tree using the features of the datum to arrive at a leaf node. The class of this terminal node is the class the test case is classified as.

IV. System Architecture

The proposed system will be designed while keeping rain statics in mind for corresponding soil. The rain data will be merged with soil data so that it will be easy to map the soil for the particular crop. The merged data will be stored separately and will be provided directly to the algorithm. The crop data will be the input for the proposed system. The crop data will be preprocessed before applying algorithm on that. The processed crop data and soil data with rain data will be processed and run through the proposed algorithm. The graphical result will be predicted which will be reported and predicted result will be displayed.

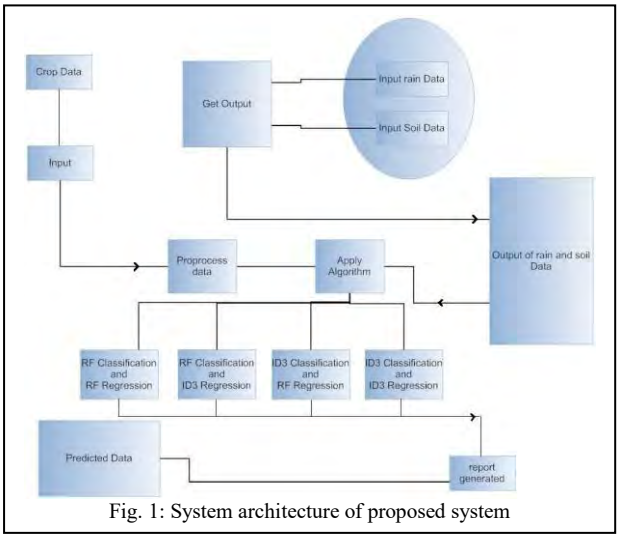


Fig. 1: System architecture of proposed system

V. Expected Outcome

The predicted result will consist of bar graph representing the success rate of predicted crop, macro nutrient and micro nutrient, rain fall history in particular soil.

VI. Conclusion

In this proposed system, by using the method of classification using ID3 combined with Random Forest Algorithm will help farmers in Vidarbha Region in deciding the crop to plough and the amount of fertilizer to be used in farm. This will surely help farmers and will enable them to be independent.

References

[1] Supriya D M "Analysis of Soil Behavior and Prediction of Crop Yield using Data Mining Approach" in International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 5, May 2017.

[2] Vaneesbeer Singh, Abid Sarwar "Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach" in International Journal of Advanced Research in Computer Science, Volume 8, No. 5, May – June 2017.

[3] Profile of Maharashtra and selected districts, shodhganga.inflibnet.ac.in/bitstream/10603/121515/13/13_chapter4.pdf.

[4] Profile of Maharashtra and selected districts, shodhganga.inflibnet.ac.in/bitstream/10603/121515/13/13_chapter4.p df.

[5] Profile of Maharashtra and selected districts, shodhganga.inflibnet.ac.in/bitstream/10603/121515/13/13_chapter4.p df.

[6] Andrew.W "Moore Professor School of Computer Science Carnegie Mellon University", Naïve Bayes Classifiers, www.cs.cmu.edu/~awm awm@cs.cmu.edu

[7] Andrew.W "Moore Professor School of Computer Science Carnegie Mellon University", Naïve Bayes Classifiers, www.cs.cmu.edu/~awm awm@cs.cmu.edu

[8] B. Bhattacharya, D.P. Solomatine "Machine learning in soil classification" Elsevier 2006 Special Issue, Neural Networks 19 (2006) 186–195

[9] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio "Generative Adversarial Nets"

*[10]Kriegel, Hans-Peter; Schubert, Erich; Zimek, Arthur (2016). "The (black) art of runtime evaluation: Are we comparing algorithms or implementations?". Knowledge and Information Systems. **52**: 341–378. doi:10.1007/s10115-016-1004-2. ISSN 0219-1377*

[11]MacKay, David (2003). "Chapter 20. An Example Inference Task: Clustering" (PDF). Information Theory, Inference and Learning Algorithms. Cambridge University Press. pp. 284&ndash, 292. ISBN 0-521-64298-1. MR 2012999.

[12]Coates, Adam; Ng, Andrew Y. (2012). "Learning feature representations with k-means" (PDF). In G. Montavon, G. B. Orr, K.-R. Müller. *Neural Networks: Tricks of the Trade*. Springer.

[13]Csurka, Gabriella; Dance, Christopher C.; Fan, Lixin; Willamowski, Jutta; Bray, Cédric (2004). *Visual categorization with bags of keypoints* (PDF). ECCV Workshop on Statistical Learning in Computer Vision.

[14]Coates, Adam; Lee, Honglak; Ng, Andrew Y. (2011). *An analysis of single-layer networks in unsupervised feature learning* (PDF). International Conference on Artificial Intelligence and Statistics (AISTATS). Archived from the original (PDF) on 2013-05-10.

[15] Schwenker, Friedhelm; Kestler, Hans A.; Palm, Günther (2001). "Three learning phases for radial-basis-function networks". *Neural Networks*. **14** (4–5)

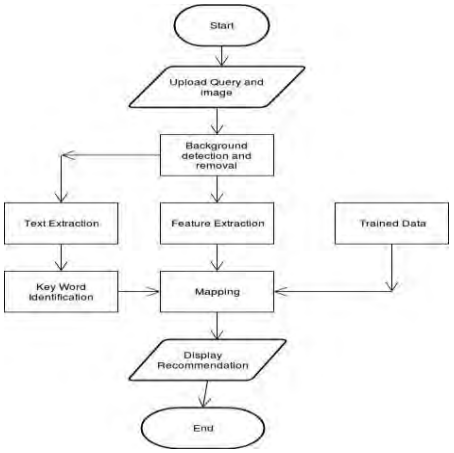
Efficient Recommendation of IT Problems using Image Recognition

Divij Shah, Paridhi Somani, Siddhita Tawde
Thakur College of Engineering & Technology Kandivali (E)
divijshah11@gmail.com, psomani27@gmail.com, tsiddhita05@gmail.com

***Abstract*—Troubleshooting is an important factor of adaptation of technology in a wider way, however currently used methods do not provide a user friendly way of troubleshooting. As per our analyses of end user requirements, a simpler way of providing recommendation to problems would be by the means of image capturing and recognition. We will train a model which will analyze the data and detect indicators such as icons and error codes. We will train a model which will analyze the data and detect indicators such as icons and error codes.We will also use different classification algorithms for more accurate results. Collect query images from different open source repositories and train the following classification for different problems and its recommendations.**

I. INTRODUCTION

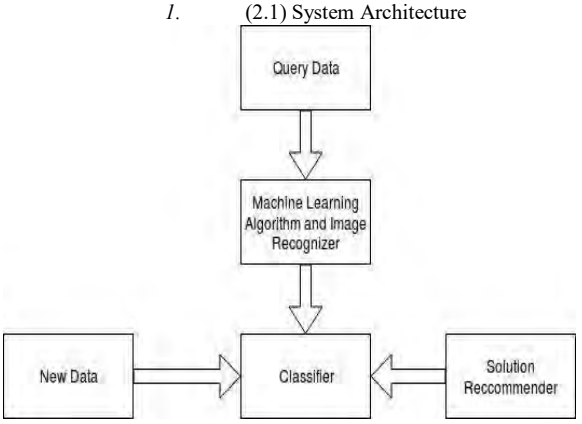
Troubleshooting is a systematic approach to problem solving that is often used to find and correct issues with complex machines, electronics, computers and software systems. The first step in troubleshooting is gathering information on the issue, such as an undesired behaviour or a lack of expected functionality. Other important information includes related symptoms and special circumstances that may be required to reproduce the issue. Once the issue and how to reproduce it are understood, the next step might be to eliminate unnecessary components in the system and verify that the issue persists, to rule out incompatibility and third-party causes. After common causes are ruled out, the troubleshooter may resort to the more systematic and logical process of verifying the expected function of parts of a system. One common method is the split-half troubleshooting approach: With a problem resulting from a number of possible parts in series, one tests halfway down the line of components. If the middle component works, one goes to the middle of the remaining parts, approaching the end. If the test finds a problem at the midpoint, one does a split towards the start of the line until the problem part is found. The split-half process can save time in systems that depend on many components. Once the problem part is identified, it may be adjusted, repaired or replaced as needed. Evidence of effective troubleshooting is indicated when the issue is no longer reproducible and function is restored one. The success of troubleshooting often depends on the thoroughness and experience of the troubleshooter. Template matching is a 'brute-force' algorithm for object recognition. Its working is simple: create a small template (sub-image) of object to be found, say a football.



Now do a pixel by pixel matching of template with the image to be scanned for, placing (centre of) the template at every possible pixel of the main image.Then, using a similarity metric, like normalized cross correlation, find the pixel giving maximum match. That is the place which has a pattern most similar to your template(football). The challenge is to merge these two concepts, that is, to implement pattern matching for troubleshooting. The collected data, because of the vague and careless nature of taking screenshots needs to be filtered to remove unnecessary information. After this initial preprocessing, all that is left is to train a model and implement it for a proof of concept.

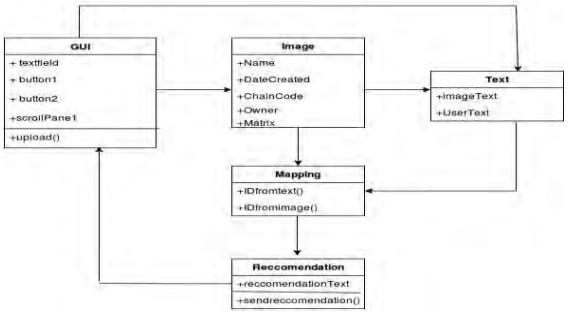
II. PROPOSED WORK

1. (2.1) System Architecture

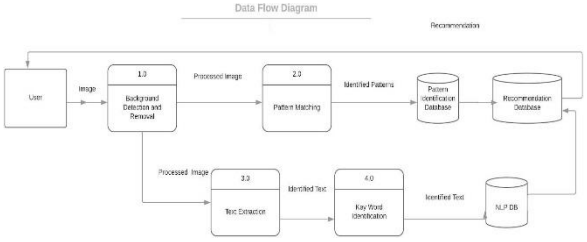


2. (2.2)Flowchart

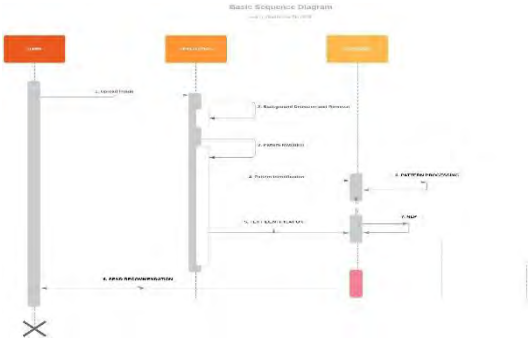
3. SEQUENCE DIAGRAM



4. CLASS DIAGRAM



5. DATA FLOW DIAGRAM



III. LITERATURE SURVEY

Technology has become an indispensable part of human life and with Industry 3.0 fully evolved and Industry 4.0 on the advent of basic computer literacy has become an unavoidable life skill. Yet, the biggest problem faced by the IT industry is difficulty for the consumer to troubleshoot basic issues. Thus, we have decided to simplify the process using ML and Image Recognition techniques.

As per the our survey, traditional methods such as windows troubleshooting techniques fall short when it comes to user convenience and this leads to wastage of time by organization such as tech support teams. It also creates a hindrance as it leads to avoidance in usage of new technologies.

IV. FEASIBILITY STUDY

A feasibility study is a preliminary study which investigates the information of prospective users and determines the resources requirements, costs, benefits and feasibility of proposed system. A feasibility study takes into account various constraints within which the system should be implemented and operated.

Thus we evaluated the feasibility of the system in terms of the following categories:

Technical feasibility:
Technology used in the front end: QtPy (Desktop App) Software Required: Anaconda3
Hardware Required: Google Cloud (AI Platform) Resources Required: Data Sets from various repositories Editor: Sublime

Economic feasibility:
A simple economic analysis which gives the actual comparison of costs and benefits are much more meaningful . Our aim is to achieve maximum accuracy from our recommendation system with the help of an adaptive environment. This will reduce the need of checking the problems manually, thus reducing efforts, time invested and ultimately cost of the whole process whilst increasing efficiency.
Market Survey: The coming revolution is spearheaded by image recognition and AI technologies. A key challenge for users of image recognition technology is finding the proper solution that will address their specific consumer goods applications. On an observation of the local requirements, recommendation systems is accessible, convenient and accurate addresses the daily technical issues hampering their routine work.

Reliability:
The system adapts according to previously provided results and expectations following a spiral methodology for requirement gathering, shortlisting and updation.

Scheduling feasibility:
Report about the researched work will be submitted by 26 March 2019.
A project will fail if it takes too long to be completed before it is useful. Typically, this means estimating how long the system will take to develop, and if it can be completed in a given period of time using some methods like payback period.

V. RESULT AND DISCUSSION

The recommendation system allows users to upload image of issue terminal and the system will provide the solution steps to the customers, more efficiently and effectively, without facing many hassles or disadvantages.The system helps the users to deal with their daily I.T. muddles with a click of a button on their device. The system can be accessed from any device with an active internet service. Thus, users can then be uninterrupted from their work for a long time, which is the current case.This project will be helpful to the peripheral device companies, as well as to the common people who want to get acquainted to technology. It will be helpful to peripheral device companies as it shall reduce the time and efforts invested by actual humans in detecting (and solving) an issue. It shall benefit the common man as the solution shall be customized as per their need and will require lesser investigation thus saving them time.

REFERENCES

[1] J. Mao, W. Xu, Y. Yang, J. Wang, A. Yuille, "Deep captioning with multimodal recurrent neural networks (m-RNN)", Proc. Int. Conf. Learn. Representations, 2015.
[2]https://www.marutitech.com/problems-solved-machine-learning (10/6/2018).
[3] K. S. Fu, Syntactic Pattern Recognition and Applications, Prentice-Hall, Englewood Cliffs, N.J., 1981.
[4] K. S. Fu and A. Rosenfeld, "Pattern Recognition and Image Processing," IEEE Trans. Computers, Vol. C-25, No. 12, Dec. 1976, pp. 1336-1346.
[5] Marilyn B. Troyer (1988) Issues and Problems in Teacher Computer Literacy Education, Journal of Research on Computing in Education, 21:2, 141-154, DOI: 10.1080/08886504.1988.10781867.

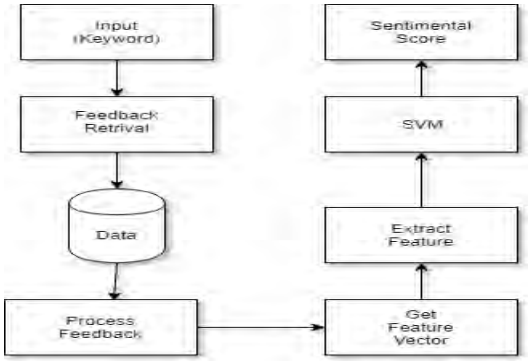
Design and Develop Sentiment Classifier Model for a Customer using Support Vector Machine

Rishabh Singh, Shreya Singh, Akshay Tripathi
Thakur College of Engineering & Technology Kandivali (E)
rishabhs343@gmail.com, shreyasingh267@gmail.com, akky16596@gmail.com

***Abstract*—Analysis of public information from social media could yield interesting results and insights into the world of public opinions about almost any product, service or personality. Social network data is one of the most effective and accurate indicatorsof public sentiment. The explosion of Web 2.0 has led to increased activity in Podcasting, Blogging, Tagging, Contributing to RSS, Social Bookmarking, and Social Networking. As a result there has been an eruption of interest in people to mine these vast resources of data for opinions. Sentiment Analysis or Opinion Mining is the computational treatment of opinions, sentiments and subjectivity of text. In this paper we will be discussing a methodology which allows utilization and interpretation of twitter data to determine public opinions. Developing a program for sentiment analysis is an approach to be used to computationally measure customers' perceptions. This paper reports on the design of a sentiment analysis, extracting and training a vast amount of tweets. Results classify customers' perspective via tweets into positive and negative, which is represented in a pie chart, bar diagram, scatter plot using php, css and html pages.**

I. INTRODUCTION

Sentiment is an attitude, thought, or judgment prompted by feeling. Sentiment analysis , which is also known as opinion mining, studies people's sentiments towards certain entities. Internet is a resourceful place with respect to sentiment information. From a user's perspective, people are able to post their own content through various social media, such as forums, micro-blogs, or online social networking sites. From a researcher's perspective, many social media sites release their application programming interfaces (APIs), prompting data collection and analysis by researchers and developers. For instance, Twitter currently has three different versions of APIs available [6], namely the REST API, the Search API, and the Streaming API. W ith the REST API, developers are able to gather status data and user information; the Search API allows developers to query specific Twitter content, whereas the Streaming API is able to collect Twitter content in real time. Moreover, developers can mix those APIs to create their own applications. Hence, sentiment analysis seems having a strong fundament with the support of massive online data.



However, those types of online data have several flaws that potentially hinder the process of sentiment analysis. The first flaw is that since people can freely post their own content, the quality of their opinions cannot be guaranteed. For e xa mple, instead of sharing topic-related opinions, online spammers post spam on forums.

Some spam are meaningless at all, while others have irrelevant opinions also known as fake opinions . The second flaw is that ground truth of such online data is not always available. A ground truth is more like a tag of a certain opinion, indicating whether the opinion is positive, negative, or neutral. The Stanford Sentiment 140 Tweet Corpus is one of the datasets that has ground truth and is also public available. The corpus contains 1.6 million machine-tagged Twitter messages.

Microblogging websites have evolved to become a source of varied kind of information. This is due to nature of micro blogs on which people post real time messages about their opinions on a variety of topics, discuss current issues, complain, and exp ress positive sentiment for products they use in daily life. In fact, companies manufacturing such products have started to poll these microblogs to get a sense of general sentiment for their product. Many time these companies study user reactions and reply to users on microblogs. One challenge is to build technology to detect and summarize an overall sentiment.

Our project Tweezer resembles the analyze of tweets by the peoples on certain products of companies or brands or performed by political leaders. In order to do this we analyzed tweets from Twitter. Tweets are a reliable source of information mainly because people tweet about anything and everything they do including buying new products and reviewing them. Besides, all tweets contain hash tags which make identifying relevant tweets a simple task. A number of research works has already been done on twitter data. Most of which mainly demonstrates how useful this information is to predict various outcomes. Our current research deals with outcome prediction and e xplores localized outcomes.

We collected data using the Twitter public API which allows developers to extract tweets from twitter

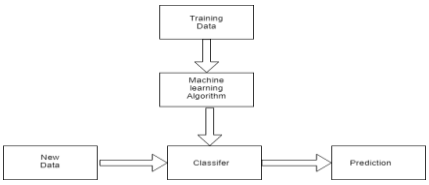
programmatically.

The collected data, because of the random and casual nature of tweeting, need to be filtered to remove unnecessary information. Filtering out these and other problematic tweets such as redundant ones, and ones with no proper sentences was done next.

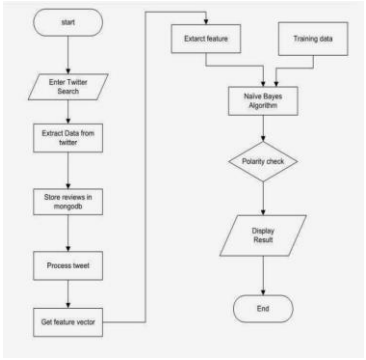
As the preprocessing phase was done in certain e xtent it was possible to guarantee that analyzing these filtered tweets will give reliable results. Twitter does not provide the gender as a query parameter so it is not possible to obtain the gender of a user from his or her tweets. It turned out that twitter does not ask for user gender while opening an account so that information is seemingly unavailable.

II. PROPOSED WORK

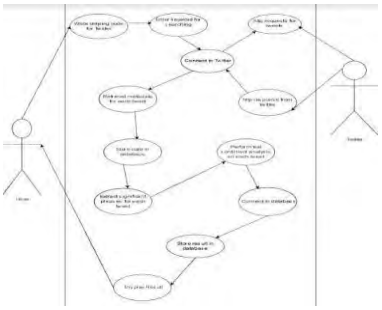
I. (2.1) System Architecture



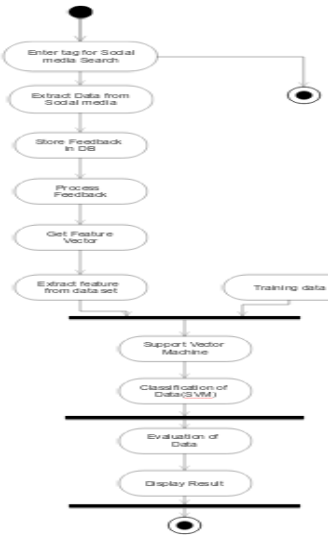
II. (2.2)Flowchart



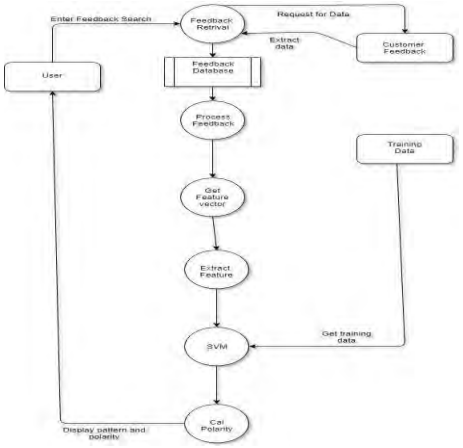
III. USE CASE DIAGRAM



IV. ACT IVITY DIAGRAM



V. DATA FLOW DIAGRAM



III. LITERATURE SURVEY

Sr.no 1

Title: Sentiment Analysis: A Survey of Current Research and Techniques

Author: Jeevanandam Jotheeswaran, Dr. S. Koteeswaran

Description: Analyzing the polarity of sentiment e xpressed in data is Opinion Mining (OM). It is a system that identifies and classifies opinion/sentiment as represented in electronic te xt.

De vel opment: Opinion Mining (OM) is natural language processing dealing with tracking the mood of people regarding a product or topic. OM combines information retrieval and computational linguistic techniques handling a document's opinions. The field's main goal is solving problems related to opinions on products, politics in newsgroup posts and review sites. It provides automatic opinion e xtraction, emotions and sentiments in text tracking attitudes and feelings on the web. People e xpress views by writing blog posts, comments, reviews and tweets on various topics. Tracking products and

brands and determining if they are viewed positively or negatively is done on the web.

Feature: Using OM, a review is evaluated at 3 levels - document, sentence and feature levels. The entire review is classified positive or negative based on opinions e xpressed in that review. The major feature based OM tasks include - (1) identifying products features in review, (2) determining opinion e xpressed by reviewer (positive, negative or neutral), (3) summarizing discovered information

Sr.no 2

Title: Classification of Sentiment Analysis on Tweets using Machine Learning Techniques

Author: Shivaraju Kethavath

Description: Sentiment analysis classification and preprocessing of twitter data. We use large number of training datasets of tweets to find polarity by using machine learning techniques. We process the twitter dataset with feature of unigram and bigram and test them with substantiated results that have been found on twitter dataset. Finally we get results about which classifier is the best

De vel opment: Ext ract opinion words referring to the target feature The opinion words are not only Adjectives like hate, love but also consist of other POS categories like Nouns (terrorism), Verbs (terrify) and Adverbs (gratefully). A nave method, like e xt racting the opinion words closest to the target feature, does not work so well when the sentence has multiple features and distributed emotions (as we will see later).

Feature:
Twitter Dataset:
Micro blogging stages are utilized by distinctive individuals to e xpress their opin
ions about diverse themes, accordingly it is an important wellspring of individuals'
Sentence Weightage:
If a tweet consists of more than one sentence, we give more weightage to sentences coming afterwards
Abbreviations and Redundant/Repeated letters :
Due to the easygoing way of Twitter dialect, a few words (as a rule conclusion words) are incorrectly spelled

or frequently over underscored because of which the classifier may not quality polarity of this word (eg. looooooooove) to the genuine word (eg.love) during training.

Sr.no 3

Title: Customer Feedback Analysis

Author: Somnath Banerjee, Riyanka Manna and Dipankar Das

Description: In general perspective, this task is a classification problem. In a global multilingual environment, the two main challenges for international companies (such as Microsoft) to automatically detect the meanings of customer feedback are: i) no widely acknowledged classes for understanding the meanings for customer feedback, ii) the classification may not be applicable in multiple languages.

De vel opment: The organizer of the customer feedback analysis provided the participants customer feedback sentences which were collected from Microsoft Office customers as part of the joint ADAPTMicrosoft research project.

Feature: We pre-processed the dataset provided for building the system. Init ially, we removed the index from the data samples. In the provided dataset, a number of data samples were tagged with multiple classes. If a data sample was tagged with multiple tags, the data sample was modified. The following exa mple shows the processing of a data sample tagged with two tags for training as a single tagged.

Original: Renovations underway...dated. ⇒comment, complaint After pre-processing:
Renovations underway...is dated. ⇒comment Renovations underway...is dated. ⇒complaint

IV. FEASIBILITY STUDY

A **feasibility study** is a preliminary study which investigates the information of prospective users and determines the resources requirements, costs, benefits and feasibility of proposed system. A feasibility study takes into

account various constraints within which the system should be implemented and operated.
Thus we evaluated the feasibility of the system in terms of the following categories:

Technical feasibility:

The machine learning approach is further useful in opinion mining than other approach due to fully automatic completion and its capability to hold large collection of Web Data. Machine Learning-Based Sentiment classification techniques can be classified into three types: supervised, unsupervised, and semi-supervised learning techniques. Aim to achieve the ma ximu m accurate feedback from customer end, solve the problem in give time using appropriate algorithm such as SVM

Economic feasibility:

A simple economic analysis which gives the actual comparison of costs and benefits are much more meaningful
. Our aim is to achieve maximum accuracy in our machine learning processing of the customer feedbacks. This will reduce the need of checking the feedbacks manually, thus reducing efforts, time invested and ultimately cost of the whole process whilst increasing efficiency.

Legal feasibility:

Data on social media which is publicly will be e xt racted so that we can apply machine learning algorithms on it and classify it according to it's sentiment. The proposed was to

make a simplified web application. It is simpler to operate and can be used in any webpages. It is free and not costly to operate.

Scheduling feasibility:

Report about the researched work will be submitted by 26 March 2019.
A project will fail if it takes too long to be completed before it is useful. Typically, this means estimating how long the system will take to develop, and if it can be completed in a given period of time using some methods like payback period.

V. RESULT AND DISCUSSION

By doing sentiment analysis on the customer feedback, an organization can find the important reactions to one's product in a simple and easy manner. Also, since the result is sorting the sentiment analysis of feedbacks according to the topics, it makes it easier for the organization to understand which topics need more attention and work or are getting a negative response. Our project would also help an organization in finding out the profitable areas as well which could be further enhanced to make it more profitable whilst working towards addressing public's opinion as well. Overall, our project would help an organization take the necessary steps for further upgrading their products or policies. Hence, this will in result make an impact over the whole society throughout, people will understand that their genuine feedbacks are given importance which would urge them to participate in it more actively. Our project would help the organizations in working towards the issues of consumers thus working for the betterment for the society and catering to their personal needs as well.

VI. CONCLUSION

We have completed our project using python as language, Php with Html and Javascript for output presentation. Although there was a problem in integration of python and php, through numbers of tutorial we were able to integrate it. We were able to determine the positivity and negativity of each Social Media Data. Based on those Data we represented them in a diagrams like Bar graph, Pie-chart and scatter plot. A small conclusion is also shown during output presentation based on product or brand entered. Our designed system is user

friendly.

All displaying results are displayed in webpage.

REFERENCESs

[1].Nan, L.a.D., D., Using te xt mining and sentiment analysis for online forums hotspot detection and forecast. Decision Support Systems archive. Volu me 48 Issue 2, January, 2010. Pages 354-368. Elsevier Science Publishers B. V. Amsterdam, The Netherlands, The Netherlands, 2010.

[2].Ian H. Witten, Eibe Frank & Mark A. Hall. Data Mining – Practical Machine Learning Tools and Techniques.

[3].Steven Bird, Even Klein & Edward Loper. Natural Language Processing with Python.

[4].Albert Biffet and Eibe Frank. Sentiment Knowledge Discovery in Twitter Streaming Data. Discovery Science, Lecture Notes in Computer Science, 2010, Vo lume 6332/2010, 1-15, DOI: 10.1007/978-3-642-16184-1_1

[5].Stefano Baccianella, Andrea Esuli, Fabrizio Sebastiani. SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In Proceedings of international conference on Language Resources and Evaluation (LREC),2010.

[6].Jurka,T.,Sentiment:ToolsforSentimentAnalysis.URLhttp: //CRAN.Rproject.org/package=sentiment. R package version 0.1., 2012.

[7].Roever C, R.N., Luebke K, Ligges U, "kla R – Classification and Visualization.". Ve rsion 0.4 -1. URL http://CRAN.R-project.org/. 2005. 8. Hsu CW, L.C., A Simple Decomposition Method for Support Vector Machines. Machine Learning, 46, 291–314. URL

[8].Ale xander Pak and Patrick Paroubek. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In Proceedings of international conference on Language Resources and Evaluation (LREC),2010.

[9].Ricgard O. Duda, Peter E. Hart & David G. Stork: Pattern Classification.

[10].Hatzivassiloglou, V., & McKeown, K.R.. Predicting the semantic orientation of adjectives. In Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL, 2009.

Visually Aware Fashion Recommendation System

Anushi Doshi , Jaimish Ashar, Shridhar Kamble Assistant Professor
Department of IT
Thakur College of Engineering and Technology Mumbai, India
doshianushi007@gmail.com, jaimish97@gmail.com, shridharkamble1@gmail.com

Abstract — In recent times, clothing recommendation systems and product images have become of increasing importance as the visual appearance of these clothes plays a major role in the consumers' decision to purchase that outfit. The primary challenge with building fashion recommender systems is the complexity and inability to quantify "fashionable" features of an outfit. Convolutional Neural Networks (CNN) extracts certain visual features. However, these cannot accurately determine user preferences since "aesthetic features" are rarely considered. To improve recommendation performance, it is necessary to use image feedback and train the recommender system simultaneously. This can be achieved using Siamese CNNs. The model shows improved performance over the already optimized Bayesian Personalized Ranking (BPR) method by using Deep Visually Aware Bayesian Personalized Ranking (DVBPR). Matrix factorization is used to extract visual features based on certain latent factors. These latent factors can be thought of as the 'preferences' of the user and the 'properties' of an item, such that a user is likely to interact with an item if their preferences are compatible with its properties. Other than images, content- and context aware models that make use of a variety of information sources, including temporal dynamics, and geographic information have been developed. A number of approaches to include deep learning into recommender systems have also been proposed such as content-based systems. YouTube video recommendation works on this principle. The most closely related work is a recent method that extended BPR-MF to incorporate visual dimensions to facilitate item recommendation tasks. An 'end-to-end' learning approach (rather than pre-trained features) like the one used in this project is shown to perform better than the traditional approach. Other proposed methods were not as accurate as DVBPR in terms of preference score, quality and diversity. Better performance is achieved by 5.13% across all datasets and 2.75% in cold start scenarios.

I. INTRODUCTION

The quantity of decisions on the web is gigantic and there is a need to channel, organize and proficiently convey important data with the end goal of reducing the issue of data over-burden. This is potential issue to numerous Internet users. This issue is comprehended with the utilization of recommender frameworks. They do this via looking through substantial volume of progressively created data to give users customized substance and options. This eruptive development in the measure of accessible computerized data and the quantity of guests to the Internet have created a potential test of data over- burden which obstructs timely access to items of interest on the Internet. Information gathering frameworks, have in part tackled this issue but prioritization and personalization of data were missing. This has expanded the interest for recommender frameworks like never before. Recommender frameworks are data separating frameworks that deal with the issue of data over-burden by sifting crucial data as per the user's inclinations, interest, or behavior about that particular item. Recommender frameworks can foresee whether a specific user would incline towards purchasing an item or not based on the profile. Recommender frameworks are valuable to both service providers and users. They lessen the expenses of finding and choosing things in a web based shopping environment. Suggestion frameworks have likewise demonstrated to enhance basic decision making and quality. In an internet business setting, recommender frameworks drive up sales, for the way that they are successful methods for offering more items. In scientific libraries, recommender frameworks bolster users by enabling them to move past catalog searches. In this way, the need to utilize effective and precise suggestion methods inside a framework that will give pertinent and tried and true proposals to users can't be over-emphasized.

II. PROPOSED WORK

Our recommender framework will give customized proposals to users, in light of expansive volumes of historical input, by revealing concealed measurements that depict the preferences of users and the properties of the things they desire. Traditionally, this implies preparing predictive algorithms that can distinguish things that are probably going to be tapped on, acquired, or given a high rating. In areas like design, this can be especially trying for various reasons: the vocabulary of things is long-tailed and new things are continually being introduced; users' preferences and item styles change over time; and the semantics that figure out what is 'fashionable' are unbelievably complex.

We seek to use pixel level image content to build recommender systems. This follows the recent trend of incorporating representation learning techniques into recommender systems. We adapt the popular formulation from Bayesian Personalized Ranking (BPR)

to include image content via a Siamese net and show significant improvements over BPR itself.

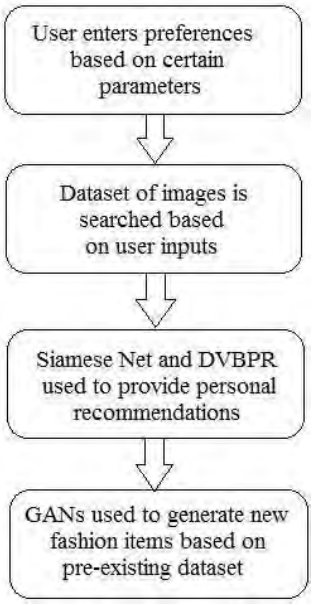


Figure 1 – Flowchart for Recommendation System

III. LITERATURE SURVEY

Lei and Li has designed a dual net deep network in which two sub networks map input images but also preferences of users into a same latent semantic space, and then the distances between image and users in the latent space are calculated to make decisions. A CDL method to train the deep network is proposed, using a pair of images compared against one user to learn the pattern of their relative distances.[1]

According to Goodfellow a framework for predicting generative models can be done via an adversarial process, in which two models are trained together. This includes a generator that captures the data distribution and a discriminator that estimates the probability that a sample came from training data rather than the generative model.[2]

Al-Halah and Stiefelhagen has discovered an approach to predict the future popularity of styles is presented from fashion images in an unsupervised manner. Using these styles as a basis, a forecasting model is trained. This represents trends over a temporal space. The model can synthesise new combinations of styles that will become popular in the future.[3]

Yu and Zang takes into consideration the impact on aesthetic preferences over time and proposes a solution in which the temporal factors are taken into account by a context-aware system which will be used as the basic

model. A study of the limitations of the existing systems has led them to create a model in which they train the data by enumerating all positive results in every iteration to optimize the model. [4]

IV. FEASIBILITY STUDY

Economic Feasibility (EFS): The cost of training the data in terms of time is directly proportional to the amount of hours an infrastructure is rented. If the required GPU is unavailable, there will be significant cost to rent a cloud based multi-core GPU using Crestle, PaperSpace, Google Cloud etc. However, the datasets will be freely available or web scraped and will incur no extra cost.

Technical Feasibility: Hardware requirements will include, ideally an Nvidia 1080i GPU. A quad core or higher core CPU will be preferred in order to train the data efficiently and quickly. 4+ GB of RAM will be ideal. The required software is all available on the Open Source Network. Ideally, a Linux distribution will be required as an operating environment. Python Machine Learning Libraries are available for free online.

Operational: The most intensive task will be the simultaneous training of data using Siamese Nets. This will require majority of the processing and GPU compute power. If worked on to transform into a full-fledged application or product, it will be able to run on commodity hardware.

Socio-cultural: This platform is intended to streamline the users' decision in purchasing fashion items based on a visually-aware recommender system. The user will be able to generate his/her own clothing items if needed. The user will effectively spend less time looking for items online as the system will automatically look for new items based on the user preferences. This will allow the user to focus on other tasks.

Legal: Initially, the project will be available only in terms of a raw implementation. As such, it will face no legal ramifications. However, if designed into a full product and combined with the commerce systems of online retailers, legalities and commission fees may need to be sorted out.

Ethical: The main concern here is the invasion of private purchase data of the consumers. To train as well as operate the system, users will have to agree to share their private purchase data in order for the system to work effectively. Appropriate safeguards will be put into place to ensure the complete privacy of the user data from external entities.

V. RESULTS AND DISCUSSION

Achieved results:
We have implemented a prototype for the ***Next?*** section. We have prepared a database of 72 images with limited

parameters as this is just a prototype. Images will be displayed based on the user's input choices: {Gender, type, style, pattern}

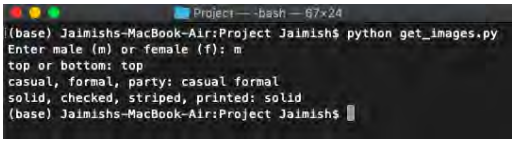


Figure 2 – Command Line

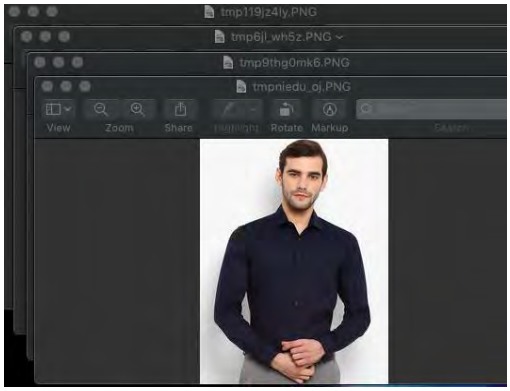
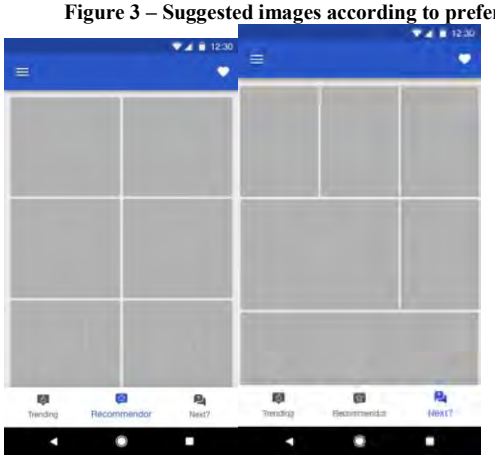


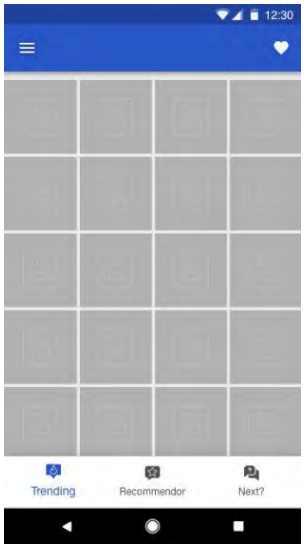
Figure 3 – Suggested images according to preferences



User Interface Design:

We have designed the basic user interface for our system. It includes but is not limited to – Trending outfits, recommender according to user preferences and suggestions for what the user can wear next. ***Trending*** tracks the most liked items, which can be used by retailers to create new designs. ***Recommender*** recommends new items to the user based on her/her liked items and preferences. ***What's next?*** acts as a fashion assistance for the user so as to guide him/her on what to wear next.

Next? (b) Recommender



(c) Trending Figure 4

Expected Results:

Sampled images from each of six categories (women/men's tops, bottoms and shoes) are compared to their nearest neighbors in the dataset. This demonstrates that the generated images are realistic and plausible, yet are quite different from any images in the original dataset

- they have common shape and color profiles, but quite different styles



A-Generated Images B- Nearest Neighbors Figure 5

Fashion design based on a prototype:

The second type of personalized design consists of making small modifications to existing items so that they more closely match users' preferences. This figure shows the results of this optimization process for six users based on a given t-shirt or pair of pants. [5]

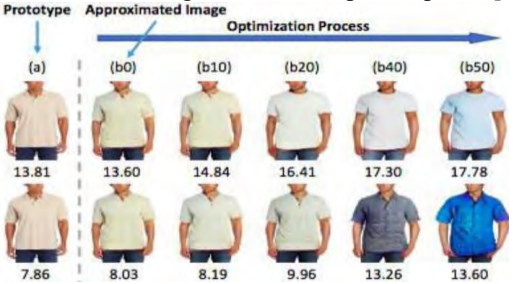


Figure 6 – Optimized Images

Discussions:

There are a few relevant observations to be made: First, the optimization process substantially increases the

preference score over that of the original image. Second, the L1- approximated images are similar to the given prototype, and most images are in the vicinity of the original image. Third, the process highlights the continuous nature of the space learned by the GAN. The types of modifications vary for different users, including but not limited to changing color, extending sleeve length, 'distressing' pants, shortening pants and other minor stylistic changes.

VI. CONCLUSION AND FUTURE SCOPE

- User preferences and tastes are ever changing and the volume of fashion items is increasing by the day. To cater to this enormous market and user requirement, a number of recommendation systems have been proposed and already exist in the market. However, due to certain underlying issues and after understanding the user with greater clarity, this novel approach was proposed.

- The proposed system is capable not only of suggesting existing items to a user, but is also capable of generating novel, acceptable fashion images that match user preferences. This suggests a new type of recommendation approach that can be used for both prediction and design. The usage of a Siamese Net employs and end-to-end learning approach about which we talked earlier and this provides the performance improvements.

- The interesting idea with this system is the usage of Generative Adversarial Nets (GANs) which was employed to use an existing image database to generate new, novel, similar pattern fashion items based on user preferences This framework can be used in a variety of recommendation scenarios, to explore the space of possible fashion items, to modify existing items, or to generate items tailored to an individual. It can also be of use to retailers and fashion designers.

Later on, we trust this opens up a promising line of work in utilizing recommender frameworks for design. Other than enhancing the nature of the created pictures and giving control of fine-grained styles, similar ideas can be applied to visual information other than fashion images,

or even to non-visual types of content. We trust that such systems can prompt more extravagant types of suggestion, where content recommendation and generation are all the more firmly connected.

A huge dataset for product aesthetic assessment can be set up after which, the systems can be prepared to extract the stylish data better. In addition, we will examine the viability of our proposed technique in the setting of explicit feedback. In conclusion, we are interested in integrating the domain knowledge about aesthetic assessment, e.g., in the form of decision rules into the recommender model.

REFERENCES

[1] Comparative Deep Learning of Hybrid Representations for Image Recommendations - Chenyi Lei, Dong Liu, Weiping Li, Zheng-Jun Zha, Houqiang Li

[2] Generative Adversarial Nets - Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde Farley, SherjilOzair, Aaron Courville, YoshuaBengio

[3] Fashion Forward - Forecasting Visual Style in Fashion: Ziad Al-Halah, Rainer Stiefelhagen, Kristen Grauman

[4] Aesthetic-based Clothing Recommendation - Wenhui Yu, Huidi Zhang, Xiangnan He, Xu Chen, Li Xiong,
Zheng

[5] Visually-aware and fashion recommendation and design with generative Images Model - Wang-Cheng Kang, Chen Fang, Zhaowen Wang, Julian McAuley

[6] Algorithmic clothing: Hybrid recommendation, from street-style-to-shop - B Sengupta, Y Qian, P Giaccone M Sasdelli, E Vazquez

[7] Trip outfits advisor: Location-oriented clothing recommendation - Xishan Zhang, Jia Jia, Ke Gao, Yongdong Zhang, Dongming Zhang, Jintao Li and Qi Tian

[8] Fashion Recommendation using text mining and multiple content attributes - Wei Zhou, Yanghong Zhou,
Runze Li and P.Y. MOK

[9] Toward Fashion-Brand Recommendation Systems Using Deep-Learning: Preliminary Analysis - Yuka Wakita, Kenta Oku, and Kyoji Kawagoe

[10] Intelligent Fashion Recommender System: Fuzzy Logic in Personalized Garment Design - L. C. Wang, X.

Implementation of Stock Market Prediction Using Regression

Ms. Ishita Gupta1, Mr. Anay Awasthi 2, Mr. Sagar Gour 3, Dr. Sheetal Rathi4
1, 2, 3Student, 4Associate Professor
Department of Computer Engineering Thakur College of Engineering & Technology
Mumbai—400101, India

***Abstract*— The stock market is a lucrative and interesting way to earn money. It revolves around buying and selling of stocks in order to generate profits. Many investment banks, hedge funds, etc. employ a lot of developers, quants, and data scientists to make models to predict market movement, compute risk to reward ratios, etc. for their positions to decide the correct amount of exposure for each of their trade. These strategies and mathematical models are not publicly shared so that it doesn't lose its uniqueness. Currently, the research papers do not use live data for trading and hence no tangible success/failure of the project can be determined. This creates a need for a project that uses real life and live data and applies mathematical models and machine learning algorithms on it to gain tangible results (Profits or Losses). The project focuses on assisting stock market traders by helping them make profit. The project works by fetching/streaming data from the National Stock Exchange (NSE) for a given stock and then applying it to linear regression model. This way, the movement of the stock can be predicted. The project also makes use of various statistical indicators such as Volume Weighted Average, Relative Strength Index, etc. It also considers Price Action strategies (such as previous day volume) in order to make a better choice. It also considers sentiment analysis for the particular stock using websites like moneycontrol. The project will then execute trades automatically and track its position and exit whenever the trade goes against the predicted outcome or when the exposure is increased above a certain capital. Finally, all the trades are tracked and a timeline or a line graph is created to track the performance of the application over a given range of time.**

I. INTRODUCTION

The stock market can be a great stream to generate supplementary or side income. The project will focus on short term (intraday) trading and apply neural networks and machine learning techniques to decide when to buy and sell a particular stock. There are a few publicly applications available that predict (or claim to) the stock market. The application would use a combination of strict risk: reward ratios, machine learning and statistics

to trade in the Indian Stock Market and analyze its efficiency. Also, primary objective is to minimize the loss per investment done by the investor.

II. LITERATURE SURVEY

1. Stock market prediction: A big data approach [1]:
In this paper, both technical and fundamental approach is considered. Fundamental analysis is done using social media data by applying sentiment analysis. Social media has high impact on our lives today. This method includes collecting news from social media and then extracting sentiments expressed by that individual's post. Technical analysis is done using historical data of stock prices by applying machine learning.

2. Performance evaluation of ANN and neuro Fuzzy system in business forecasting [2]:
Majorly, two prediction techniques have been compared in this paper i.e. Adaptive Neuro Fuzzy inference system (ANFIS) and Artificial Neural Network (ANN). To check the performance, they have considered two examples. First, they take into consideration the sales dataset of cold drinks, the cold drink sales are computed by the system and results are generated. Second is the prediction done by the two techniques on the stock market, the data is taken from the NSE which is provided by the broker. Based on this live dataset, optimal results are generated to achieve successful trades.

3. A methodology for stock market analysis utilizing rough set theory [3]:
The behavior between market and the investor is simple randomization which is difficult for the AI/Neural Network to analyze. In this paper they have described about rough set theory which is applied in AI so, that the AI could work efficiently in a dynamic environment. Basically, the normal stock market prediction is based on general quants. The problem for AI is to interpret these quants and also apply its own strategy based on learning. So, they have implemented rough set theory which is the overall collection of all the random actions happened in day to day stock

exchange and maintained in a dynamic database. Using this database, the AI will be able predict the stock market without any ambiguity.

4. Stock Market Prediction of S&P500 Using BFO Technique [4]:
The present paper introduces the bacterial foraging optimization (BFO) technique to develop an efficient forecasting model for prediction of various stock indices. The connecting weights of the adaptive linear combiner-based model are optimized by the BFO so that its mean square error (MSE) is minimized. The short and long-term prediction performance of the model is evaluated with test data and the results obtained are compared with those obtained from the multilayer perceptron (MLP) based model.

5. Survey of Stock Market Prediction Using Machine Learning Approach [5]:
The paper talks about non-linear nature of the Stock market. It also talks about people's sentiments and their approach. It talks about predicting the market using tips from other websites or brokers without any solid reason backing a particular prediction. In this paper well-known efficient regression approaches are explored to predict the stock market price from stock market data based. A gap or improvement in the paper can be that the results of multiple regression approach could be improved using more number of variables.

6. Stock Market Prediction Using Linear Regression [6]:
Linear Regression is applied on a publicly available dataset which contains the closing price of the shares of Tata Consultancy Services(TCS). The paper focuses on providing the investors and corporate stakeholders with a method to forecast daily behavior of stock market. It also analyzes different regression algorithms and concludes that linear regression was able to fit the data in the most precise manner and hence shall be preferred instead of other algorithms.

7. How to Predict Stock Prices Easily [7]:
The video is an introduction to Deep Learning and helps in the understanding of how to create models for predicting the stock market and how to check their efficiency. However, it needs to be noted here that since the signal to noise ratio is very low, the model ends up fitting noise and not the signal and hence cannot be expected to be correct all the time.

7. Things You Learn After 1 Year of Day Trading for a Living [8]:

The article talks about the struggles of day trading as a career. Also, it explains that the signal to noise ratio for stocks is very low. Hence, no system is accurate a 100% of time. Also, sometimes, the more intricate and complex the system, the less accurate it becomes. The author, Mr. Andrew Kreimer explains the different intricacies of trying to predict a second-degree chaos system such as the Stock Market.

III. PROPOSED SYSTEM

The system will use linear regression, statistics and money management to predict the price movement of selected stocks in the Indian Stock Market. Then, it will automatically place the given trades at the given price with a strict take profit and stop loss. Also, the application will have dashboard to view statistics and analyze overall profit and loss over a given period of time. This will enable the trader to analyze the correct and wrong trades and also for tax purposes.

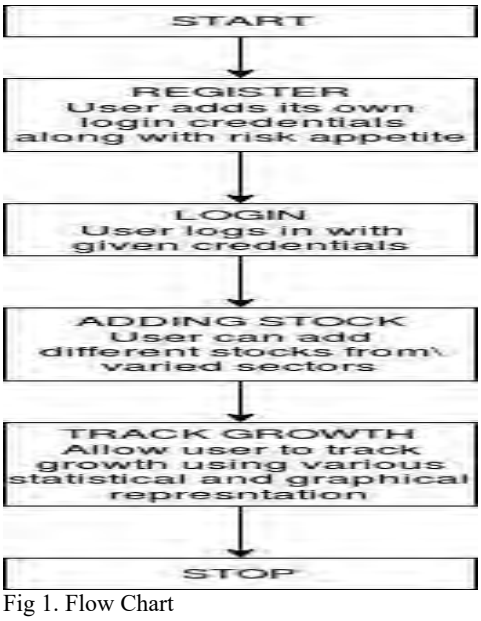


Fig 1. Flow Chart

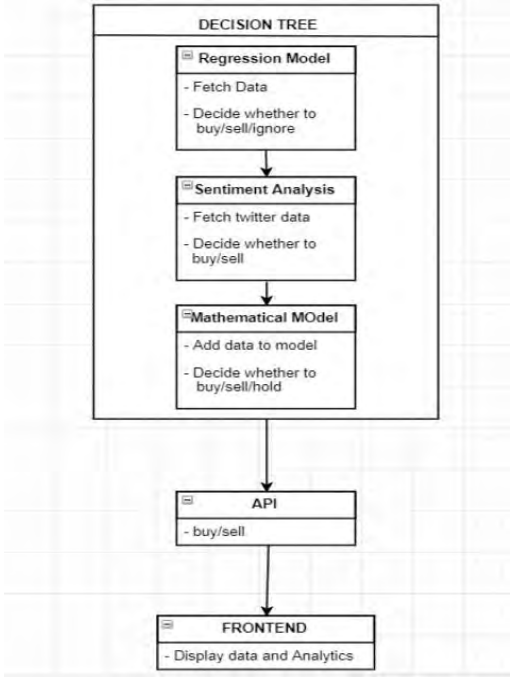


Fig. 2. Block Diagram of the Proposed System

The user registers to the system using their Email Address/Username, Password, Demat Account Details. Then, the user logs in. They are then presented with a dashboard through which they can track the performance for the past day and their profit/loss for the day. Also, they can add new stocks which will be monitored. It will also have various analytics tools to track growth of capital and profit realization over time. On the backend side, various stocks are selected and monitored. Once a few conditions of the model are satisfied, it will apply sentiment analysis on stock news which can be taken from websites such as moneycontrol and twitter followed by regression. This three-way confirmation would not only improve probability of success but filter out wrong signals and false positives.
The system starts by generating an access token early morning. Then, after the market opens, the system waits till 10 am to scrap the National Stock Exchange of India's website using Selenium, it then creates a list of the top gainers and losers of the day. After that, it uses the Upstox API to get the candlestick data for the past 7 days of the stocks shortlisted. It then, analyses whether various statistical, regressional and sentimental conditions of the selected stock and then shortlists the best candidate for the trade. Finally, the system uses the

API to automatically place a trade in the market. The process repeats at 1 o'clock in the afternoon as well.

IV. METHODOLOGY

The project uses the popular SEMMA model developed by SAS Institute. SEMMA is an acronym for Sample- Explore-Modify-Model-Assess.
This allows the team to divide the task into different parts and focus on developing a fast and minimum viable product instead of a monolith without proper testing and performance overview.
The process can further be dealt in an agile manner where updates and improvements are continuously deployed based on new and relevant findings.

V. TECHNOLOGY

The project focusses on Machine Learning and AI and hence would require Python along with Scikit-learn library for the development of Machine Learning Models and linear regression algorithm. NLTK (Natural Language Toolkit) library would be used for sentiment analysis using Naïve-Bayes algorithm from twitter/moneycontrol data. Also, the project requires heavy use of statistics for the creation of mathematical models along with an emphasis on a large dataset which would require usage of the Numpy library for python to implement a Mathematical Model. For scraping data for sentiment analysis, tweepy library would be used for getting tweets and beautiful soup 4 for scraping data from other websites. Due to high cost of training models, the best and economical solution would be to use a cloud GPU such as the one provided by Google under the "Colaboratory" Project.

VI. FEASIBILITY STUDY

I. Technical Feasibility:

Technology used at front end: CSS, bootstrap, HTML, Materialize CSS, jQuery
Technology used at Back end: Python
Resources Required: Manpower, Programmers, testers, debuggers
Software required: Testing Tools (to perform black box and white box testing),
GitHub
Editor: Sublime Text IDE: PyCharm
Libraries: scikit-learn, Numpy, etc.
Hardware required: PC for development, server for deployment
Infrastructure: For scaling up the project.

II. Operational Feasibility:

The proposed system is concerned with minimizing the losses incurred by traders and also automating their trading setups in the stock market. The project will have a frontend user interface which will allow the user to see which trades are being taken on their behalf and also show the current profit and loss for any trade. It will also have settings to optimize the risk appetite, frequency, etc. The backend of the application will use a custom prediction model based on sentiment analysis and Linear Regression algorithm to predict the price of a stock. Multiple algorithms might further reduce the number of false positives thereby increasing the profit. Based on the prediction, the system will automatically place the trade order and monitor it for a change of trend and exit the trade when required.

VII. RESULT AND DISCUSSION

The results of the project can be measured in terms of Return on Investment (ROI) from the initial amount over the period of a month, quarter, or an year. The project focusses on small quantities and retailer-level positions in the market instead of operator level positions. Hence, the target audience is youngsters who want to invest their pocket money or other small savings for a short period of time, or new traders who want to trade in small quantities. The measurable return can be further compared with other indexed funds, mutual funds, and even the entire stock market itself (For example, the returns given by Nifty over a 1-month period). The following is some preliminary result that was obtained upon applying Linear Regression on the closing prices of the stock Vedanta Ltd. over the timeframe of an year:

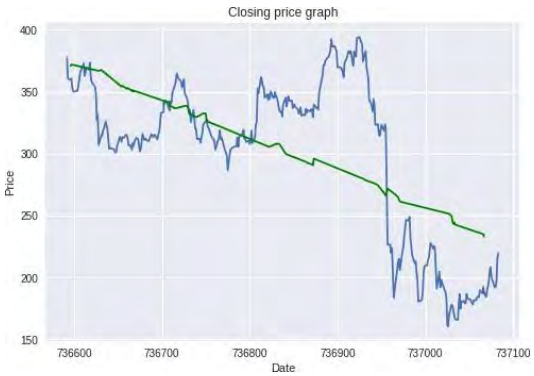


Fig. 3. The closing price for the stock Vedanta Ltd.

As it can be inferred from the above image, the line in blue represents the actual price of the stock whereas, the green line represents the predicted price. It can be clearly

understood that linear regression with time and stock price as the two related parameters do not exactly trace the actual stock price. However, upon clear observation, it can be inferred that the predicted price does not follow the actual price but rather behaves the same. The rises and falls in the actual price are clearly in unison with the rises and falls in the predicted price. This would help predict the direction in which the trade needs to be taken rather than finding an exact, quantifiable value of the stock price.

Upon further research, the VWAP indicator (Volume weighted average price) was selected due to its close relation with the stock price and volume of the stock. It was further observed that upon crossing of the VWAP from either direction by the stock price, the stock seems to move further in that particular direction for long enough to make it a profitable trade. Also, the stock seems to struggle to move against that direction beyond the VWAP. This allows a comfortable support for the stock loss to not be hit.



Fig. 4. 5-minute Chart of Yes Bank Ltd.

In the above figure, it can be seen that at points 1,2 and 3, when the stock closed below the VWAP, the stock further rallied down and would be a shorting opportunity.

VIII. FUTURE SCOPE

The project is limited to small, retail-level investments and similar to any system, is not scalable beyond a specific capital

amount. The project could be modified further to incorporate other brokers such as Zerodha. Currently, only Upstox is supported due to its API access. Also, multiple strategies and machine learning, deep learning models can be applied on the different stocks and back-tested to find which stock follows which strategy. Then, while trading, the strategies can be hot-swapped based on the stock that is selected in the watchlist. Further, the strategy can also be applied to derivative trading (such as futures) instead of being limited to only equities as and when the position size increases. Legal feasibility of a project should also be studied before mass roll-out is done.

IX. CONCLUSION

The project would not only benefit new traders but also those who want to trade in the Indian Stock Market but cannot do so because of their full-time job. This platform allows them to not just venture into the world of intraday trading and make profits which can be used as a secondary income source or as contingency or retirement funds. The project can also be flexibly used to check different strategies by modifying conditions in the decision center and also on tracking the overall growth of capital over a given span of time. From a Data Science standpoint, the project can be thought of as a great avenue into the world of statistics, machine learning and artificial intelligence building the foundation for further development and research.

REFERENCES

[1] Girija V Attigeri, ManoharaPai M M, Radhika M Pai, Aparna Nayak, “Stock market prediction: A big data approach”, Tencon 2015- 2015 IEEE Region 10 Conference, Macao, China, November 2015.

[2] Sharifa Rajab and Vinod Sharma, “Performance Evaluation of ANN and Neuro-Fuzzy system in business forecasting”, 2015 2nd International Conference on Computing for Sustainable Global Development(INDIACom), New Delhi, India, March 2015.

[3] R.H. Golan and W. Ziarko, “A Methodology for Stock market analysis utilizing rough set theory”, IEEE Conference on Computational Intelligence for Financial Engineering(CIFEr), New York, NY, USA, April 1995.

[4] Ritanjah Majhi, G. Panda, G. Sahoo, P.K. Dash and D.P. Das, “Stock Market Prediction of S&P500 using BFO Technique”, 2007 IEEE Congress on Evolutionary Computation, Singapore, Singapore, September 2007.

[5] Ashish Sharma, Dinesh Bhuriya, Upendra Singh, “Survey of Stock market prediction using Machine Learning Approach”, 2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, April 2017.

[6] Dinesh Bhuriya, Girish Kaushal, Ashish Sharma, Upendra Singh, “Stock market prediction using a linear regression”, 2017 International Conference of Electronics, Communication and Aerospace Technology(ICECA), Coimbatore, India, April 2017.

[7] https://www.youtube.com/watch?v=ftMq5ps503w (Visited on 30th December, 2018)

[8] https://blog.usejournal.com/things-you-learn-after- 1-year-of-day-trading-for-a-living-a97bbc8d19fa (Visited on 2nd January, 2019)

Ensuring Safety of an Individual by using EEG and Machine Learning

Prof. Dr. Sharmishtha Desai, Prof. Shamla Mantri, Adesh Raut, Kshitij Nirgude, Mrunali Ruikar, Aarohi Mali

Computer Engineering MIT College of Engineering Pune, India

sharmista.desai@mitcoe.edu.in, shamla.mantri@mitcoe.edu.in, rautadesh98@gmail.com, nirgudekshitij@gmail.com, mali.arohi@gmail.com, mrunalivruikar@gmail.com

***Abstract*— Brain–computer interface (BCI)-based fear signal detection and subsequent self-defense system has to be developed. Self-defense is a countermeasure that involves protecting the health and well-being of oneself from detriment by others including human beings, animals. The system aims at designing an automated alert mechanism that operates involuntarily by taking into consideration the brain wave signals of a human being without the knowledge of the victim. It includes a small device that may be embedded in cap or clothes, which monitors the human brainwaves to detect the fearful condition of a person when he is in danger. Upon detection of fear signals, the system sends emergency alert information including the location of the user via GPS to some predefined mobile numbers. The system is aimed to be implemented using brain sensors along with a Machine Learning Algorithm – Support Vector Machine and Neural Networks.**

Keywords*— *Artificial Intelligence, Non-invasive Electroencephalogram, Machine Learning, Dataset for Emotion Analysis using Physiological signals, Emotion Recognition, Fear Detection, .

I. INTRODUCTION

Sexual harassment is a form of discrimination that consists of three types of harassing behavior first is gender harassment (verbal and nonverbal behaviors that convey hostility,objectification, exclusion, or second-class status about members of one gender), second is unwanted sexual attention (unwelcome verbal or physical sexual advances, which can include assault) and third is Sexual coercion (when favorable professional or educational treatment is conditioned on sexual activity). The distinctions between the types of harassment are important, particularly

because many people do not realize that gender harassment is a form of sexual harassment.

The major reason which drove us to select this topic was The Asifa Rape Case. The impact of sexual harassment can vary. 'Anger, annoyance, and embarrassment were the most common emotional responses, with 45% of women feeling anger, 41% annoyance, and 36% embarrassment. Furthermore, close to one in three women (29%) who has experienced sexual harassment have said that they felt fearful as a result of the most serious incident, while one in five (20%) victims say that the most serious incident made themselves feel ashamed of what had taken place. In other situations, harassment may lead to temporary or prolonged stress or depression depending on the recipient's psychological abilities to cope and the type of harassment and the social support or lack thereof for the recipient. Psychologists and social workers report that severe or chronic sexual harassment can have the same psychological effects as rape or sexual assault. Victims who do not submit to harassment may also experience various forms of retaliation, including isolation and bullying. As an overall social and economic effect every year, sexual harassment deprives women from active social and economic participation and costs hundreds of millions of dollars in lost educational and professional opportunities for mostly girls and women. The most effective weapon against sexual harassment is prevention. Harassment does not disappear on its own. In fact, it is more likely that when the problem is not addressed, the harassment will worsen and become more difficult to remedy as time goes on.

This paper has six sections. The section on literature survey discusses six papers with their methodologies in brief.. After literature survey, is the problem definition. The next section discusses the architectural diagram of the proposed system explaining the use of different modules. The section on methodologies discusses various methods that can be used to implement the proposed system. The last section gives us idea about SVM classifier used to train dataset. The important steps are mentioned in brief. The last section discusses the conclusion and future scope.

II. LITERATURE SURVEY

Literature survey is done so as to get an idea of previously implemented projects along with their methodology, advantages and limitations. This helps in understanding the shortcomings of the previously studied and created EEG systems for safety and motivates us to overcome them.

An emotion is a complex psychological state that involves three distinct components: a subjective experience, a physiological response, and a behavioral or expressive response [14], [2]. Emotions have been described as discrete and consistent responses to events (external or internal) with significance for the organism [3]. They are brief in duration and correspond to a coordinated set of responses, which may include verbal, behavioral, physiological and neural mechanisms.

There are two main areas of the brain correlated with emotional activity: the amygdala (located close to the hippocampus, in the frontal portion of the temporal lobe); and the pre-frontal cortex (covers part of the frontal lobe). Although there is no consensus about a possible lateralization of the amygdala, its activation seems to be more related to negative emotions than positive ones [4]. Changes in alpha power and asymmetry between the hemispheres of the brain are related to emotions. A relative right frontal activation is associated with withdrawal stimuli or negative emotions, such as fear or disgust. A relatively greater left frontal activation is associated with an approach stimuli or positive emotions, such as joy or happiness. Thus, the asymmetrical frontal EEG activity may reflect changes on the valence [4], [5], [6], [7]. Beta bands are also related to valence

[7]. Pre-frontal and parietal asymmetry in the alpha band and temporal asymmetry in gamma band are present for valence recognition, while pre-frontal asymmetry in alpha band and temporal asymmetry in the gamma band are observable for arousal recognition [8]. Changes in the gamma band are related with the emotions happiness and sadness, and so is the decrease in the alpha wave in different sides of the temporal lobe (left for sadness, and right for happiness) [9], [10]. Finally, the ERP components of short (N100 and P100) to middle (N200 and P200) latencies have been shown to correlate with valence, whereas the components of middle to long (P300 and SCP) latencies have been shown to correlate with arousal [11].

Previous studies have suggested that men and women process emotional stimuli differently. They suggest that men rely on the recall of past emotional experiences to evaluate current emotional experiences, whereas women seemed to engage the emotional system more readily [12]. There is also some evidence that women share more similar EEG patterns among them when emotions are evoked, while men have more individual differences among their EEG patterns [13]. In summary, [14] concludes that the frontal and parietal lobes are the most informative about the emotional states, while the alpha, gamma and beta waves appear to be the most discriminative. The gender-related findings are consistent with the common belief that women are more emotional than men, which suggests possible gender-related neural responses to emotional stimuli.

Although EEG has a poor spatial resolution and requires many electrodes placed at various sites on the head, it provides great time resolution, allowing researchers to study phase changes in response to emotional stimuli. Furthermore, the use of EEG is noninvasive, fast, and inexpensive, making it a preferred method in studying the brain's responses to emotional stimuli [15]. Nowadays, due to their wearability, price, portability and ease-of-use, new wireless EEG devices are coming to the market. Thus, it is now possible to use EEG-based emotion recognition in different areas such as entertainment, e-learning, virtual worlds, or e-healthcare applications [16], [17]. It may be used for many purposes, such as instant messaging, online games, assisting therapists and psychologists while doing their job.

The frontal lobe is the part of the brain that controls important cognitive skills in humans, such as emotional expression, problem solving, memory, language, judgment, and sexual behaviors. It is, in essence, the "control panel" of our personality and our ability to communicate.It is also responsible for primary motor function, or our ability to consciously move our muscles, and the two key areas related to speech, including Broca's area.

The frontal lobe is larger and more developed in humans than in any other organism. As its name indicates, the frontal lobe is at the front of the brain. The right hemisphere of the frontal lobe controls the left part of the body, and vice versa.

The frontal lobe also plays an important part in integrating longer non-task based memories stored across the brain. These are often memories associated with emotions derived from input from the brain's limbic system. The frontal lobe modifies those emotions to generally fit socially acceptable norms. The function of the frontal lobe involves the ability to project future consequences resulting from current actions. Frontal lobe functions also include override and suppression of socially unacceptable response as well as differentiation tasks.

No	Publisher	Year	Advantages	Disadvantages
1.	IEEE Transactions on Affective Computing	2018	more discriminative EEG feature extraction	Dynamic GCN is not much explored
2	Proceedings of International Ethical Hacking Conference 2018, Advances in Intelligent Systems and Computing	2018	--	--
3	International Conference on Intelligent Computing, Instrumentation and Control Technologies	2017	high learning capacity, classify a high dimensional feature set	Several key parameters need to be set accurately
4	IEEE	2017	GELM yields a much better performance in comparison with other models	cannot encode more than 1 layer of abstraction
5	Journal of Mechanics in Medicine and Biology	2012	used for analyzing large volume of data for real-time applications	often converge on local minima, overfit data
6	IEEE : Affective Computing	2016	SVM is frequently used	--

Table 1: Literature Survey

III.PROBLEM DEFINITION

The main aim is to ensure safety of an individual by using EEG and ML.

1. To detect and analyze brain wave patterns for fear using EEG and Machine Learning algorithms
2. To design and implement a real time system for generating alerts after detecting the state of extreme fright in mind.
3. A small device that may be embedded in a cap, which monitors the human brainwaves to detect the fearful condition of a person when he is in danger. Upon detection of fear signals, the system automatically send emergency alert information including the location of the user via GPS to some predefined mobile numbers. The system aims to be implemented using EEG sensors to do the necessary steps.

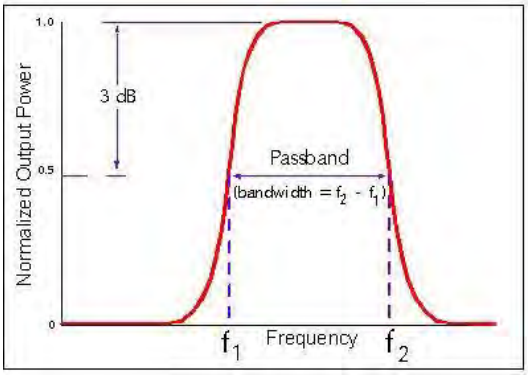
IV.ARCHITECTURAL VIEW OF THE SYSTEM

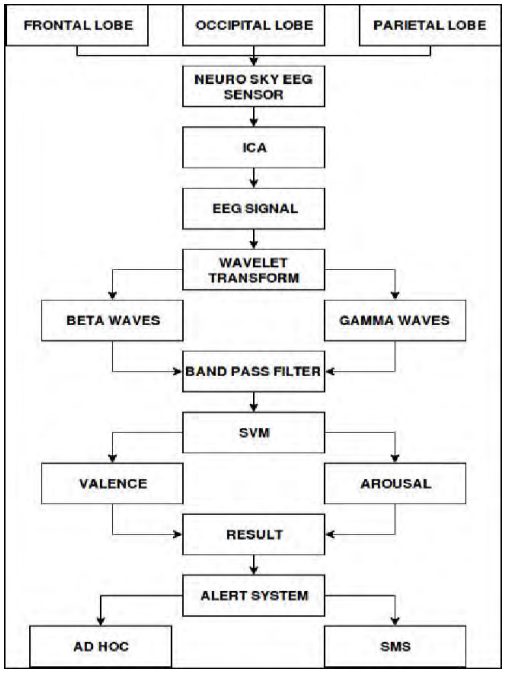
The functionality starts with detection of radiations from frontal lobe of brain using Neurosky EEG sensor. Independent Component Analysis is performed on the captured signals. Now the output is passed to wavelet transform to classify it into

gamma waves and beta waves. These waves are passed through band pass filter to remove the noisy data. Further the data is passed on to SVM Classifier which is trained already using

sent by SMS or

ad hoc networks





.

V. FEATURE SELECTION AND EXTRACTION TECHNIQUES

A. Wavelet Transform:

The wavelet transform take advantage of the intermediate cases of the Uncertainty Principle. Each wavelet measurement (the wavelet transform corresponding to a fixed parameter) tells you something about the temporal extent of the signal, as well as something about the frequency spectrum of the signal. That is to say, from the parameter w (which is the analogue of the frequency parameter k for the Fourier transform), we can derive a characteristic frequency k(w) and a characteristic time t(w), and say that our initial function includes a signal of "roughly frequency k(w)" that happened at "roughly time t(w)". Reasons for choosing Wavelet Transform:

1. A signal is said to be stationary if it does not change much over time. Fourier transform can be applied to the stationary signals.

2. However, like EEG, plenty of signals may contain non-stationary or transitory characteristics.

3. Thus it is not ideal to directly apply Fourier transform to such signals. In such a situation time–frequency methods such as wavelet transform must be used.

B. Band Pass Filter:

A band-pass filter, also bandpass filter or BPF, is a device that passes frequencies within a certain range and rejects (attenuates) frequencies outside that range. Even though until now preprocessing has been done on the gathered data, the extracted signals may contain the noisy data due to eye blinking or many more factors. Hence to avoid this errors or noisy data, band pass filter has been used.

Figure 1 : Band Pass Filter

C. Independent Component Analysis (ICA):

ICA is a blind source separation technique for linearly separating independent sources of activity constituting the data. Since artifacts and activity related brain data have separate physiological sources, the basic ICA assumption of statistically independent components is valid .The ICA of data was obtained by multiplying it with the weight matrix (W) or the de-mixing matrix to generate independent components .The artifact components are rejected after careful examination of these components to obtain error-free data.

ICA Data (Source Data) = W*Recorded EEG Data_Distance

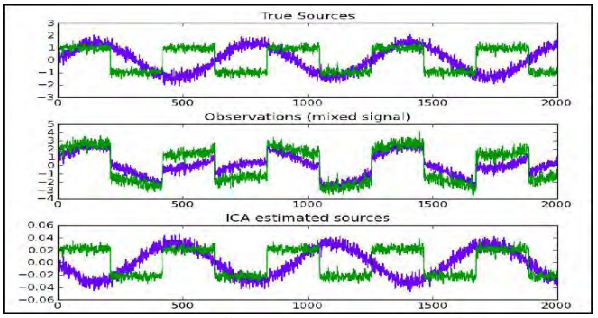


Figure 2 : Independent Component Analysis.

VI. SVM CLASSIFIER

SVM is a discriminative classifier formally defined by a separating hyperplane which can be effectively used to classify a high dimensional feature set. The hyper plane should be such that it maximizes the interclass separation. It is a supervised learning model where predefined class values are used for training. The EEG data can also be converted to a higher dimensional mapping. The mapping function can be found using Kernel functions. The most popular kernel function is RBF kernel.

VII. CONCLUSION AND FUTURE SCOPE

Fear is such a human emotion which occurs when a person is in danger. This self-defense system is based on monitoring human brain waves during prevalence of fear. When a hazard is encountered, our brain sends unique combination of electrical impulses to our endocrine system through the nerves. The most significant concern is the design of a small size device in the form of a cap that may be worn by the victim for detection of fear. On receiving the emergency information, proper safeguard of the victim is ensured. This device promises freedom of movement of common men, local population, and above all assures the nation a safe surrounding. A wireless EEG system using non-contact sensors can be implemented.

1. The system will consist of a set of capacitive electrodes manufactured on a standard printed circuit board.
2. This system can operate through fabric easily.
3. The application scope can be broadened for ensuring healthy mind state of an individual.
4. Depending upon the brain state, the system can suggest some corrective measures to the user. (Listening to music, having a walk, consulting a psychiatrist, etc.)

REFERENCES

[1] D. Hockenbury and S. Hockenbury, Discovering psychology. Macmillan Publishers, 2007.

[2] I. B. Mauss and M. D. Robinson, “Measures of emotion: A review,” Cognition & Emotion, pp.209–237, 2009.

[3] E. Fox, Emotion Science: Cognitive and Neuroscientific Approaches to Understanding Human Emotions. Palgrave Macmillan, 2008.

[4] G. Chanel, J. J. M. Kierkels, M. Soleymani, and T. Pun, “Short-term Emotion Assessment in a Recall Paradigm,” International Journal of Human-Computer Studies, pp. 607–627, 2009.

[5] C. P. Niemic, “Studies of Emotion: A Theoretical and Empirical Review of Psychophysiological Studies of Emotion,” Journal of Undergraduate Research, pp. 15–18, 2002.

[6] C. Hondrou and G. Caridakis, “Affective, natural interaction using EEG: sensors, application and future directions,” in SETN, 2012, pp. 331–338.

[7] M. Ali, A. H. Mosa, F. Al Machot, and K. Kyamakya, “EEG-based emotion recognition approach for e-healthcare applications,” in ICUFN, 2016, pp. 946–950.

[8] D. Huang, C. Guan, K. K. Ang, H. Zhang, and Y. Pan, “Asymmetric Spatial Pattern for EEG-based emotion detection,” in IJCNN, 2012, pp. 1–7.

[9] M. Li and B.-L. Lu, “Emotion classification based on gamma-band EEG,” in IEEE EMBS, 2009, pp. 1223–1226.

[10] K. S. Park, H. Choi, K. J. Lee, J. Y. Lee, K. O. An, and E. J. Kim, “Emotion recognition based on the asymmetric left and right activation,” International Journal of Medicine and Medical Sciences, pp. 201–209, 2011.

[11] M.-K. Kim, M. Kim, E. Oh, and S.-P. Kim, “A Review on the Computational Methods for Emotional State Estimation from the Human EEG,” Computational and Mathematical Methods in Medicine, 2013.

[12] T. M. C. Lee, H.-L. Liu, C. C. H. Chan, S.-Y. Fang, and J.-H. Gao, “Neural activities associated with emotion recognition observed in men and women,” Molecular psychiatry, pp. 450–455, 2005.

[13] J.-Y. Zhu, W.-L. Zheng, and B.-L. Lu, “Cross-subject and Cross-gender Emotion Classification from EEG,” in IUPESM, 2015.

[14] Soraia M. Alarcao, and Manuel J. Fonseca, “Emotions Recognition Using EEG Signals: A Survey” in IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, DOI 10.1109/TAFFC.2017.2714671

[15] M. Balconi and G. Mazza, “Brain oscillations and BIS/BAS (behavioral inhibition/activation system) effects on processing masked emotional cues: ERS/ERD and coherence measures of alpha band,” International Journal of Psychophysiology, 2009.

[16] Y. Liu, O. Sourina, and M. K. Nguyen, “Real-time EEG-based Emotion Recognition and Its Applications,” Transactions on Computational Science XII, pp. 256–277, 2011.

[17] N. Jatupaiboon, S. Pan-ngum, and P. Israsena, “Emotion classification using minimal EEG channels and frequency bands,” in JCSSE, 2013, pp. 21–24.

[18] Hosseinifard Behshad, Mohammad Hassan Moradi, and Reza Rostami, "Classifying depression patients and normal subjects using machine learning techniques and non-linear features from EEG signal", Computer Methods and Programs in Biomedicine,2013.

[19] Ashish Panat,Anita Patil “Analysis of emotion disorders based on EEG signals of Human Brain”,IJCSEA Vol.2, No.4, August 2012.

[20] A. Alkan, M.K. Kiymik, “Comparison of AR and Welch methods in epileptic seizure detection,” Journal of Medical Systems 30 (6) (2006) 413–419.

[21] M. Rajya Lakshmi,Dr. T. V. Prasad ”Survey on EEG Signal Processing Methods”,International Journal of Advanced Research in Computer Science and Software Engineering,Volume 4, Issue 1, January 2014.

[22] Yu M. Chi and Gert Cauwenberghs,”Wireless Non-contact EEG/ECG Electrodes for Body Sensor Networks”,2010 International Conference on Body Sensor Networks.

[23] B. Liu, M. Wang, L. Yu, Z. Liu, H. Yu, “Study of Feature Classification Methods in BCI Based on Neural Networks”, In proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, September 1-4, 2005.

Plagiarism Checker Using Natural Language Processing

Gautamkumar Ojha, Mihir Rane, Vivek Saroj, Dr. Rekha Sharma
Computer Engineering, TCET
University of Mumbai
ojha.gautam97@gmail.com, ranemihir45@gmail.com, viveksaroj42@gmail.com, rekha.sharma@thakureducation.org

ABSTRACT-The plagiarism of technical journals, papers and other documentation is a widespread and growing problem in the academic as well as research process. The traditional manual detection of this kind of plagiarism by human is hefty, not accurate, and overloading process. This project aims to create a plagiarism-detection system that can help academicians and researchers to make a better judgment of the work.There are currently many existing detection algorithms. Some of them have been studied, analysed, and compared. Based on this analysis and comparison, the most suitable one for our system has been chosen for the implementation. The system under development also aims to maximise the plagiarism score by thoroughly analysis and comparing the requested document. Given that, there are already some existing systems or software for plagiarism detection, but they are all not free and not open- source.Currently in the initial phase of the project the proposed system checks plagiarism between two documents. Both Syntactic and Semantic similarity is calculated and the similarity index is shown in result stating whether the two sentences are similar or not

Keywords: External Plagiarism; Natural Language Toolkit; Intelligent Plagiarism; Karp-Rabin; All-to-All Matching; Siamese Neural Network; WordNet.

1. Introduction:

Literary theft is the "illegitimate allocation" and "taking and distribution" of another creator's "dialect, considerations, thoughts, or articulations" and the portrayal of them as one's own unique work.

Literary theft is viewed as scholastic untrustworthiness and a break of journalistic morals. It is liable to sanctions like punishments, suspension, and even ejection. As of late, instances of "extraordinary unoriginality" have been recognized in the scholarly world. The cutting edge idea of counterfeiting as shameless and inventiveness as a perfect developed in Europe in the eighteenth century, especially with the Romantic development. Literary theft isn't in itself a wrongdoing, yet can comprise copyright encroachment. In the scholarly community and industry, it is a genuine moral offense. Written falsification and copyright encroachment cover to a significant degree, yet they are not equal ideas, and numerous kinds of literary theft don't establish copyright encroachment, or, in other words copyright law and might be arbitrated by courts. Literary theft isn't characterized or rebuffed by law, yet rather by establishments (counting proficient affiliations, instructive foundations, and business substances, for example, distributing organizations). There are distinctive kinds of content copyright infringement, which are comprehensively arranged as exacting and savvy literary theft. In the previous, the liars don't invest much energy secluded from everything the scholastic wrongdoing they submitted. For instance, they just reorder the content from the Internet. In canny copyright infringement, the counterfeiter endeavor to sell out perusers by changing the commitments of others in a keen way and influence it to show up as their own work. Shrewd copyright infringers endeavor to conceal, muddle, and change the first work in different wise ways, including content control, interpretation, and thought reception. Content control is identified with the exchanging of the words in the first content and making a phony one. In thought appropriation the entire thought from a content is duplicated. In this manner astute literary theft cases are more perplexing to identify and consequently difficult to work with. The work proposed in this paper manages location of copyright infringement cases utilizing Natural Language Toolkit (NLTK) and WordNet(Lexical Database) for semantics and syntactic similarity between the documents.

2. Background and Related Work:

All the approaches available in external plagiarism detection mainly concentrate on improving efficiency of detection system. Most of the approaches present today

follow a common methodology, only a few deviates. The general approach includes the following steps: pre-processing, candidate document selection, document comparisons, passage boundary detection and evaluation. While there are many tools that automate the detection of plagiarism in source code, users still need to inspect the results and decide whether there is plagiarism or not. Moreover, users often rely on a single tool (using it as "gold standard" for all cases), which can be ineffectiveand risky. Hence, it is desirable to make use of several tools to complement their results. However, various limitations exist in these tools that make their usage a very time- consuming task, such as the need of manually analyzing and correlating their multiple outputs. In this paper, we propose an automated system that addresses the common usage limitations of plagiarism detection tools. The system automatically manages the execution of different plagiarism tools and generates a consolidated comparative visualization of their results.

.1 Desirable properties

We believe a copy-detection algorithm should have three properties:

Whitespace insensitivity In matching text files, matches should be unaffected by such things as extra whitespace, capitalization, punctuation, etc. In other domains the notion of what strings should be equal is different—for example, in matching software text it is desirable to make matching insensitive to variable namesNoise suppression Discovering short matches, such as the fact that the word the appears in two different documents, is uninteresting. Any match must be large enough to imply that the material has been copied and is not simply a common word or idiom of the language in which documents are written.Position independence Coarse-grained permutation of the contents of a document (e.g., scrambling the order of paragraphs) should not affect the set of discovered matches. Adding to a document should not affect the set of matches in the original portion of the new document. Removing part of a document should not affect the set of matches in the portion that remains.

2.2 Karp-Rabin String Matching

Karp and Rabin's algorithm for fast substring matching is apparently the earliest version of fingerprinting based on k- grams [4]. Their problem, which was motivated by string matching problems in genetics, is to find occurrences of a particular string s of length k within a much longer string. The idea is to compare hashes of all k-grams in the long string with a hash of s. However, hashing strings of length k is expensive for large k, so Karp and Rabin propose a "rolling" hash function that allows the hash for the i + 1st k-gram to be computed quickly from the hash of the ith k- gram. Treat a k-gram c1 ...ck as a k-digit number in some base b. The hash H(c1 ...ck) of c1 ...ck is this number: c1 ∗ b k−1 + c2 ∗ b k−2 ∗ ... + ck−1 ∗ b + ck To compute the hash of the k-gram c2 ...ck+1, we need only subtract out the high-order digit, multiply by b, and add in the new low order digit. Thus we have the identity: H(c2 ...ck+1)=(H(c1 ...ck) − c1 ∗ b k−1) ∗ b + ck+1 Since bk−1 is a constant, this allows each subsequent hash to be computed from the previous one with only two additions and two multiplications. Further, this identity holds when addition and multiplication are modulo some value (e.g., the size of the largest representable integer), so this method works well with standard machine arithmetic [1].

2.3 All-to-all matching

The first scheme to apply fingerprinting to collections of documents was developed by Manber, who apparently independently discovered Karp-Rabin string matching and applied it to detecting similar files in file systems [2]. Rather than having a single candidate string to search for, in this problem we wish to compare all pairs of k-grams in the collection of documents. The all-to-all nature of this comparison is a key difficulty in document fingerprinting. To illustrate, consider the problem of all-to-all matching on ASCII text. Since there is a k-gram for every byte of an ASCII file, and at least 4-byte hashes are needed for most interesting data sets, a naive scheme that selected all hashed k-grams would create an index much larger than the original documents. This is impractical for large document sets, and the obvious next step is to select some subset of the hashes to represent each document. But which hashes should be selected as fingerprints? A simple but incorrect strategy is to select every ith hash of a document, but this is not robust against reordering, insertions and deletions (requirement (3) above). In fact, prepending one character to a file shifts the positions of all k-grams by one, which means the modified file shares none of its fingerprints with the original. Thus, any effective algorithm for choosing the fingerprints to represent a document cannot rely on the position of the fingerprints within the document. The scheme Manber chose is to select all hashes that are 0 mod p. In this way fingerprints are chosen independent of their position, and if two documents share a hash that is 0 mod p it is selected in both documents. Manber found this technique worked well. In, Heintze

proposed
choosing the n smallest hashes of all k-grams of a document as the fingerprints of that document[3]. By fixing the number of hashes per document, the system would be more scalable as large documents have the same number of fingerprints as small documents. This idea was later used to show that it was possible to cluster documents on the Web by similarity [2]. The price for a fixed-size fingerprint set is that only near- copies of entire documents could be detected. Documents of vastly different size could not be meaningfully compared; for example, the fingerprints of a paragraph would probably contain no fingerprints of the book that the paragraph came from. Choosing hashes 0 mod p, on the other hand, generates variable size sets of fingerprints for documents but guarantees that all representative

fingerprints for a paragraph would also be selected for the book. Broder [1] classifies these two different approaches to fingerprinting as being able to detect only resemblance between documents or also being able to detect containment between documents.

2.4 LSTM Siamese Neural Netwo

It present a siamese adjustment of the Long Short-Term Memory (LSTM) organize for marked information included sets of variable-length groupings. This model is connected to survey semantic comparability between sentences, where we surpass best in class, beating deliberately high quality highlights and as of late proposed neural system frameworks of more noteworthy unpredictability. For these applications, furnishing word-inserting vectors enhanced with synonymic data to the LSTMs, which utilize a settled size vector to encode the basic significance communicated in a sentence (independent of the specific wording/language structure). By limiting ensuing tasks to depend on a straightforward Manhattan metric, urging the sentence portrayals learned by the model to frame a profoundly organized space whose geometry reflects complex semantic connections. The outcomes are the most recent in a line of discoveries that feature LSTMs as intense dialect models equipped for errands requiring perplexing comprehension [3].

2.5 Sentence Similarity based on Corpus statistics and WordNet

Sentence similitude estimates assume an inexorably vital job in content related research and applications in zones, for example, content mining, Web page recovery, and discourse frameworks. Existing strategies for figuring sentence similitude have been received from approaches utilized for long content reports. These strategies procedure sentences in a high-dimensional space and are therefore wasteful, require human information, and are not versatile to some application areas. This paper centers straightforwardly around figuring the closeness between short messages of sentence length. It displays a calculation that assesses semantic data and word arrange data suggested in the sentences. The semantic closeness of two sentences is figured utilizing data from an organized lexical database and from corpus insights. The utilization of a lexical database empowers our technique to demonstrate human presence of mind learning and the consolidation of corpus insights enables our technique to be versatile to various areas. The proposed strategy can be utilized in an assortment of utilizations that include content learning portrayal and disclosure. Analyses on two arrangements of chose sentence sets exhibit that the proposed strategy gives a comparability measure that demonstrates a noteworthy connection to human instinct [4].

Comparative Analysis:

Table 1 : Comparison of different algorithms

Parameter	Karp-Rabin	All to All Matching	NLTK Tokenizing	LSTM Siamese	WordNet Corpus based
Words	Yes	No	Yes	No	Yes
Similarity Detection	Only Semantics	Only Semantics	Only Semantics	Both Semantics and Syntactic	Partially Semantics
Sentence	No	Yes	No	Yes	No
Complexity	O(mn)	O(n^2)	O(mn)	O(n*t*(ij+jk+kl))	O(mn)

3.1 Syntactic Similarity

The methods used to detect plagiarism between

3. Proposed Methodology:

documents were fulfilled by using some softwaressuch as Anaconda Navigator, NLTK library, and Python Programming Language.

```
def process_content(l,tokenized):
    try:
        file=[]
        file_list_separated = []
        stop_words = set(stopwords.words('english'))
        for i in tokenized[:l]:
            #word separation and removing punctuation
            tokenizer = RegexpTokenizer(r'\w+')
            word_tokens = tokenizer.tokenize(i)
            words = [w for w in word_tokens if not w in stop_words]
            file.extend(words)
            #use append when separate list for separate sentences
            file_list_separated.append(words)
        return file,file_list_separated
    except Exception as e:
        print(str(e))
```

Code for File1

There are pictures of Jupyter Notebooks which illustrates the methodology used to implement the Plagiarism checker. In process content function, only words are

```
def show_plagiarised(file1,file2):
    f1 = set(file1)
    f2 = set(file2)
    return f1.intersection(f2)
```

```
plagiarised_words = show_plagiarised(file1,file2)
print("Plagiarised words are :")
for i in plagiarised_words:
    print(i,end =' ')
```

```
Plagiarised words are :
town boasts soybeans Engineering acres college 11 Neutron farmland similar hogs yearly rural Quark Economics 10
000 known famous Hog several Gridlock Agricultural used beautiful growing hometown lies department This mainly
devoted local My also surrounded Watching population program Subnormal 15 center attending cattle First student
swells Festival College campus people Corn many corn Calling
```

Here we pass two sentences as input to compute semantics similarity between them as shown in figure below

Show Plagiarism Function is used to find intersection between two files which give the common words betweetwo files. **'sentence_by_sentence'** function checks all the sentences whether they contain the words which are present in common words. If yes, then counter is raised

Here File1 is read containing one paragraph. Paragraph is broken into sentences. Sentences are passed to process_content function.

Same process is followed for file 2 as shown in above.

```
In [31]: import nltk

In [32]: from nltk.tokenize import sent_tokenize, word_tokenize
         from nltk.corpus import stopwords,wordnet
         from nltk.stem import WordNetLemmatizer
         from itertools import product
         import numpy
```

Figure 3: Importing Necessary modules.

```
print("Sentence by sentence analysis for file 1:")
sentence_by_sentence(file_list_separated1,plagiarised_words)
print()
print("Sentence by sentence analysis for file 2:")
sentence_by_sentence(file_list_separated2,plagiarised_words)
```

```
Sentence by sentence analysis for file 1:
Sentence 1 : 5 words plagiarised out of 7
Sentence 2 : 3 words plagiarised out of 4
Sentence 3 : 6 words plagiarised out of 8
Sentence 4 : 7 words plagiarised out of 9
Sentence 5 : 10 words plagiarised out of 10
Sentence 6 : 2 words plagiarised out of 9
Sentence 7 : 10 words plagiarised out of 10
Sentence 8 : 8 words plagiarised out of 10
Sentence 9 : 2 words plagiarised out of 6
Sentence 10 : 12 words plagiarised out of 14
Sentence 11 : 17 words plagiarised out of 18
Percentage plagiarism is 78.0952380952381

Sentence by sentence analysis for file 2:
Sentence 1 : 6 words plagiarised out of 7
Sentence 2 : 5 words plagiarised out of 5
Sentence 3 : 4 words plagiarised out of 5
Sentence 4 : 11 words plagiarised out of 13
Sentence 5 : 15 words plagiarised out of 16
Sentence 6 : 17 words plagiarised out of 19
Sentence 7 : 9 words plagiarised out of 10
Sentence 8 : 10 words plagiarised out of 12
Sentence 9 : 12 words plagiarised out of 12
Percentage plagiarism is 89.8989898989899
```

Figure 2: Percentage Similarity

3.2 Semantic Similarity

For Semantic Similarity the approach is based on Wordnet which is a lexical database for the English dialect. It bunches English words into sets of equivalent

words called synsets, gives short definitions and use models, and records various relations among these equivalent word sets or their individuals.



Finally the output shows similarity index and based on similarity threshold(assumed in this case, if similarity of 0.8 then the sentences are similar and if it is between 0.6 and 0.8 else if the index is below the given range then the sentences are not similar).



Figure 6 : Final Output of Semantics Similarity

plagiarism is calculated by: **Sentence level plagiarism =**

$$\frac{(\text{words in sentence} \cap \text{common words})}{(\text{total number of words})}$$

Here, total plagiarism is calculated by all words plagiarized in each sentence to total number of words in file.

5. Combine the sentence level score to calculate overall plagiarism score.
6. Display sentence wise plagiarism score and overall plagiarism score.

4.2 For semantics analysis and computation, we make use of the following algorithm:

After taking the input we preprocess it using techniques such as Stopwords removal, Lemmatization, Tokenization and Stemming. Thus, at last we compute Similarity index of each word as below

Figure 5: Compute Similarity

1: Input: A set of documents to check.

2: Output: The plagiarism report of the documents.

3: while the user clicks the button to scan for plagiarism do 4: generate the plagiarism report and display similarities

5: from DB according to size

6: generate the plagiarism report and show all matching

7:display the plagiarism report and show the matched 8:

1. Conclusion:

The Project of Plagiarism checking using natural language procession aims to check the plagiarized content for the requested document. Since the field of research and development is blooming and it requests the urgency for system that evaluates the uniqueness of the technical journals, paper and many other documentations.

[3] Sentence Similarity Based on Semantic Nets and Corpus Statistics, Y.Li, D.Mclean, Z.A. Bandar, J.D. O'Shea, K. Crockett. Published in: IEEE Transactions on Knowledge and Data Engineering (Volume: 18 , Issue: 8 , Aug. 2006).

There are already existing systems that compares and evaluates the plagiarism of a particular document but that comparison is done only on the basis of the syntactical similarities and judging a document to be plagiarized only for syntactical similarities is not a fair platform.

In the first phase of our project we have partially implemented our proposed worked on sentence similarity level basis. Wherein, we have computed syntactic and semantic similarity between two sentences and displayed the results. In the next phase of our project we plan

to
implement the proposed work on entire document and
check it's respective plagiarism with respect to web and finally generate a plagiarism report for the document. Therefore, our system will be checking and evaluating the document for plagiarism by comparing both syntactic and semantic similarities between them. By doing so we restore the true virtue of plagiarism and benefit research and development sector. We also look forward to deploy the developed system on a global platform through a free and open source website.

7.REFERENCES

[1] Karp-Rabin String Matching, Richard M. Karp and Michael O. Rabin. Pattern-matching algorithms. IBM Journal of Research and Development, 31(2):249–260, 1987.

[2] All-to-all matching, Andrei Broder, Steve Glassman, Mark Manasse, and Geoffrey Zweig. Syntactic clustering of the web. InProceedings

Big Data Analysis for Associated Cloud Service using Parallel Trust Computing: A Survey.

Monika Gedam.[1] Bharati Ainapure.[2]

[1] *(Student, M. E. Computer Science and Engineering (Data Science and Analysis) MIT- WPU, pune)*
[2](Professor, *Computer Science and Engineering, MIT- WPU, pune)*

Abstract:-

This survey paper presents on-going research on the development of Intercloud Architecture Framework (ICAF) that examines the drawbacks of multi-provider multi-domain miscellaneous cloud that are based on infrastructure services and applications integration and interoperability.

Cloud services model that merge frequently used cloud service models, like IaaS, PaaS, SaaS, in one multilayer model with corresponding inter-layer interfaces. Intercloud Controlling and Management Plane are based on applications interaction that supports cloud. To accessing trustworthy cloud services is the main task for any cloud computing platform.

Keywords: - *Intercloud, Cloud Computing, Cloud Computing Security, Trust Computing, Big Data Analysis.*

I. Introduction

Cloud technologies bring applications and infrastructure services mobility and Complementary components of the proposed Intercloud Architecture:

Multilayer Cloud Services Model (CSM) - is functional layers and components of the general cloud-based services infrastructure are define by integration and compatibility respectively.

The three vertical planes infrastructures can define following functionality of CSM;

- Control and Management Plane.
- Operations Support System.
- Security Infrastructure.

(2) Intercloud Control and Management Plane (ICCMP) - it is used for inter-cloud applications and physical/hardware platform independence to the present distributed computing and networking applications.

INTERCLOUD:-

The networks known as the "Internet"; in a world of Cloud Computing, content, storage and computing is ubiquitous and interoperable in a network of Clouds known as the "Intercloud".

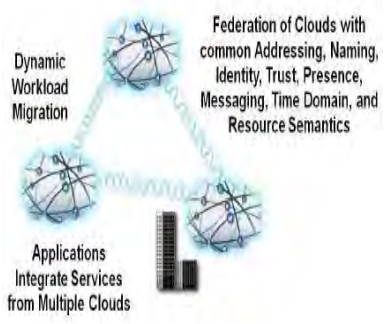


Figure: - The Intercloud Vision [1]

Individual Intercloud is capable for communicating with each other, as clients, through the server environment which is hosted by Intercloud Roots and Intercloud Exchanges.

commonly used for vertical cloud services , both relations between cloud service models (such as IaaS, PaaS, SaaS) and other required infrastructure control. The management also includes inter-applications signaling, synchronization and session management, configuration, monitoring, run time infrastructure optimization including VM migration, resources scaling, and jobs/objects routing.

The ICCMP Interfaces support some functionality that is listed below;

- Inter-/cross-layer control and signaling.
- Monitoring.
- Location service.

- Topology aware infrastructure management.
- Configuration and protocols management.

(3) Intercloud Federation Framework (ICFF) - It allow independent clouds and its related infrastructure components federation to manage independently the cloud based infrastructure components belonging to different cloud providers and/or administrative domains. It should support to the level of services, business applications, semantics, and namespaces or assuming necessary gateway.

Table: - Operations and functionalities of Intercloud Federation Framework (ICFF)

Federated inter-cloud infrastructure supports following operations	**The ICFF Interfaces support the following functionalities**
Trust and service brokers	Publishing
Service Registry	Naming, Addressing and Translation (if/as needed)
Identity provider (IdP)	Attributes management
Trust manager/router	Discovery
Service Discovery	Trust/key management

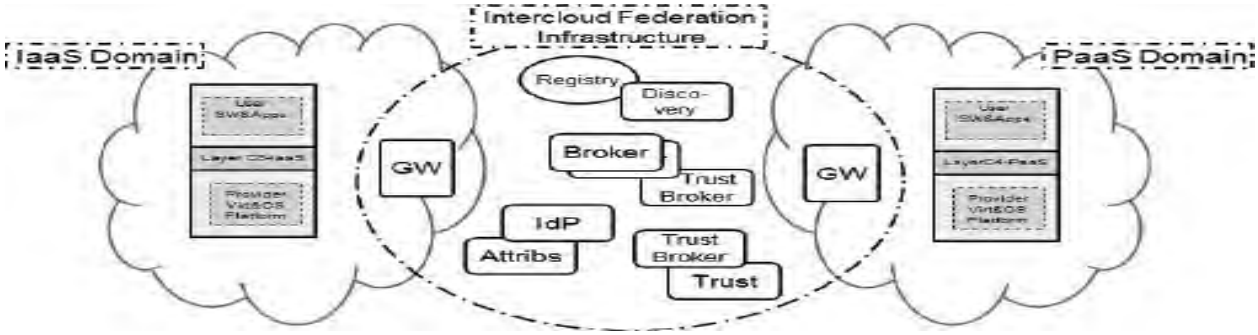


Figure. ICFF components[3] .

Table: - Functional components and functionalities of Intercloud Operation Framework (ICOF)

Main functional components	**ICOF interfaces should support the following functionalities**
Service Broker	Service Provisioning
Service Registry	Deployment
Cloud Service Provider, Cloud Operator	Decommissioning (or Termination)
Cloud (physical) Resource provider, Cloud Carrier	SLA management and negotiation
-	Services Lifecycle and metadata management

MONITORING AGENT MANAGEMENT

When a cloud provider registers its resources in the cloud broker, the registered resource demands to download a monitoring agent and bind this monitoring agent with its

IP. If a computing task is distributed to this resource, the monitoring agent will be activated and then begin its monitoring task.

TRUST COMPUTING BASED ON BIG DATA ANALYSIS

To accelerate pre-processing and perceiving the trust behavior indicators, two types of software agents are exploited.

1) **SMAs (security monitoring agents) :** -

SMAs are responsible for collecting security-related data, such as the authentication type, the authorization type, the self-security competence and the number of malicious

Section III, we discuss the different technologies that are used for big data analysis for cloud service field. Section IV contains advantages and finally, concluding remark will be given in Section V.

II. Literature Review

[1] David Bernstein et al Intercloud Security Considerations Cloud computing is a new design example for large, distributed datacenters. In cloud computing users can store their data using cloud infrastructure and even they can access the applications like email, search, and social networks, Service providers offer These all services.

Advantage: -

The key benefit of the system is its deterministic behavior and the fair balancing of load among the peers.

[10]Yuri Demchenko et al introduced a system that represents the research work for developing the Intercloud Architecture Framework (ICAF). This ICAF system defines four corresponding components of Intercloud integration and interoperability: multilayer Cloud Services Model; Intercloud Control and Management access.

2) **QMAs (QoS monitoring agents):** -

QMAs responsible for collecting and pre-processing QoS-related information, most of which are the indirect trust indicators, and need calculation and pre-processing based on a given time period.

In this work, we will present an updated survey on recent development and focus on the future research and development trends in Big Data Analysis for Associated Cloud Service field. We have initially discussed Introduction in section I, then secondly the literature review in Section II.

Plane; Intercloud Federation Framework, and Intercloud Operation Framework. The architecture provides the model for developing Intercloud middleware for making possible clouds interoperability and integration. In multilayer Cloud Services Model uses IaaS, PaaS, SaaS service models.

Advantage:-

Developing Intercloud middleware and facilitate clouds interoperability and integration.

[7] Hamid Mohammadi Fard et al the system will propose a pricing model and truthful mechanism for dynamic scheduling of a single task in commercial multi-Cloud environment. The system will schedule the scientific workflows with respect to optimization of two objectives, makespan and monetary cost, and proved the truthfulness of the mechanism theoretically. The proposed system used the (BOSS) Bi-Objective Scheduling Strategy for dynamic scheduling.

Advantage:-

The goal is to satisfy an application deadline by extending the local computational capacity with Cloud resources.

[8] Xiaoyong Li et al this paper present a service

operator-aware trust scheme (SOTS) for resource matchmaking across multiple clouds. According to multi-dimensional resource the operator, model trust and evaluate the process. The multi-attribute decision-making, and develop an adaptive trust evaluation approach based on information entropy theory. This approach can overcome the limitations of traditional trust schemes, and the trusted operators are weighted manually or subjectively.

Advantage:-

The middleware trust management framework is used for reducing the user burden and improving system dependability.

[3]Shirlei Aparecida de Chaves et al This work describe an architecture for and implementation of a private cloud monitoring system. This is quite high-level architecture and it contains three layers: an Infrastructure layer, an Integration layer and a View layer respectively. The implementation is modular in design and consists of different components that are mainly focused on the integration layer of the architecture. Recently, it is compatible with Eucalyptus as a IaaS implementation therefore, it is mentioned that it could be extended to work with alternative IaaS implementations in the future. It appears to rely quite heavily on Nagios for its monitoring functionality.

Advantage:-

In private clouds, enterprises usually use their own (proprietary) data center to take the benifit of cloud computing while inside their firewalls.

[13]Alberto Fernández et al focus on the system for large-scale analytics based on the MapReduce scheme and Hadoop, its open-source implementation. It also identifies several libraries and software projects that have been developed for aiding practitioners to address this new programming model. This Cloud Computing and its features stress its elasticity in the use of computing resources and space, less management effort, and flexible costs.

Advantage:-

The perspective of deploying a huge cluster of machines configured such that the load can be distributed among them.

Keqing He et al discuss about the Special Issue (SI) on Cloud Services Meet Big Data is to solicit innovative and promising methods and different techniques related to cloud services in the era of Big Data. The issue can promote the visibility and relevance of this noteworthy direction in the interdisciplinary field between Services Computing and other emerging disciplines.

Advantage:-

The evaluation of state-of-the art, using both real-world QoS data and synthetic Web service data.

[12]Marcos D. Asuncion et al discusses approaches and environments for carrying out analytics on Clouds for Big Data applications. It encircles around four main areas of Big Data analytics, namely (i) data management and supporting architectures; (ii) model development and scoring;

(iii) visualization and user interaction; and (iv) business models.

Advantage:-

Managing and gaining insights from the produced data is a key to competitive benefit.

III. Technologies Used:-

Pretty Good Privacy (PGP):-

Pretty Good Privacy (PGP) was originally developed by Phillip Zimmerman to provide a means of secure communication in an insecure electronic environment. The term “Pretty Good” is an underestimated framework

which is based on, Public Key Infrastructure (PKI) and its encryption standards. It can use Diffie-Helman or RSA algorithms of varying strengths that have been subjected to rigorous cryptanalysis[16].When a user encrypts plaintext using PGP, then PGP initially compresses the plaintext. Data compression saves modem transmission time, disk space, and more importantly, strengthens the security. Most cryptanalysis techniques accomplish patterns that are present in the plaintext to crack the cipher. Compression reduces the patterns in the plaintext, thereby greatly enhancing resistance to cryptanalysis[15, 16].

PGP Encryption-

Then PGP creates a session key, which is a one-time-only secret key. This key is a random number generated by the random movements of your mouse and the keystrokes you type. This created session key works with a very secure, fast conventional encryption algorithm to encrypt the plaintext; and generated result is cipher text. Earlier the data is encrypted the session key is again encrypted to the recipient's public key. This public key-encrypted session key is then transmitted along with the cipher text to the recipient[15]

.

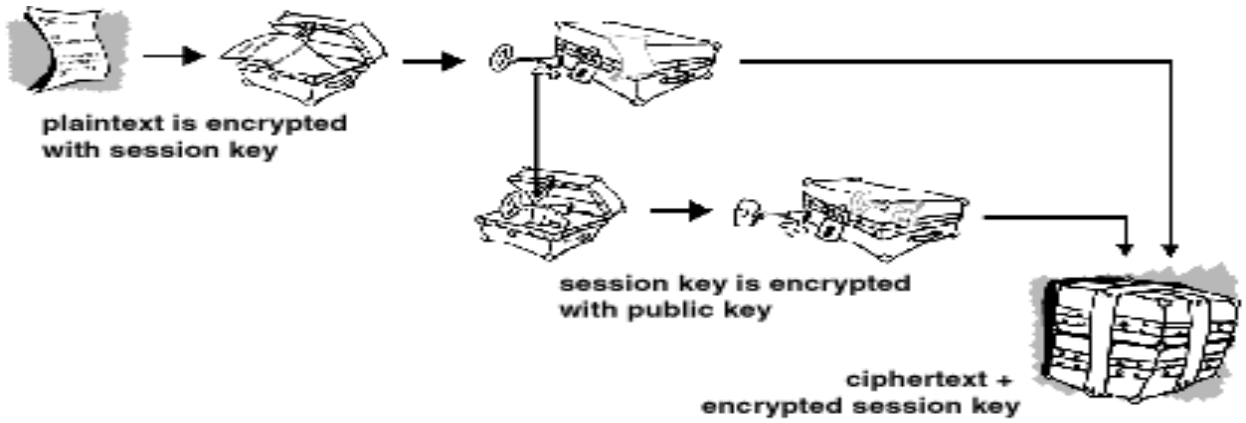


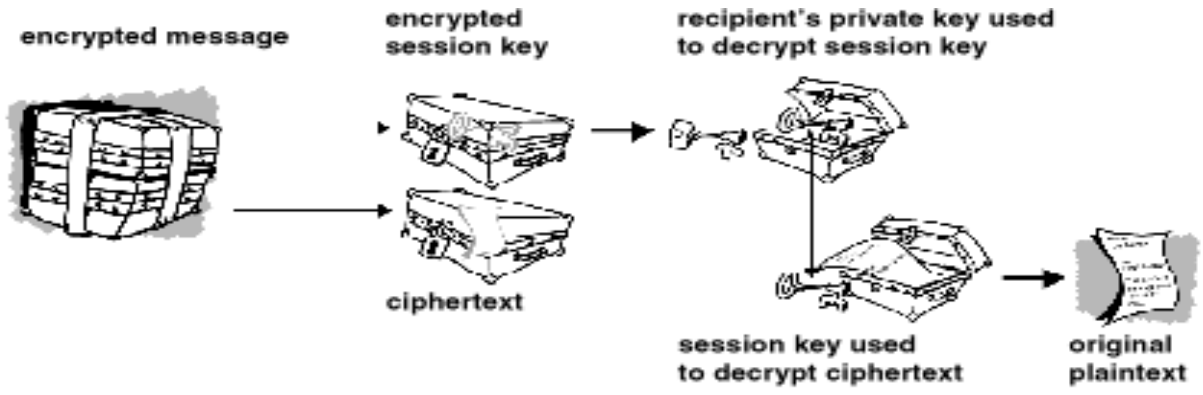
Figure: - PGP Encryption.

PGP Decryption-

Decryption works in the reverse process. The recipient's copy of the PGP uses his or her private key to recover the temporary session key. Then the PGP decrypt the conventionally- encrypted cipher text

.

Figure: - PGP Decryption.



Real-trust Trust Degree (RTD)

RTD is used to evaluate recent cloud resource service operators, and RTD is evaluated by knowledge of a resource's quality of service. Hence, RTD is a time window-based trusted indicator for service operators, and it should be more sensitive to new operators. RTD is generated in the time period when an interaction takes place among a user and a resource respectively. This approach may lead to misinformation and could preclude an accurate evaluation of trustworthiness [2]. Thus, avoiding the effect of individual favoritism on the weight allocation of trust attributes is a key task. An adaptive trust evaluation model based on information entropy [5], which can overcome the shortage in traditional trust models, wherein the trusted attributes are weighted manually or subjectively.

IV. Advantages

Provides trustworthy & secure cloud service to the user.Calculate real-time trust value of cloud resources at the different time.Provide access control mechanism to the cloud services.

V. Conclusion

This paper is a survey of Intercloud Control and Management Plane that supports cloud based applications and infrastructure Services interaction; Intercloud Federation Framework that defines infrastructure components for independent cloud filed federation; and Intercloud Operation Framework that defines functional components and procedures to support cloud based services and operations. Research work of different researchers is explored in literature survey with its advantages. Different technologies that are used for big data analysis using parallel trust Computing is also demonstrated.

ACKNOWLEDGMENT

The authors would like to thank the publishers and researchers for making their resources available. We also thank the college authority for providing the required infrastructure and support.

References

[1] David Bernstein and Deepak Vij "Intercloud Security Considerations" International Conference on Cloud Computing Technology and Science 2010 IEEE.

[2] X. Li, F. Zhou, X. Yang, "A Multi-Dimensional Trust Evaluating Model for Large-scale P2P Computing", Journal of Parallel and Distributed Computing, vol.71, no.6, pp.837-847, 2011.

[3] Shirlei Aparecida de Chaves, Rafael Brundo Uriarte, Carlos Becker Westphall "Toward an Architecture for Monitoring Private Clouds" 2011 IEEE.

[4] H. Shen and G. Liu, "An efficient and trustworthy resource sharing platform for collaborative cloud computing," IEEE Trans. Parallel Distributed. System., vol. 25, no. 4, pp. 862–875, Apr. 2014

[5] H. Zhou, G. Zhang, G. Wang, "Multi-objective decision making approach based on entropy weights for reservoir flood control operation", Journal of Hydraulic Engineering, Vol. 38, no. 1, pp. 100-106, 2007.

[6] F. Paraiso, N. Haderer, P. Merle, R. Rouvoy, and L. Seinturier, "A federated multi-cloud PaaS infrastructure," in Proc. 5th IEEE Int. Conf. Cloud Comput. (CLOUD), Jun. 2012, pp. 392– 399.

[7] Hamid Mohammadi Fard, Radu Prodan and Thomas Fahringer "A Truthful Dynamic Workflow Scheduling Mechanism for Commercial Multi-Cloud Environments" TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS 2012 IEEE.

[8] Xiaoyong Li, Huadong Ma, Feng Zhou and Xiaolin Gui "Service Operator-aware Trust Scheme for Resource Matchmaking across Multiple Clouds" 2013 IEEE.

[9] I. Butun, M. Erol-Kantarci, B. Kantarci, and H. Song, "Cloud-centric multi-level authentication as a service for secure public safety device networks," IEEE Commun. Mag., vol. 54, no. 4, pp. 47–53, Apr. 2016.

[10] Y. Demchenko, M. X. Makkes, R. Strijkers, and C. de Laat, "Intercloud architecture for interoperability and integration," in Proc. IEEE 4th Int. Conf. Cloud Comput. Technol. Sci. (CloudCom), vol. 1. Dec. 2012, pp. 666–674.

[11] D. Niyato, Z. Kun, and P. Wang, "Cooperative virtual machine management for multi- organization cloud computing environment," in Proc. Int. Workshop Game Theory Commun. Netw. (Gamecomm), Paris, France, May 2011, pp. 528–537.

[12] M. Singhal et al., "Collaboration in multicloud computing environments: Framework and security issues," Computer, vol. 46, no. 2, pp. 76–84, 2013.

[13] Alberto Fernández, Sara del Río,Victoria López, Abdullah Bawakid, María J. del Jesus, José M. Benítez and Francisco Herrera "Big Data with Cloud Computing: an insight on the computing environment, MapReduce, and programming frameworks" © 2014 John Wiley & Sons, Ltd. Volume 4, September/October 2014.

[14] Ali Gholami and Erwin Laure "BIG DATA SECURITY AND PRIVACY ISSUES IN THE CLOUD" International Journal of Network Security & Its Applications (IJNSA) Vol.8, No.1, January 2016.

[15] https://www.slideshare.net/pavanaarrya/pgp-pretty-good-privacy.

[16] https://users.ece.cmu.edu/~adrian/630-f04/PGP-intro.html - imp PGP working.

Livestock Monitoring in Agriculture Using IOT

Khushbu Shah, Kinjal Shah , Bhavin Thakkar
1(Department of Information Technology, Thakur college of Engineering and Technology, India) 2(Department of Information Technology, Thakur college of Engineering and Technology, India) 3(Department of Information Technology, Thakur college of Engineering and Technology, India)

Abstract: Agriculture is the major source of income in countryside areas. Animals like cow, buffalo, sheep, goat etc. play an important role in life of rural. They are used as a source of income. Hence animal husbandry becomes a most important concern. Lots of farmers are now torment from different killing diseases and increased breeding costs. It is therefore essential for farmers to adapt to efficient and technical methods to increase productivity and reduce diseases. In this paper, we illustrate Internet of Things technology for the farmers. We use sensors to collect and transmit the data about parameters, such as temperature, humidity and heartbeat and then transfer it to the microcontroller. The Arduino Uno receives the data from the sensor and develops the content accordingly. Wi-Fi module will transfer the data to the monitoring website. All the hardware components and sensors will read the results through the computer monitor which also tracks the current status and location of the livestock.

Key Words: Internet of Things, Agriculture, Monitoring, Sensors.

I. INTRODUCTION

The Internet of things (IOT) refer to growing network of physical devices and other internet-based items with electronics, software,sensors,monitors,transmitters,receivers and network connectivity which enable these objects to collect and store data [2,4]. Internet of Things technology has gone further in livestock management which allows farmers to embed internet connected sensors on their livestock that does not cause them discomfort. A primary advantage for monitoring livestock is to help farmers track their cattle, their temperature, humidity, heartbeat which will help to manage the cattle in an easier and efficient manner.

Parameters	Normal	Less	More
Temperature	38-42 ◦C	Indigestion, milk infection	Influenza, anthrax
Humidity	1-72%		->80% severe stress
Heartbeat	48-84	Low blood pressure	High blood pressure

Farming industry gives significant income to the Indian nation. But livestock in the farms are mostly affected by a number of diseases. The regular temperature of cow is 38-42 ◦C. When the temperature is below 38.5-39.5◦C the diseases arises are indigestion, milk infection etc. and when the temperature is above 42 ◦C the diseases arises are influenza and anthrax [5,8]. When the temperature of the animal is very high on that time it may die. When the humidity is between 1-72% (No stress), 72- 79%
(Mild stress), 80-90% (Moderate stress), 91-99% (severe stress) [8]. Humidity can reduce heat exchange and have enervating impact on the cattle. The adult cow has a heart rate of between 48 and 84 beats per minute. An elevated heart rate may be a sign of pain, and is seen in conjunction with several diseases.

This Paper illustrates Wireless Sensor Network technology to the farmers. It proposes that a wireless sensor networks should be installed on farms to gather ecological data which shall then facilitate farmers not only in monitoring the livestock via the Web from outside the farm, but also help the control of farm environmental in remote locations. Electronic means of monitoring cattle health or animal health are developed. Using information from these sensors, farmers will be able to monitor the overall health of the livestock. In the absence of the use of monitoring technology and software, a problem with a particular livestock may not be noticed until it has reached an advanced stage and requires more considerable intervention which wouldn't have been necessary if the issue had been detected earlier.Background
Applications used do not provide the location of the Livestock [1,3]. Systems do not have a built in Wi-Fi module [5]. The results that were generated are not efficient enough due to the loose connection of the sensors to the Arduino Uno [1]. From the literature survey conducted & resulting research gaps acquisition, the proposed paper aims to overcome all the drawbacks and

present a monitoring system that is user-friendly and guarantees to provide results for all livestock used for agriculture.

II. MATERIAL AND METHODS

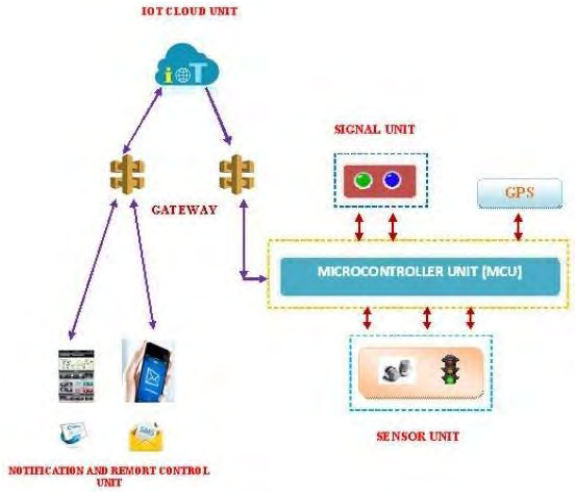


Fig: 1 Block Diagram

Temperature sensor is used to sense the body temperature of the cattle. Numbers of infections are accompanied when the body temperature of the cattle changes. Humidity sensor is used to determine stress level of the animal. The adult cow has a heartbeat of between 48 and 84 beats per minute. And heartbeat sensor will give stress as well as animal anxiety. Using information from these sensors, farmers will be able to monitor the overall health of the animal. If any of these parameters where to veer outside of accepted ranges, farmers would be able to more quickly provide treatment to the animal. The Arduino Uno is flexible and readily available for a wide variety of applications. The Arduino Uno is used because of their flexibility, simple programming, and low cost, huge collection of application data and large availability of open source developer tool. The signals arriving from the sensors are sent to the WI-FI module through Arduino Uno. The WI-FI module approaches a complete and self-controlled Wi-Fi interacting solution. The Wi-Fi

module is low cost, small and maintain Wi-Fi connection and encryption in client mode and access point mode. Wi-Fi module communication is done through simple serial RX and TX lines using "AT" i.e. attention type commands and data. A GPS module is a simple device which receives information from GPS satellites and then calculates the device's geographical position. It is used to provide suitable location of the livestock on the map

Procedure methodology

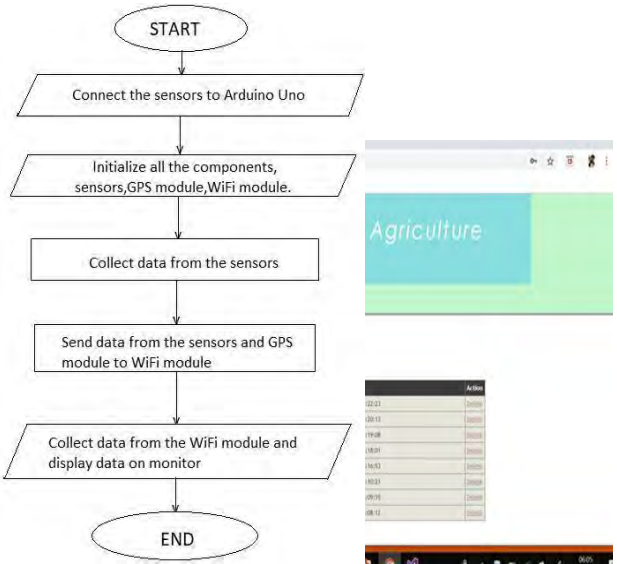


Fig: 2 Flowchart of Livestock Monitoring using IOT

This flowchart describes the flow of the Livestock Monitoring in Agriculture using IOT. It represents the flow of the method from the connection of hardware to the collection and display of data.

III. RESULT

The proposed paper describes a website that will notify the farmer with an alert message regarding any fluctuation from the normal range provided for livestock through an email notification or a message. The website also keeps a track of the database and the interface is user friendly which can be used by any farmer who is unaware about the current technologies.

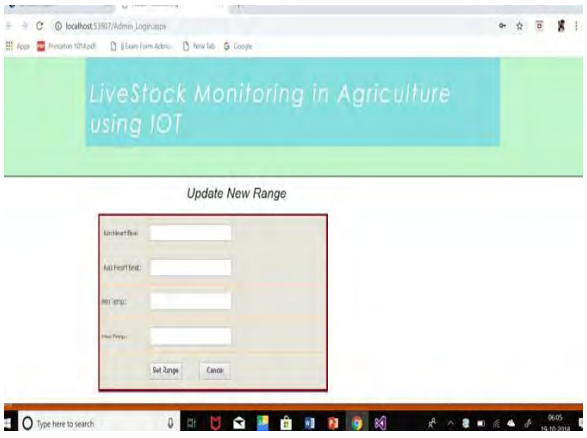


Fig: 3 Set Range of Website

Figure 3 shows that the ranges of temperature, heartbeat can be set as follows:

Min Heartbeat, Max Heartbeat, Min Temperature, Max Temperature. Here we can set the range as and when required according to the values we want to enter. Thus, this is dynamic in nature and further used in future.

Fig: 4 Results obtained using sensors

Figure 4 shows some of the results obtained from the sensors on our website. Heart rate, Temperature, Humidity and the date and time of the data are available.

IV. CONCLUSION

Using the website developed we can monitor the health of livestock to get accurate results and track the current location of the livestock. By this we will accurately track the changes in the environment ofthe Livestock.

REFERENCES

[1]. Advance Cattle Health Monitoring System Using Arduino and IOT. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (An ISO 3297: 2007 Certified Organization) Vol. 5, Issue 4,

April 2016. e-ISSN; 2395-0056

[2]. Internet of Things: A Hands-On Approach- By Arshdeep Bahga, Vijay Madisetti

[3]. International Journal of Scientific Research and Education. Volume 3, Issue 5, Pages-3386-3392,

May-2015, | ISSN (e): 2321-7545

[4]. Internet of Things: Principles and Paradigms -

Amir Vahid Dastjerdi, RajkumarBuyya

[5]. Leena Narayanan et al IJSRE Volume 3 Issue 5 May 2015 Page 3386. Animal Health Monitoring System Using Raspberry Pi. (IRJET) e-ISSN: 2395 -0056 Volume: 04 Issue: 03 | Mar -2017

[6]. Chong, C.Y.; Kumar, S.P.; Hamilton, B.A. Sensor networks: Evolution, opportunities, and challenges. Proc. IEEE 2003, 91, 1247-1256.

[7]. K. Mayer and K. Taylor, "An Embedded Device Utilizing GPRS for Communications," presented at International Conference on Information Technology and Applications, Bathurst, Australia, 2002.

[8]. Zoetis, Dairy Zoetis Proud to Protect. Zoetis Blog, 2014. Available online:

http://www.zoetis.ie/diseasesandconditions/dairy -cattle/index.aspx

Search Ranking Factors using Organic and Paid Marketing

Nazneen Pendhari, Arwa Katib, Javeria Ansari, Uzma Sayed
M. H. SabooSiddik COE Mumbai, India
mailnazneen@gmail.com, arwakt21@gmail.com, javeriaansari343@gmail.com, nidasayed33.ns@gmail.com

***Abstract*— Search Engine are robots that follow a particular pattern to show the relevant answer for a particular query. Websites are optimized based on multiple factors. Accurate Keyword research forms the foundation for SEO marketing. Proper keyword selection ensures more number of visitors to our website. SEO and SEM are the methods by which the rank of the web page can appear on top. Digital marketing includes a large range of tools and techniques by which a authority of a website can be increased as well as to boost the ROI (Return on Investment
) of any kind of business .On page as well as Off page SEO are taken under consideration for organically increasing the website authority but this takes a longer duration of time than paid marketing. Firstly, we mention all the techniques and tools required for On-page and Off-page optimization including all meta-tags, keywords, in-bound and out-bound links etc. We have also mentioned how Google Ad-words is used for paid marketing. Secondly, we highlight the comparative study between Organic and paid marketing and which one of them gets us faster results. Thirdly, this Research Paper provides the analysis of Search Engine Optimization techniques that can help to improve the ranking of a website and to get more number of viewers to increase visibility of a website.**

Keywords*— *Search Engine Optimization, Search Engine Marketing, organic search, paid search, keywords, page rank.

I. INTRODUCTION

Search Engines have become a major part of our lives as they give answers to all our queries. The number of websites on the internet is increasing rapidly and with the increasing number it becomes more complex for the search engines to show the most relevant content for a particular query. Hence there are various techniques and algorithms used by the search engines in order to rank a website on the internet.

Google has the largest market share out of all the search engines. Small position changes, means increased traffic. Optimized site maintains traffic and rank. There are factors that are considered by Google to rank a page on the top. It can be done by SEO (Search Engine Optimization) and SEM (Search Engine Marketing).Search Engine Optimization refers to the natural way by which search ranking can be increased whereas Search Engine Marketing refers to paid marketing.Assume that we are searching for a new house, we have two options, we can either buy the house or pay monthly rent .Now relating this to our context ,renting the house is same as SEO and buying the house is same as SEM. SEO is a way of improving the ranking with the help of optimizing website with the help of keywords, backlinks, meta-tags, meta description, etc. SEM is another word for paid marketing, buying a fixed place on Google so that when people search for a particular keyword, their results appear on the top. When you search a query on the internet, the results are sorted according to the relevancy. Relevant results appear on the top of the list and this includes both the organic as well as paid results [1]. Ranking of the organic results are done based on multiple factors including On-page and Off-page SEO [2]. The purpose of our research is to study the difference between organic marketing with uses SEO techniques and paid advertising which uses SEM techniques.

The section I gave the basic introduction about our review paper. Section II will include the Survey that we did based on the papers that we reviewed. Section III focuses on the SEO and SEM techniques. Section IV gives a brief Analysis and comparison between paid and organic marketing .Section V focuses on the approach and the applications in order to rank a website on top. Section VI gives the conclusion about which techniques are effective according to the techniques mentioned in the above sections.

II. State of Art / Literature Review

A. How Search Engines Work?

Search Engines are robots that crawl the web pages. They carry out the following steps:

1. Discover: Search engines follow links to look around the internet using programs known as web crawlers or spiders.
2. Analyze: Web crawlers evaluate and analyze meta data and keywords.
3. Index: The findings from each page are combined by the search engines and index are built in large databases.

4. Receive the query: Search engine uses its various algorithms to make sense of what the user searches for.

5. Return the results: Web crawlers then pull out the relevant results.

Search engine help drive traffic to your website. When you search a query for example if you search for “Sports Shoes” ,relevant results appear on the top .The first three links that appear are called as the Golden Triangle [2]. These websites appear on the top because they are optimized according to the search engines. Placing relevant keywords, Meta tags, meta description increases the authority of the website and helps it rank higher. The websites lower than these websites don’t get as much traffic as the top 3 websites [2]. Search engines display organic as well as paid results. There are 90% of the internet users who rely on search engines to find out what they want instead of directly going to a particular website. There are various search engines available for use, but 83% of internet users rely on Google.

The SEO skilled and customers each ought to perceive that SEO takes constant effort and time.

It could take months to move ahead in the ratings, or to build a large number of backlinks. Additionally, if youstop optimizing for a few time,most likely you will experience a considerable drop in ratings. SEO is an important element in the process of digital marketing. Ranking first on Google is being difficult day by day as Google updates and makes changes in its search algorithm. More than 90% of the people click on the websites that appear on the first page, hence the difficulty to optimize a websites increases.Data from a website [8] says that, most of the search engine follows this model. The search engines search for relevant keywords on the web pages but the rank to the website is given based on the back links that particular website consists of and the authority of those links. Back links are the links provided on your website that redirect to other relevant websites having good authority. Higher the authority better is the chance of ranking the website high. Hence we can say that Google discovers your website based on the keywords and is able to trust your website based on the back links on the website.

III. TECHNIQUES OF SEARCH ENGINE OPTIMIZATION

Search Engine Optimization techniques include On-Page, Off-Page and Technical SEO. Each of them consists of a number of factors that are taken into consideration while ranking on Google.

A. On-Page SEO

On-page SEO is optimization of individual web pages to earn more relevant traffic on search engines. On- page SEO is about the content and HTML source code of a page that can be optimized, while off-page SEO that has links and other external signals. SEO techniques focus on overall user experience along with search engine rankings
.

Another SEO technique that has proven to work effectively is focusing on local SEO. Recently it was learned that 'near me' searches have increased as users think the results will be automatically relevant to their location[9]. Mobile devices poses physical addresses that give the exact location of the user whether itis their home, office or another location. Users always look for services that are close to their location. Thus mobile devices prove to be very useful especially in mapping mobile app. Hence, on page SEO is a technique that focuses on users to find your business.

A simple technique is to include your contact details and directions on your website. The users usually like to spend more time on a site that provides them experience they were hoping for or better which improves search engine rankings. Our main aim is to increase search engine rankings. This is what we are focusing on in this paper, the more your website has search engine ranking the more higher it will rank.*Content*

The literal meaning for Content is "a state of satisfaction", and satisfaction is what we require from our users. Content building is very important part in SEO. The content is the raw material of your website page; optimized content is any content build with the main goal of attracting search engine traffic. The website that has extremely good content provides better environment to its user in search engine optimization. It should have valued information so that the user visits the website frequently and spends more time on the website. For example, making content while not knowing what your audience is searching for can leave you with none important traffic. And, creating content without knowing what your audience is looking for will leave you without any significant traffic. Beyond that, high quality content is what earns you back links, and it's also what search engines crawl when they arrive at your site.

Keyword Optimization: grasp wherever and the way to use keywords in your content for optimum search ability.

Content Organization: The content on your site should be organized in a logical way. This is not only good for SEO, it also helps visitors on your site find other related content easily.

Content Promotion: The content on your web site ought to be organized during a logical means. This is not solely smart for SEO, it also helps visitors on your site find other related content easily.

1. Keyword

Keywords are words that describe a part of some topic that are words that popup in mind of users while they search on the internet. For example when a user searches 'Maruti Car', they specify that they need a car only from the brand 'Maruti' rather than any other brand like 'Toyota or

Hyundai'. The more specific a user is the more appropriate results they get. These keywords are used in SEO to increase the page rankings we will see how this technique helps in SEO. Initially we identify a set of words that will be used in optimization as keywords. After finding these set of words, we can determine which keywords are mostly used by the users and analyze the frequency of their occurrences. As we discussed above the more specific keywords a user types the more appropriate results they get, example alone with the brand name they specify car model 'Maruti Suzuki Celerio', it increases the number of search result and limit the search diameter for the user.

Another concept in SEO keyword is duplication that is repeated words in URL's or content of your website

pages. This affects the search rankings of your page. In duplication there are more number of similar chunks or as Google calls it, "appreciably similar, these content being in more than one location on the Internet can make it tough for search engines to determine results to a given search query .Duplicate content can cause issues that are not majorly harmful but trouble search engines and site owners in different ways. It gets hard to determine versions and how to rank them for searches, as for site owners it cause rankings and traffic loss, this makes it difficult for the owners to provide the best search experience and search engines thus are forced to choose which version is most likely to be the best result[10]. Example there exists URL's like 'YAHOOO' or 'YAAHOO' or 'YAHHOO' duplicated to a famous unique id 'YAHOO' this transfers the traffic from the main page to duplicate once if the users have a typing mistake this can also cause security issues. Google clearly says that the method of creating your text distinctive, using low- quality techniques like adding synonyms and related words. The positioning of integrated keywords also affects positively for better search results [10].

2. Meta Data

Metadata is data that describes about other data. It is a set of small connections between your site to all the information about a page that you send to a search engine. It works in the background within the HTML pages. Meta data have meta tags that are actually attributes of a tag. For example :<meta name="" content=""/>, here name is the webpage name and content is the meta description that you provide about the webpage.

3. Meta tags

In order for the Search engines to understand the Meta data we use meta tags. They are snippets of text that are small text or source code that can be inserted into the existing code of a program or Web page. It is easy for common and frequently used code or functions to be implemented into larger sections of code. It provides a small region of re-usable source code, machine code, or text.

4. Title Tag

Title tags are major component of on-page optimization strategy. Each page includes title tag optimized for SEO, they represent the title of a webpage. These are the metadata elements, represented by <title> element. It is crucial to choose right title for your page as it acts as the main attraction to your page and highly affects the search engine ranking. Words at the beginning are of most importance, mostly title tags consist of

keywords that describes your business and the brand name of your business. Titles must be short, consistent and easy to remember around 55 characters or less. They can vary depending on the keywords of the particular page on your website that are also displayed on the Webpage head. Making the title tag longer is not forbidden but it will only make them difficult to be displayed in the search results. As the size of the devices on which the webpage is viewed is varied, the size of title that is displayed also varies. Using hyphens to separate keywords and phrases show different results[11] .

5. Website/URL :

The website URL follows HTTPS (Hyper Text Transfer Protocol Secure is security protocol that ensures security to your website. They provide end to end encryption between sites which is necessary to ensure trust between the communicating parties. The "Area in Green" means the website is safe to use for any kind of transaction, purchase, personal information etc. It is a guarantee that the website is not a fake as it is certified by Certificate Authority or CA. Although green is sign for GO there exists many fake sites that can buy certification like any other true sites, For example "Amazon.com" the actual website and "Amazoon.com" typing one wrong alphabet can get the user in trouble, and you will find users spending money and time over nothing. Therefore it is required that your website has verified legal identity and it is visible to the user. Extended Validation Certificate (EV) is an extra security provided to your website, they are more expensive. They don't play a large role in SEO but are important as well.

Any nonstandard characters separate the keyword forward slashes included, they are just ignored by Google Search engine, so if you search for "Amazon/Women/shoes" Google will just replace it with a space "Amazon Women shoes"

Changing URLs for SEO is not always the Best option. While you can always change the URL after the fact, remember, that page is likely already ranking under the existing URL. Changing it later would mean it would lose some of the history and authority it has built up. If you want to change URL then change it with a permanent 301 URL. A 301 URL redirects users and search engines from your old site or a site to another site that is different from the one they originally typed into or searched for. If a website changes it's URL from the current URL, it is necessary to ensure the users get sent to the right address [13]. This can help maintain search rankings of the website incase when we need to change the URL for any reason.

6. Meta Description

Meta description is summary of the page of your website (Google recognizes this according to the query searched). It is a block of text under the title and URL. Meta Description basically describes what page is about and that block of description is shown on the search engine page. Meta description cannot be seen on the actual webpage as it is written in the source code. You can even view the Meta description by using a particular browser add on.

For example:

Title: "Related Synonyms, Related Antonyms | Thesaurus.com",

URL: "https://www.thesaurus.com/browse/related",

Description: "Synonyms for related at Thesaurus.com with free online thesaurus, antonyms, and definitions. Find descriptive alternatives for related".

The words that are bold are of our search query For example: you searched for "words that start with vowels", each of these words together or individually will be bolded in the meta description of the website that the search engine has optimized. It should always describe the content. It has a character limit (160 characters), if it's too long it will be cut by the search engine.

B. Off-Page SEO

Off-page optimization is the criteria which consist of factors related to the website's environment like the number of incoming links, the page rank score, the domain, directories, etc. Basically it can be defined as the things you do outside of your site to improve the traffic on your website. It concerns establishing relations between websites, advertising and making your website known, linking with other sites affects your website widely in a positive direction.

1) Actions to take outside of your website to improve the website

It indicates indirect influence of SEO by engaging your user base. Social Media Factors don't affect your ranking. Google crawls all accessible links. Tweets are crawled and indexed so the more people who tweet about you or your brand, the more opportunities a reference to your brand with a link to your site has of showing up in search results. Won't boost site authority though can improve link sharing. Increase social media visibility in web searches.

*Focus on Public Relation.*Gaining authority to off-site SEO is Similar. An algorithm that analyzes web links to determine the relative importance of websites. When a website links to other website it is counted as a vote. Each page is assigned a page rank value (0-10). Higher the rank, higher it is indexed on Google. The more page rank a site has the more it can link to other sites and this passing of page ranks is known as LINK JUICE. This concept refers as page authority in today's world. Importance of links has decreased, quality of linked is taken into consideration.

2) Acquiring inbound links.

Links should be earned naturally. Google has improved on weeding out poor quality links. Main Link judging factors. Amount of links your website has, quality of the links, relevance of the links, placement of the links- links found in and to the body are considered to have the most value since they are directly related to the purpose of the site and appear most authentic. Keyword anchor text should be good but not used too much.

C. Technical SEO

Technical SEO focuses on how well search engine spiders can crawl your site & index your content and maximize the usability. It is important to issue technical glitches on your website. Content might be undiscovered without the proper foundation. Your website should support on page SEO recommendations. It is necessary for creating optimized website. Hence we agree that technical SEO is the necessary basis of search engine rankings. Check indexing that is the number of your website's pages that are indexed by search engines. Make sure important resources of your website are crawl able and Optimize crawl budget. Audit internal links, they help spread ranking power and provide a good structure basis. Test and improve page speed and make it mobile friendly.

1) Improving Website

Make Content Relevant to the theme. Match content to topics that explain your website better. Organizing the blogs within sub directories helps the search engines to identify the categories of the web pages. Unique content is very important it attracts the user and makes them stay on the site longer. Don't duplicate content from other sites as Google ranks them lower. Break up the keyword use for example: "Movie reviews" these two words can be used differently unless occur naturally. Use synonyms for more simple words so that every user ids familiar with the content. Link other relevant pages. Don't over optimize with same keywords.*How a site understands and is scanned by Search Engines.*

An important element of technical SEO is creating site architecture. It begins with an appropriate hypertext transfer protocol; it provides security and protection to your website. Next use breadcrumb, which is a type of navigation that shows the location of a user on your website. It is a type of website navigation which presents website hierarchy and indicates where user currently is. It also reduces the number of actions a user has to take when they need to go back to the homepage, different section or any other page. We can have a

clear and consistent URL structure, they describe the page both for users and search engines. They can also act as a ranking factor if used correctly. The robot.txt file and sitemap are elements of wisely designed site architecture.

2) Structure and code of the website.

Sitemap helps search engine to point pages on your site. A Sitemap is an XML file that lists URLs for a site along with additional metadata about each URL. XML site map as well as HTMLS sitemap can be created on Xml-sitemaps.com but only 500 pages are allowed to be crawled and Screaming frog crawls 500 pages . They are easy to read and understood by users, made for search engines, read by users, read by search engine robots. It is a simple page that provides links to other pages within a site. Provides Search engine with extra data about a page when page is updated, and also gifts information on how often the pages are changed. They are extra useful for new undiscovered sites. We can upload the xml site map on Google or Bing to make the search engine aware that the website exists.

IV. TECHNIQUES OF SEARCH ENGINE MARKETING

Empowered by continuously evolving information technologies, Internet search engines have been both an information-seeking vehicle and a flexible online marketing instrument for all kinds of businesses [6]. Search Engine Marketing Professional Organization (SEMPO) states that [7], the total search engine marketing industry size estimate in 2011 is $19.3 billion, with an increase of 16% from that of 2010 ($16.6 billion)Generally, "search engine marketing" refers to paid search marketing (PSM), a system where businesses pay Google to show their ads in the search results. Search engine optimization, or SEO is different from SEM because businesses don't pay Google for traffic and clicks; rather, they earn a free spot in in the search results by having the most pertinent content for a given keyword search.

A. SEM Keyword Research

You need to conduct comprehensive research as part of your keyword management strategy before choosing which keywords to use in your search engine marketing campaigns. First, you need to identify keywords that are related to your business and that target customers are likely to use when searching for your products and services. This can be accomplished by using any of the various Keyword Tools available.

In addition to helping you find keywords you should be bidding on, thorough keyword research can also help you identify negative keywords – search terms that you should exclude from your campaigns. Negative keywords aren't terms with negative implications, but rather inapplicable terms that are highly unlikely to result in conversions. For example, if you sell cakes, you might want to exclude the keyword "cake recipes", as users searching for cake recipes are unlikely to be in the market for your product. This term is known as search intent. There is a high commercial intent for some keywords and have high chances that the buyer will purchase the product.

One of the most enduring misbelieve about search engine marketing is that whomever has the largest advertising budget wins. Although a larger advertising budget can certainly be desirable, especially when targeting highly competitive keywords, but it's far from a requirement for success with search engine marketing. This is because all ads go through a process known as the ad auction before appearing in the search results.

B. How the Ad Auction Works

Every single time someone enters a search query into Google, the ad auction process takes place. To be entered into the ad auction, advertisers pick out the keywords they want to bid on, and state how much they are willing to spend (per click) to have their ads appear in the results relating to those keywords. If Google determines that the keywords you have bid on are included within a user's search query, your ads are entered into the ad auction.

C. How Ads 'Win' the Ad Auction

Not each single ad can seem on each single search.

This is because the ad auction takes a variety of factors into account when determining the placement of ads on the SERP, and because not every keyword has enough commercial intent to justify displaying ads next to results. However, the 2 main factors that Google evaluates as a part of the ad auction method square measure your most bid and also the Quality Score of your ads.

Maximum bid is the maximum amount you have stated you are willing to pay for a click. Quality Score is a metric based on the final or net quality of your advertisement. Google calculates these metrics during the ad auction to decide placement of advertisements. The results of this calculation is understood as ad rank. Given that Google Ad Words' Quality Score comprises half of the ad rank formula, it is one of the most critical metrics search engine marketers can focus on. High Quality Scores can help you reach better ad position at lower costs, because Google favors ads that are highly relevant to user queries.

Search engine marketing's greatest strength is that it presents advertisers the opportunity to put their ads in front of motivated customers who are ready to buy at the precise moment they're ready to make a purchase. No other advertising method can do this, which is why search engine marketing is so successful and such an amazingly powerful way to grow your business.

V. COMPARISON TABLE BETWEEN SEO AND SEM

Parameters	SEO	SEM

Time	Takes around 90 days for the results.	Results are instant and can easily create traffic for website.
Algorithm	Changes are made from time to time.	It is not affected by algorithm updates.
Cost	It is free of cost and doesn't rely on the click volume.	It is paid and regular investment has to be made.
User Behavior	Natural or organic results are more trusted by the user.	People don't trust paid listing that much.
Return on Investment (ROI)	It will increase in future because website will be ranked higher by the search engine.	Return on Investment can increase but there is no guarantee.

VI. APPROACH TOWARDS SEO AND SEM

Organic search results are the "natural" results found beneath the ads within the SERPs. They are natural within the sense that the computer program formula analyzing all the pages across the net felt these pages contained the foremost relevant data for your search.

The formula has set that these websites can have the foremost helpful data inline
with what you're attempting to find that is why they rank extremely for the keywords you typed in your query.

The page is scanned for numerous things (like correct title tags, Meta description, URL, keywords, amazing content, etc.) and the more relevant the information, the higher the ranking.

To optimize your page so as to rank within the organic search results, you wish to use SEO (search engine optimization) techniques to form your page as fascinating to go looking engine algorithms as possible.

Some firms dish out heaps of cash to specialized SEO agencies so as to rank on the primary page – or the number one goal, to rank as the first organic result.

However, ranking with SEO may be a long run method, and sometimes takes upwards of six months to climb the ranking Paid search results are essentially ads obtained by businesses so as to rise higher than the organic leads to the quickest approach attainable.

Marketers invest cash into SEM (search engine marketing) so as to spice up their web site to (ideally) the primarypage once somebody sorts associate inquiry victim ization specific keywords. For ads showing up on Google, marketers use Ad Words to craft ads and bid on placement opportunities in the auction.

On the SERP, you'll be able to typically tell the paid results from the organic ones with visual clues that distinguish them.

Paid results will usually appear first on the page and have a small outlined 'Ad' label on the left-hand side of the URL .It's assumed that whoever pays the most in the Ad Words auction will have the best visibility, but much like the rules surrounding who ranks organically, Google has processes in place (quality score, copy rules, landing page quality, and more) to make sure that the ads that rank are still highly relevant for searchers.

Marketers ought to use each paid and organic search ways to contend in today's digital landscape.

Organic search results, typically speaking, have an evergreen presence.

That means they'll still climb up the rankings over time, assuming the information remains relevant.

Paid ads, on the opposite hand, stop displaying as soon as you stop paying.
While there are execs and cons to investment in paid and organic search ways, the best strategy of them all is to find a balance between the two.

It's necessary to focus on customers at each the highest and therefore the bottom of the funnel (and everyplace in between) so as to capture additional leads for your business.

By optimizing the site using SEO tactics, one can ensure that your website content is showing up when people are digging through the organic results – potentially for years to come.

Not forgetting about those nearing the bottom of the funnel, a strong paid search game ensures that the business is front and center for those ready to make a purchase. If the goal is to make it easy for potential customers to discover the business online, then a strong marketing strategy would utilize both organic and paid search techniques. A combination will deliver the best results.

VII. CONCLUSION

Optimization is a very important internet marketing strategy that is the primary need of every company. The optimized

website appears at a better rank of the search engine result. This research paper is based on review in the available SEO and SEM techniques, i.e. On-site and Off-site techniques to optimize the website to make the website search engine friendly. These techniques can help to improve the ranking of a website if used in a proper manner as suggested in the paper. To ensure that the website's target users can continue to easily find their work through search engines, web developers should keep up to date with the evolving search algorithms, SEO practices, their website's traffic, and their competition. The paper shows the comparative study between the various SEO and SEM techniques. We have also suggested some observed methods which we believe to be better than others for optimization. The lack of visibility of accurate content can be reduced, thus making the user experience good in terms of the organic searches.

REFERENCES

[1] Kumar, L., Kumar, N.: SEO technique for a website and its effectiveness in context of Google search engine. Int. J. Comput. Sci. Eng. (IJCSE) 2, 113–118 (2014).

[2] . Sagot, A.-J. Fougeres, E. Ostrosi, P. Lacom, "Search engine optimization: from analysis based on an engineering meta-model towards integrative approaches", 2014 International Conference on Information Society, pp. 274-281, 2014

[3] Bansal M, Sharma D. Improving Webpage Visibility in Search Engines by Enhancing Keyword Density Using Improved On-Page Optimization Technique. International Journal of Computer Science and Information Technologies 6(6): 5347–5352 , 2016.

[4] Search Engine Optimization [online] Available, [online] Available: http://www.denningesolutions.com/website-design/search-engine-optimization/seo-process/.

[5] W. Dou, N. Zhou JoyceYosephLemos, R. Joshi Abhijit, "Search Engine ptimization to Enhance User Interaction", International Conference on I-SMAC ,IoT in Social, 2017

[6] , K.H. Lim, N. Cui, and C. Su, "Brand Positioning Strategy Using Search Engine Marketing", Management Information Systems Quarterly,SEMPO, "The State of Search Engine Marketing 2011, Survey of Advertisers and Agencies Search Engine Marketing Professional Organization (SEMPO)", July, 2012

[7] Search Engine Optimization [online] Available, [online] Available: www.digitaldeepak.com

[8] How 'Near Me' helps us find what we need, not just where to go[online] Available, [online] Available: https://www.thinkwithgoogle.com/consumer-insights/near-me-searches/

[9] Duplicate Content SEO Advice From Google[online] Available, [online] Available: https://www.hobo-web.co.uk/duplicate-content-problems/

[10] Google: Hyphens In Search Queries Can Change Results[online] Available, [online] Available: https://www.seroundtable.com/google-hyphens-in-search-queries-23965.ht Redirect, and When Should You Use One?[online] Available, [online] Available: https://blog.hubspot.com/blog/tabid/6307/bid/7430/what-is-a-301-redirect-and-why-should-you-care.aspx

[11] Do Social Signals Drive SEO? Available, [online] Available: https://www.convinceandconvert.com/digital-marketing/do-social-signals-drive-seo/

[12] [14]A. J. Su et al., "How to improve your search engine ranking: Myths and reality", ACM Transactions on the Web (TWEB), vol. 8, no. 2, pp. 1-25, 2014.

[13] [15]D. Gaur, P. Dagar, "SEO techniques: New dimensions for popular search engines", Int. J. on Advanced Comput. Theory and Eng., vol. 3, no. 4, pp. 2319-2526, 2014.

[14] [16]K. Rehman et al., "The foremost guidelines for achieving higher ranking in search results through search engine optimization", Int. J. of Advanced Sci. and Technology, vol. 52, pp. 101-110, 2013.

[15] [17]D. K. S.V. Omprakash, "Concept of search engine optimization in web search engine", Int. J. of Advanced Eng. Research and Stud., vol. 1, no. 1, pp. 235-237, 2011.

[16] [18]T. P. Novak, D. L. Hoffman, and Y.-F. Yung, "Measuring the customer experience in online environments: A structural modeling approach," Marketing Science, vol. 19, pp. 22-42, 2000.

Implement System to Predicttrends from Business Report Using Machine Learning

Vishwa Solanki, Raj Shah
Engineering Computer Thakur College Of Engineering And Technology
Mumbai, India solankivishwa273@gmail.com, shah.raj1407@gmail.com

***Abstract*—Business trends prediction is the act of trying to determine the future value net worth and sales traded on a financial exchange.It also calculates and predicts year on year growth of the company based on the previous datasets- on the previous datasets Machine learning are applied for accuracy. In this project the client will first login into the application and then he can see the graphical view of his company's p- ast sales and by analyzing the past datasets of the company application will predict what ac- tions they need to take to increase their busin- ess in global market.**

I. INTRODUCTION

Predictig the trends has became one of the most important task in financial market.More and Mor business houses are implementing predictions of trends using technologies like (Machine Learing Data Analytics and many more)in order to incre- ase their stake in the world market.The objective of this project is to develop a web Application & mobile application for third party who can anytim e for a third party who can anytime use their past data sets to analyze the trends and then would be able to make right decision to increase their stake and in turn would be able to increase their busin- ess across the world Using this application third party would login using their credentials and then anytime at any place they could see the graphical analysis of their past data sets.

Technology used are HTML,CSS,JS,ANGULAR JS..

II. EASE OF USE

A. User Friendly

Identify applicable funding agency here. If none, delete this text box.
This application will be user friendly in this the

User has to just enter his or her credentials and

Login in that's it and he will get the desire resu-

B. It will work in offline mode

It will also work when there is no internet conn-

ection or if the user is in the flight.

C. Abbreviations and Acronyms

HTML-Hyper Text Markup Language

a standardized system for tagging text files to achieve font, color, graphic, and hyperlink effects on World Wide Web

CSS-Cascading Style Sheet

Cascading style sheets are used to format the layout of Web pages.

.JS-Java Script

JavaScript is a programming language that started off simply as a mechanism to add logic and interactivity to an otherwise static Netscape browser.

D. Equations

It is one of the most widely known modeling technique. Linear regression is usually among the first few topics which people pick while learning predictive modeling. In this technique, the dependent variable is continuous, independent variable(s) can be continuous or discrete, and nature of regression line is linear.

Linear Regression establishes a relationship between dependent variable (Y) and one or more independent variables (X) using a best fit straight line (also known as regression line).

It is represented by an equation Y=a+b*X + e, where a is intercept, b is slope of the line and e is error term. This equation can be used to predict the value of target variable based on given predictor variable(s).

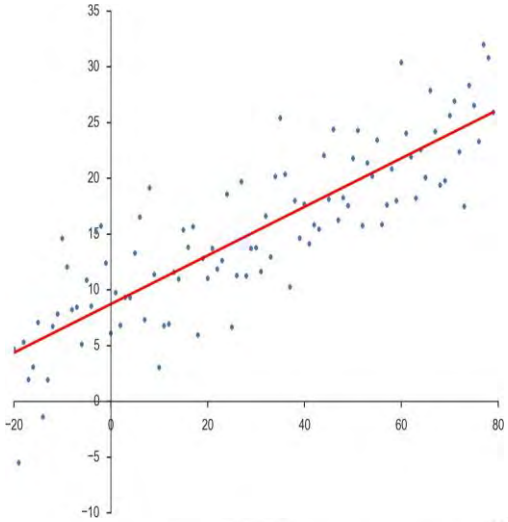
C. Literature survey

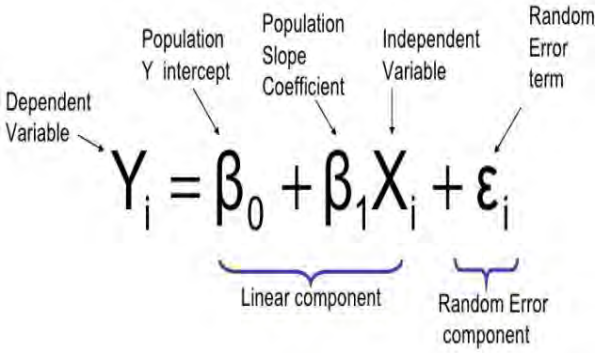
Predicting the trends of financial markets is one of the most important tasks for investors. They have tried to predict the trends using various m- ethods. Technical analysis and fundamental ana- lysis and fundamental analysis are employed in analyzing the trends of Business technical analy sis is one of thetraditional analytical methods th- at makes historical prices and trading volumes to determine the trend of future inventory prices prices. This analysis is based on supply and dem ands.Itcan also be applied to firms with bad fina- ncial condition Fundamental analysis predicts stock prices by using intrinsic values.The marke t values are determined by the financial val statements and economic factors of firms.Investo rs estimate the profits of firms and evaluate whet
her they are proper. However,this approach cannot reflect other factors that affect stock prices, such a s the emotional factor of market participants.Rece ntly,several studies analyze financial markets with the sentiments of investors,such as blogs,news and news, and social network services, emerge

,
\

Figure

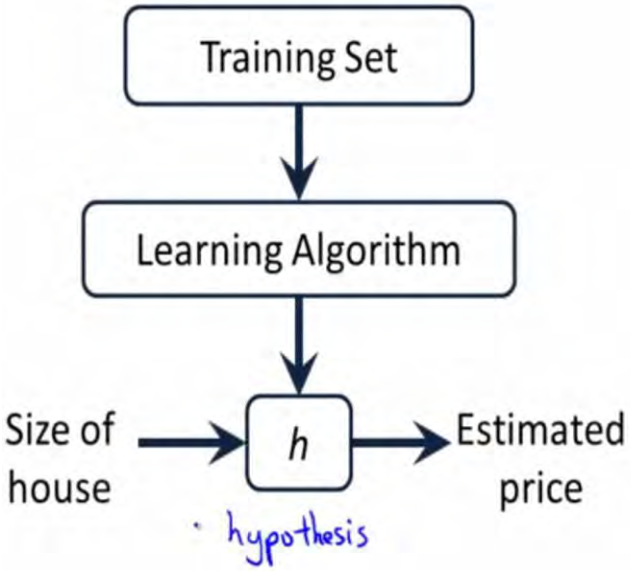
Linear Regression





The following are the advantages of application:
Easy
prediction
improve
business
More
customers

Example



III. CONCLUSION

Predicting stock prices has been a major issue to investor and researchers. Many methodologies from various acade mic fields have been introduced for prediction, and some methods have been used in real financial markets for trade strategies.Recently,researchers have analyzed the trends of stock prices using machine learningmodels, and they have shownconsiderable performances by usingmeaningful input factors.

IV. REFERENCES

[1] Rechenthin, Michael David. Machinelearning classification techniques for the
analysis and prediction of highfrequency stock direction. The University of Iowa, 2014.
[2] Milosevic, Nikola. “Equity forecast: Predicting long term stock price movement using machine learning.” arXiv preprint arXiv:1603.00751 (2016).
[3] Shen, Shunrong, Haomiao Jiang, and Tongda Zhang. “Stock market forecasting using machine learning algorithms.” Department of Electrical Engineering, Stanford University, Stanford, CA (2012): 1-5.
[4] Bonde, Ganesh, and Rasheed Khaled.
“Extracting the best features for predicting stock prices using machine learning.” Proceedings on the International Conference on Artificial Intelligence.
[5] Freitag, “Design methodologies for LSI circuitry”, IBM Tech, Place, Rep TR41736, pp 80-82, 1983.
[6] Hegazy, Osman, Omar S. Soliman, and Mustafa Abdul Salam. ”A Machine Learning
Model for Stock Market Prediction.” arXiv preprint arXiv:1402.7351 (2014).
[7] Dai, Yuqing, and Yuning Zhang. “Machine Learning in Stock Price Trend Forecasting.”
(2013

Emotion Classification in Speech Signals Using Feature Space Reduction and SVM

Mr. Vishwajit B. Gaikwad, Mrs. Ujwala V. Gaikwad
(Terna Engineering College, Nerul, Navi Mumbai, India)

ABSTRACT : *Emotion recognition from speech signal is a challenging task due to the complexity of emotions. The performance is heavily dependent on the effective selection of features extracted from the speech. Most features are sensitive to emotionally irrelevant factors, such as the speaker, environment, and speaking styles. Speech emotion recognition has considered as an important research area over the last two decades. Many researchers have been attracted due to the automated analysis of human affective behavior. Therefore a number of systems, algorithms, and classifiers have been developed and outlined for the identification of emotional content of a speech from a person's speech. Inspired by this, a method is proposed based on reduced combinational features and support vector machine to recognize emotions from input speech signal from Berlin Dataset. The experimental results reveal that the combination of MFCC and DWT based features has improvement in accuracy.*

I. INTRODUCTION

In past two decades, the researchers highly attracted towards emotion recognition in the field of speech signal processing and analysis. There is an enormously important role of emotions in human life. As per human's perspective, feelings or emotions are essential medium of expressing his or her psychological state to others. Humans has normal ability to recognize an emotion through spoken words but since machine does not have capability to analyze emotions from speech signal, machine emotion recognition using speech signal is very difficult task. Automatic emotion recognition paid close attention in identifying emotional state of speaker from voice signal[1]. An emotion plays important role for better decision making and there is a desirable requirement for intelligent machine human interfaces[1,3]. The detection and identification of the user emotional state through his/her voice signal or extracted feature from speech signal means emotion recognition through speech.

It is principally useful for applications which require natural machine human interaction such as E-tutoring , electronic machine pet, storytelling, intelligent sensing toys,

The last step of speech emotion recognition is classification, it involves classifying the raw data into particular class of emotion on the basis of features extracted from the data. There are a variety of machine learning algorithms that have been examined to classify. In recent years in speech emotion recognition, researchers proposed many classification algorithms, such as Gaussian Mixture Model(GMM)[11], Hidden Markov Model (HMM)[11], k-nearest neighbours (KNN), Support Vector Machine (SVM)[12], Artificial Neural Network (ANN)[9] and Recurrent Neural Networks (RNN) [8,14]. In fact, there has been no agreement on which classifier is the most suitable for emotion classification. In recent year, the extremely complex nature of human emotional states makes this problem more complicated in terms of feature selection and classification.

There are various challenges in emotion recognition from the speaker's voice due to certain reasons such as, existence of the differ in speaking styles, speakers, sentences, languages, speaking rates introduces accosting variability affected different voice features this a particular features of speech are not capable to distinguish between various emotions also each emotion may correspond to the different portions of the spoken utterance. The same utterance may also in the car board system application where the detected emotion of user makes it more practical[1]. Emotion recognition from speech signal is useful for enhancing the naturalness in speech based human machine interaction. Automatic emotion recognition through speech provides some other applications such as speech emotion recognition system used in aircraft cockpits to provide analysis of Psychological state of pilot to avoid accidents, speech emotion recognition systems also utilized to recognize stress in speech for better performance in lie detection, in Call center conversation to analyze behavioral study of the customers which helps to improve quality of service of a call attendant also in medical field for Psychiatric diagnosis, emotion analysis conversation between criminals would help crime investigation department.

II. LITERATURE SURVEY

There are different features utilized for recognizing emotion from speech signal such as spectral features and Prosodic features[3,11,13]. The feature extraction methods like MFCC (Mel Frequency Cepstrum Coefficient) provide the highest accuracy on all databases provided using the linear kernel[6] and the spectral coefficients derived from LPC (Linear Predictive Coding) [7], Many researchers[4,6] have proposed important speech features which contain emotion information, such as energy, pitch, formant frequency, Linear Prediction Cepstrum Coefficients (LPCC), Mel-Frequency Cepstrum Coefficients (MFCC) and modulation spectral features (MSFs)[6].

Palo HK, Mohanty MN[10] used the DWT and MFCC with Stability-Plasticity Dilemma Neural Network. In this work, three different types of Daubechies wavelets (db4, db6 and db8) were analysed. Features such as range, pitch average, intensity, pitch change, maximum value, minimum value and a fusion of all the specified features were used to obtain the feature vector[11].

show different emotions & hence recognized emotional states which are not clear[4]. Significant research was done in these last two decades and there are significant results, but still new and better solutions for emotion detection are required and expected.

III. EMOTIONAL SPEECH DATABASE



Figure 1. Plutchik's wheel of emotions

As per his classification, there are 8 fundamental emotions (Table 1), grouped in 4 pairs of opposites.
All of them visible in various degrees of intensity and their combination results in secondary emotions.

Table 1. The fundamental emotions by Robert Plutchik [1,2]

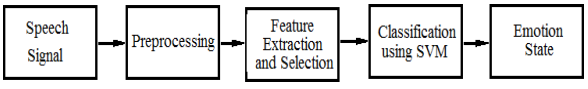
which includes different emotional states of human being. R. Cowie et.al constructed their own English language

sadness etc. Researcher M. Liberman, et al utilizes the database consists of 9 hours of speech data which contains speech in 15 emotional categories, such as hot anger, cold anger, sadness, elation, panic-anxiety, despair, happiness, interest, boredom, shame, pride, disgust and contempt, constructed at the University of Pennsylvania. Berlin emotional speech database contains about 500 simulated speech samples, which are simulated by professional actors [7,8]. Most of the researcher have used this database for evaluation of their contribution in emotion recognition. Berlin database in which each speech sample corresponds to one emotion is used to evaluate performance of proposed method for emotion recognition.

IV. PROPOSED METHOD

Automatic Emotion recognition system through speech is similar to the typical pattern recognition system. In this study, the block diagram of the emotion recognition system using speech considered illustrated in Figure 2. some

Figure 2: Proposed speech



American psychologist Dr. Robert Plutchik has classified emotions in different schemes. The Emotion Wheel was developed by him which represents complex scheme in graphical form is shown in figure 1, known as Plutchik's wheel[1]

Joy	Sadness
Trust	Disgust
Fear	Anger
Anticipation	Surprize

The effectiveness of recognition is highly depends upon the naturalness of database used in the speech emotion recognition system. The collection of suitable database is most crucial task concerning to an emotion recognition system using speech signal. An important parameter to evaluate the performance of emotional

speech recognizer for detecting emotions is the degree of naturalness of the database used. Design of a good quality of database is required to evaluate effectiveness of emotion recognition system. If poor database used then inaccurate recognition occurs. Thus for better classification the design of the database plays significantly important role in classification. More realistic and natural speech samples from real life situations are expected in database. But it is difficult to collect such samples due to some natural parameters such as noise included at the time of recording. Different researchers contributed to design good databases

emotional speech database for 6 emotional states such as fear, anger, happiness, disgust, neutral,

Proposed system is based on prosodic features and some spectral features of voice signal. It consists of the emotional state speech sample pre-processing , train & test sets of inputs, generation & selection of features, classification of Emotional state using Support Vector Machine and recognition of emotion state as the output.

Survey shows that the MFCC gives good results as compare to other features for speech based emotion recognition system[11]. MFCC is the most axiomatic and popular feature extraction technique for speech emotion recognition. The speech signal decomposed in sub bands by using DWT in three levels. Mean energy is calculated for each sub band. The feature vectors are obtained from the detailed coefficients applying statistics mean, standard deviation. Also features such as skewness, and kurtosis are used to obtain the feature vector. The final feature vector is obtained by applying PCA on augmented normalized feature vectors. These feature values will be provided to the SVM classifier as a test input. Then classifier classifies the test sample into one of the emotion & gives output as recognized emotion from the above mentioned five emotional state as Angry, Disgust, Happy, Neutral and Fear.

emotion recognition system framework

Features	Angry (%)	Disgust (%)	Happy (%)	Neutral (%)	Fear (%)	Average accuracy (%)
MFCC	64	63	60	55	57	59.8
MFCC, DWT	86	82	79	77	78	80.4
MFCC, DWT, PCA	84	79	78	77	76	78.8

V. EXPERIMENTAL RESULTS

In this study, five different emotional states are considered and from different emotional state speech input signal features were extracted. Berlin emotional speech database is used to evaluate performance of proposed method. MFCC Three levels of DWT of DB4 applied on speech signal to obtain mean, standard deviation, skewness, and kurtosis features. These features were combined with MFCC and PCA to reduce feature vector dimension. Proposed method is simulated on MATLAB 2017 and results compared with methods from literature.

As shown in the table 2, the classification results for emotion recognition using SVM classifier with respect to particular mode of different emotional state is calculated in which for Anger state classifier correctly classified testing speech sample with the recognition rate of 84% as anger. For emotional state happy classifier correctly classified at the recognition rate of 73%. The fear state was classified as fear at 72%. The disgust state was classified as fear at 78%. The neutral state were correctly classified at 70% as neutral state.

Table 2: Comparison of classification accuracy with traditional and proposed features.

The impact of feature reduction with the misclassification percentage is shown in above table. This table also provide a comparison on percentage accuracy for classification of these emotions with our proposed feature sets. It indicates improvement in performance with reduced features than simple features hence strengthens the novelty of the proposed feature sets.

VI. Conclusion

In this paper, utilization of spectral and prosodic features were extracted from speech signal with five different emotional state through speech signal using SVM classification were analyzed. Speech Emotion Recognition has a promising future and its accuracy depends upon the combination of features extracted, to increase the performance of system combined features utilized. The feature were extracted from emotional speech samples such as mean, standard deviation, skewness, and kurtosis after applying three level DWT, Mel frequency cepstrum coefficient (MFCC) feature. PCA is used to reduce vector dimension to provide better classification and efficiency. The added advantage is that since feature vector dimension is very small, it can be used for near to real time recognition applications. The efficiency of system is extremely depends on proper database of emotional speech sample. Emotional speech database with affective states, improving the performance and engagement of the current interfaces with machine. Therefore it is necessary to create a proper and correct emotional speech database. Emotion recognition system can provide more efficiency with combination of different classifier or implementing hybrid classifiers for better recognition rate. Future research is focused to develop Marathi language speech database for emotion recognition.

References

[1] Plutchik, Robert, The nature of emotions, American Scientist 89 (2001), page 344.

[2] Dimitrios Ververidis and Constantine Kotropoulo, " A Review of Emotional Speech Databases", avialble online on http://delab.csd.auth.gr/bci1/Panhellenic/560ververidis.pdf (accessed on 25/08/2018

[3] Zhou y., Sun Y., Zhang J, Yan Y., Speech Emotion Recognition using Both Spectral and Prosodic Features", IEEE, 23(5), 545-549, 2009.

[4] Ayadi M. E., Kamel M. S. and Karray F., "Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases", Pattern Recognition, 44 (16), 572-587, 2011.

[5] Wang K, An N, Li B N, Zhang Y., "Speech emotion recognition using Fourier parameters". IEEE Trans Affective Comput 2015; 6(1):69–75. January-March.

[6] Galvo, Rafael and Sidney DMello, Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications, IEEE Transactions on Affective Computing, VOL. 1, NO. 1, 2010.

[7] Khanna P, Kumar MS. Application of vector quantization in emotion recognition from human speech. ICISTM, CCIS 141. Springer; 2011. p. 118–25.

[8] Weninger, Felix, Martin Wllmer, and Bjrn Schuller, "Emotion Recognition in Naturalistic Speechand Languagea Survey", in book: Emotion Recognition: A Pattern Analysis Approach, Published Online: Wiley Online Library, pp.237-267, JAN 2015.

[9] Michael Neumann, Ngoc Thang Vu, "Attentive Convolutional Neural Network based Speech Emotion Recognition: A Study on the Impact of Input Features", Proceedings of Interspeech 2017, Sweden, Stockholm, August 20-24, 2017.

[10] PALO HK, MOHANTY MN. "Wavelet Based Feature Combination for Recognition of Emotions", AIN SHAMS ENG J (2017), http://dx.doi.org/10.1016/j.asej.2016.11.001

[11] Sara, M., Saeed, S., and Rabiee, A. "Speech emotion Recognition Based on a ModifiedBrain Emotional Learning Model", Elsevier, pages 32–38, 2017.

[12] B. Schuller, G. Rigoll, and M. Lang, "Speech Emotion Recognition Combining Acoustic Features and Linguistic Information in a Hybrid Support Vector Machine – Belief Network Architecture," Acoust. Speech, Signal Process., vol. 1, pp. 577–580, 2004.

[13] Tejal Udhan and Shonda Bernadin, "Speaker-Dependent Low-Level Acoustic Feature Extraction for Emotion Recognition, The Journal of the Acoustical Society of America 143, 1747 (2018); https://doi.org/10.1121/1.5035705, 17 April 2018

[14] P. Staroniewicz, "Influence of Speakers' Emotional States on Voice Recognition Scores", in Proc. Anal. Verbal Nonverbal Commun. Enactment, Process. Issues, Budapest, Hungary, 2010, pp. 223-228.

A Technological Solution For Cleaning And Maintenance Of Sewerage System

Heena Ganesh Shirke Prathamesh Vijay Rawool Akshata Anand Salvi Information Technology Department Information Technology Department Information Technology Department Saraswati College of Engineering Saraswati College of Engineering Saraswati College of Engineering
Kharghar, India Kharghar, India Kharghar, India heenashirke09@gmail.com prathameshvijay.rawool20@gmail.com akshata.salvi27@gmail.com
Diksha G Kumar Information Technology Department Saraswati College of Engineering
Kharghar, India diksha.gkumar@gmail.com

***Abstract*—Thousands of sewer cleaners are risking their lives every month in the Indian business hub of Mumbai because of a lack of basic safety gear. The workers, who are mostly from low- caste communities, die before age 60 because of work- related health problems. In Mumbai, an average of 20 sewer workers die each month from accidents, suffocation or exposure to toxic gases, the study found that they have to deal with filthy conditions and dangerous gases. Workers has to clear drainage and sanitation lines using metal scrapers or their bare hands[6]. Current alternative measures proposed include Robots and Sewer jetting which have their own drawbacks in the form of cost-effectiveness, trained technicians, ineffective in cleaning muddy areas, damage to fragile pipes due to high pressure in jetting, inappropriate for large sewer networks, designing control algorithms as per fluctuation needs etc. So to overcome above mentioned issues we propose a system that gives a complete solution for the clogged drains as well as its timely detection.**

***Index Terms*—clogged drain, blockage, sewerage pipe**

I. Introduction

"Former defense minister Manohar Parrikar, in his reply to a parliamentary question in 2015, said that a total of 171 army personnel were "martyred" fighting terrorists between 2012 and 2014 country wide. In the same period, a total of 142 sewer workers died."

As extension of the current sewerage system can be difficult and costly. Due to the pipeline conditions, it is very challeng- ing for the closed pipeline if it got blocked or even clogged that can affect its overall performance. Prior research has identified a number of causes that can influence the pipeline performance. Usually, it is not solid objects that swirl down the drain that causes a clog. The fact is that the clog is gradually built from its internal diameter. After certain period of time, the accumulated clog will reduce the pipeline diameter that consequently restricts free flow of water in the pipeline. Meanwhile, sludge is a formation of sticky concoction of soap scum, grease, hair, food particles or even the dissolved minerals in the water. Once the layer of sludge formed, more sludge tends to accumulate in that segment of the pipeline, increasingly nar- rowing the size of the pipe that will reduce the efficiency of the pipeline. Maintenance costs are high compared to decen- tralized systems and consists mainly inspection, unblocking, repair and maintenance requires human to enter manhole.Currently, the cleaning of sewers and septic tanks is done with the use of jetting and suction machines that use water pressure to clean out sludge. However, suction devices are often incapable of cleaning out blockages in manholes or septic tanks. The same with jetting, which uses the force of water to wash out sewer lines. At these times, it becomes imperative for human beings to enter and clean the clog. The greatest disadvantage of the existing method is that it fails to clean the mud in the form of rock solids due to the deposition of slits[1]. As well as the blockages formed inside a sewer pipe still needs a human entry. In order to overcome the above issue, an attempt has been made to fully automate the maintenance and controlling of the sewer pipes. The proposed machine shall be inserted into the sewerage pipe through the manhole and it will attempt to move the blockage to the nearby manhole by using water jet. If this doesn't work then it will drill the heap so that it is dispersed and then water jet can be used to move the garbage heap to the next manhole.

II. Existing System

The earliest form of sewer cleaning was hand excavation whereby laborers loaded sediment into barrows which were moved down the sewer and then lifted out at manholes by bucket. It is a dreadful process where the drainage wastes are cleaned by the people that may include the basic tools like buckets, brooms for disposing the drainage wastes. Water jet is used to cut a wide variety of material using a very high- pressure jet of water, Sewer jetters, also known as "hydro- jetters" or "water jetters", are powerful drain cleaning ma- chines that use high-pressure water jets to clear obstructions in

residential and commercial drain pipes as well as larger municipal sewer systems.

III. Proposed System

Our proposed system aims at providing 3 major features lacking in the current system. First, timely detection of some

issue or choke up in sewer system. This will facilitate in reducing the incubation period of a major drainage issue. Secondly, this system will be connected to an online portal for each constituency as in each municipal corporation office will have a complete notion of the present sewer's conditions and when an issue nears to a major danger and crosses a threshold then an indication through HMI. And lastly, a machine that is capable of cleaning the clogs generated inside the sewers.

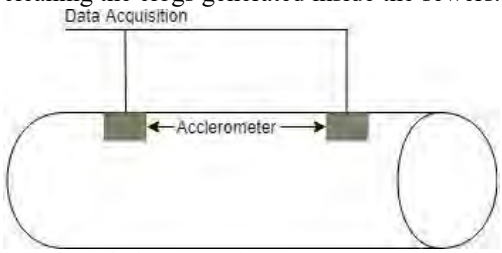


Fig. 1. Blockage Detection

A. Human Machine Interface

1) Accelerometer: An accelerometer is a device that mea- sures changes in gravitational acceleration in a device it may be installed in. Accelerometers are used to measure acceleration, tilt and vibration in numerous devices. It observes vibration at different locations.

2) Transmitter: The whole setup of transmitter part is placed inside the
sewerage pipe. A single transmitter is connected to accelerom- eters of multiple pipes. It acquires data from the accelerome- ters and sends it to the processor.

3) Receiver: Receiver receives the data acquired by trans- mitter and feeds it to the processor.

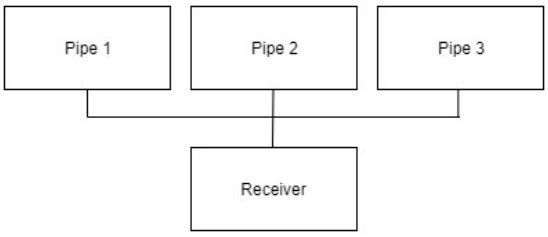


Fig. 2. Data Acquisition

4) Processor: Processor analyses if there is any instance of blockage. The pressure changes are calculated from the velocity profile and shown in Table 1. It can be seen that pressure during blockage drops as the restriction of flow increases. When fluid flow through an orifice, high flow velocity and small restriction in pipe will cause pressure to drop.

5) Human Machine Interface: Human-machine interface (HMI) is a system of software and hardware elements that allows us to:

1) Monitor, gather, and process real-time data, Directly interact sensors, through HMI software, Record events into a log file.

2) The HMI system quickly notifies an operator that a specific area is showing a high incidence of blockage.

B. Blockage Detection

For detection of blockage Accelerometer is put on two different locations. These different points of measurement will help in evaluating the vibration signal behavior throughout the flow. The corresponding acceleration signal response is analyzed.

Fig. 3. Methodology for blockage detection

The streamlines of fluid flow in the pipe is the simulation result of pipe without blockage. The streamlines are distributed almost evenly within the inner pipe area. The streamlines of the velocity profile shows a uniform pattern. Based on [8], a fully developed laminar pipe flow occurs when there is no blockage in the system. As the blockage is getting larger, the streamlines get closer and the gaps between the lines are narrower. These behaviors can be seen clearly when the water flow passes the blockage. The streamlines start to adjust as the fluid begins flowing through the obstacle. Water velocity inside the pipe increases as the flow area decreases.

C. Blockage Cleaner

Rather than installing multiple units in all areas, we are proposing a multi functional machine. As it doesn't need to be installed at a particular area it is dispatch able and also a cost-effective solution. The robot will be capable of cleaning the soft clogs as well as the harder ones. Generally, 70% of the blockages are caused due to soft clogs and the rest of the 30% are caused due to the hard clogs.

Once the blockage has been detected we are supposed to dispatch the machine in the nearest manhole available. The machine will then adjust its size according to the diameter of the sewer pipe. Machine will be compatible to various diameters of the sewer pipe. This grinder works on umbrella mechanism. So, if the size of the pipe is detected to be larger than its current settings the On-Off panels will start working so as to fit into the pipe. If the size of the pipe is smaller than its current settings the panels will contract themselves to make it compatible to

enter the pipe.

The machine will then traverse through the sewer pipe. Once the machine is in proximity to the location where blockage has occurred it will first try to get rid of the blockage by jetting the water at high pressure at an angle. If the jetting doesn't work, then it will start drilling. The head of the machine will have one-degree freedom i.e. moving forward and backward. This degree of freedom will help the drilling

part to be more efficient. The machine will then travel through the pipe towards the next manhole.

1) Water Jet: Water jet is the process of shooting high- powered streams of water through the drain, down into the sewer to blast away any debris blocking the passage of water. This is more effective than using a snake, blades, or even drain rods because, first the water is shot at such a high intensity that the force isn't even comparable to manual labor, secondly the water is much more capable of bending around curved or angular pipes to reach all the tight spots. A water jet is composed of a controlled high-pressure water source such as a pressure washer or reciprocating displacement pump, a flexible high-pressure line of up to hundreds of meters in length, the Mini-Reel and a nozzle that uses hydraulic force to pull the line into sewer drains, clean the sides of pipes, and flush out residue.

2) Drill: If the jetting doesn't work, then the machine will start drilling. The head of the machine will have one-degree freedom i.e. moving forward and backward. This degree of freedom will help the drilling part to be more efficient.

3) On-Off Panel: These panels enable the machine drill to increase its size to accommodate the various sized pipes.

IV. Implementation Strategies

A. Advantages:

- Human life is saved and the proper disposal of drainage wastes[1].
- Timely detection of some issue or choke up in sewer system. This will facilitate in reducing the incubation period of a major drainage issue..
- This system will be connected to an online portal for each constituency as in each municipal corporation office will have a complete notion of the present sewer's conditions and when an issue nears to a major danger and crosses a threshold then an indication through HMI
- The machine is capable of cleaning the clogs generated inside the sewers.

B. Disadvantages[1]:

- Cost effective
- This method requires trained technicians to operate the robot

V. Conclusion

We propose an innovative, fully autonomous, and un-tethered sewer inspection and clearance robot. Inspection is carried out by live streaming of the pipeline. Our machine can adapt to different pipe diameter. The proposed one unit model eliminates the human entry in sewage pipelines. Also, the work of cleaning and removing obstacles would be done automatically. Our proposed system is capable to handle all possibilities. We are glad to work on such an important problem faced by government.

References

[1] S.Sharmili, T.Saarika and R.Malar," Intelligence Drainage Cleaning Using Arm Robot",International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)ISSN: 0976-1353 Volume 24 Issue 7 – APRIL 2017.

[2] Dr. Rajesh Kanna S.K., Ilayaperumal K. and Jaisree A.D, "Intelligent Vision Based Mobile Robot for Pipe Line Inspection and Cleaning",International Journal of Information Research and Review Vol. 03,Issue, 02, pp.1873-1877, February, 2016

[3] Muhammed Jabir.N.K', Neetha John', Muhammed Fayas", Midhun Mohan, Mithun Sajeev, Safwan.C.N"Wireless Control of Pick and Place Robotic Arm Using an Android Application",International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering(IJAREEIE)Vol. 4, Issue 4, April 2015.

[4] Ana SaNura Zainal Abidina,k, Muhammad Hamizan Zainia, Muhammad Farhan Aiman Mohd Pauzia, Mo- hammad Muaz Sadinia, Sim Chia Chiea, Shahrol Mo- hammadana, Annisa Jamalia, Rasli Muslimena, Muhd Fadzli Asharia, Mohd Syahmi Jamaludina ,Chong Yee Minga,"Development of Cleaning Device for InYpipe Robot Application",2015 IEEE International Sympo- sium on Robotics and Intelligent Sensors (IRIS 2015).

[5] Srinivas Panguluri Shaw Environmental & Infrastruc- ture, Inc. Cincinnati, OH Gary Skipper Brown & Cald- well San Diego, CA Steve Donovan Brown and Caldwell Cincinnati, OH," DEMONSTRATION OF INNOVA- TIVE SEWER SYSTEM INSPECTION TECHNOL- OGY",(U.S.

Environmental Protection Agency).

[6] Vaani I1, Simerna J. Sushil2, Vani Kunjamma U.S3, Akshaya Ramachandran4, Dr. Thulasi Bai V.5, B. Thyla 6,"BhrtyArtana (A pipe cleaning and inspection robot)",2017 IEEE 3rd International Conference on Sensing, Signal Processing and Security (ICSSS).

1) Jing Yan1,3, Zao Feng 1,3, Jiande Wu 1,3, Jun Ma 2,"Research on Identifying Drainage Pipeline Blockage based on Multi-Feature Fusion"(29th Chinese Control And Decision Conference, 2017).
2) M. Toophanpour-Rami, E.R. Hassan, R. M. Kelso and J.P. Denier,
"Preliminary investigation of impulsively blocked pipe flow", in 16th
Australasian Fluid Mechanics Conference (AFMC), 2007, pp. 967-970

Migration: From RDBMS to NoSQL Database

Er .Zainab Mirza(*HOD*), Mr. Taha Bikanerwala, Mr. Mohd. Shibli Qureshi, Mr. Husain Amreliwala
Information Technology
M.H.Saboo Siddik College Of Engineering.
Mumbai, India
zainab.mirza@mhssce.ac.in, tahabikanerwala@gmail.com, mohdshibli27@gmail.com, husain.amreliwala52@gmail.com

***Abstract*:-Data today is increasing on tremendous scale making it a challenge for existing systems to handle it properly. Statistics state that on an average about 2.5 exabytes that is, 2.5 billion gigabytes (GB) of data is generated every day as of now. Traditional databases such as SQL databases start to fail miserably once the size of data crosses the maximum data handling threshold. With expanding the size of the data, we also observe the variation in the structure of data. NoSQL database technology was developed to handle this expansion and variety with simplicity and ease. Although one can efficiently develop and test the applications using structured data,but when unstructured and semi-structured operational data comes into the picture, NoSQL technology is the optimal choice in case of efficiency. This paper provides such examples and proofs that argue how one can migrate from RDBMS (SQL) to NoSQL and also one can enhance existing applications by opting data migration to NoSQL.**

Keywords— NoSQL, SQL, MongoDB, Big Data, RDBMS .

I. INTRODUCTION

Structured Query Language (SQL) databases are structured databases based on the concept of relational databases. Tables are the basic unit of SQL databases. Till 2005 - 07, almost every internet application that hosted data creation, updation, deletion and display services i.e in all data management services harnessed their data handling strengths from SQL based RDBMS (Relational DataBase Management System). SQL is the standard language that is supported by such RDBMS for the basic CRUD (Create, Read, Update, Delete) operations as well as the advance operations that consists of JOINS, TRIGGERS, Transactions, etc. Thus SQL technology was the most popular and widely used database technology at that time. But, since 2005 the size of data began to grow on an unexpected pace due to the advent of technology. Sources of data generation multiplied in numbers, playing the role of one of the major contributors to increasing size of data. This turning point in data technology, lead to the introduction of Big Data. The primary obstacle was that SQL was falling inefficient in terms of handling data that was increasing beyond its handling capacity. Although its capacity was

stretched to occupy big data, but this put forth a lot of constraints on the actual functioning of SQL. One of the major issues were not only the size but also the widening of variety. Big Data constitutes the data that spreads in no one domain, but multiple domains. Data gradually began to lose structure and extensive applications endorsing image files, video files made it more difficult to store the actual files in directory filesystem and store their respective paths in the database. This proved cumbersome when retrieval operations came to light. As a counter attack to these issues, already introduced in 1998 NoSQL was reintroduced in early 2009 as a concept of database systems that are non - relational and did not expose the SQL. NoSQL functions at the cost of ACID transactions, but it provides extensible scalability, partitioning and distributed data storage. This paper presents findings and proofs that state the inefficiencies of SQL considering some examples, and efficiencies of NoSQL in that examples as a result of migration from relational algebraic systems to more Big Data-oriented NoSQL systems.

II. LITERATURE REVIEW

Jagdev Bhogal and Imran Choksi in their IEEE publication "HANDLING BIG DATA using NoSQL"(published in 29th International Conference on Advanced Information Networking and Application Workshops, 2015) introduced primary and quite significant comparisons between the NoSQL and SQL systems in the big data context. NoSQL (abbreviation for not only SQL) is the emerging technology in data management domain and serves as a viable alternative to SQL based systems. The fact that is highlighted in this publication is that the major problem to solve is not; how NoSQL can replace SQL, but how NoSQL and SQL can be used in a complementary fashion together. One needs to understand the requirements of data structure and accordingly decide to opt for whether SQL or NoSQL. The author demonstrates that there happens a significant trade-off between three major requirements of data management paradigm name Consistency, Availability, Infinite Scaleout. Although data

stored in NoSQL systems is available anytime, anywhere (depending upon its server), and also they can be scaled on in infinite (preferably huge) scaling magnitude, but NoSQL systems trade-off with consistency and ACID transactions. On the contrary, SQL systems provide consistency and availability at its best, but they make a trade-off with scalability. Big Data cannot be handled using relational databases i.e. SQL. The author demonstrated these peculiarities by implementing a small prototype application using both SQL and NoSQL. For SQL, Oracle APEX was considered and for NoSQL, the document-oriented MongoDB was selected. The author illustrated this difference by comparing two different visual schema definitions of the same application. One was the fixed schema to be implemented in SQL and another was the Document Schema to be implemented in MongoDB. When technically analysed, the author attributed a fact why NoSQL provides better performance is that NoSQL schema can reduce the hassle of multiple joins and complex queries. Normalizing the schema definition in SQL provides clean and segregated structure, but this also requires a number of joins and complex nested queries to retrieve even small amount of relationally linked data, on other hand NoSQL being of no fixed schema, can easily retrieve data with simple function calls (for eg: in MongoDB). The author concluded by suggesting applications of NoSQL in strategic business data and IoT based sensor networks that are nothing but the real-time data. Applications of referential integrity and which possess no

concrete structure are more preferred to be handled by NoSQL systems. Dynamic data model applications are most suitable for NoSQL.

Yishan Li and Sathiamoorthy Manoharan in the IEEE publication "A performance comparison of SQL and NoSQL databases" make an effective comparison between the existing SQL technology for handling big data and the rapidly emerging NoSQL technology. The author puts forth the idea that not all NoSQL base database management techniques are as efficient as SQL but also attributed to the fact that SQL lacks some major features that NoSQL provides. In this publication, the author compares the key - value implementations of both NoSQL and SQL databases. The comparison is based on four major data management operations viz. Read, write, delete, retrieval. The author provides introductory information about the initial implementations of NoSQL technology namely Google's Big Table and the Amazon's Dynamo. The author further provides information about the various existing implementations of NoSQL, among them are MongoDB, Hypertable, Apache CouchDB, Apache Cassandra, RavenDB and Couchbase. The author has compared above NoSQL implementations with the Microsoft's SQL Server Express using an experimental setup to test the time taken by each of these implementations in performing operations like Instantiate, Read, Write, Delete. Also, retrieval of the data is considered. For instantiate operation, the RavenDB NoSQL implementation ranks fastest while SQL Express ranks slowest. For reading operation, the Couchbase and MongoDB perform the best while RavenDB performs the worst. SQL Express ranks on the third positions in this comparison. For a write operation, Couchbase and MongoDB again bag the top 2 positions, while RavenDB and CouchDB finish at second to last and last respectively. SQL Express makes it to the fifth position on this run. For a delete operation, the results are quite similar to that of the results of the read operation. Couchbase, MongoDB and SQL Express prove themselves the frontrunners while RavenDB and CouchDB take the last positions. The maximum number of records that are read, written and deleted as well as fetched are 100000 records. In the final test i.e fetch all records operation, SQL Express ranked the fastest while other NoSQL implementation faired well. CouchDB still was the last rank holder while Couchbase was excluded in this test as there are no APIs available for fetching all the keys. The author concludes by stating that NoSQL is peculiar because this technology is optimized for its key - value implementations. While SQL is not yet optimized. The author asserts that there are a wide variety of NoSQL implementations available and existing for applications. One can easily observe how different NoSQL implementations perform best for respective operations and worst for respective operations. The overall performance of MongoDB, a document-oriented implementation of NoSQL performed the best if the inability of Couchbase to fetch all records is considered, otherwise Couchbase is the front rank holder. The author also argues that these database comparisons and standings may not hold true when these are tested for more complex operations, although these implementations undergo various changes and revisions that may lead to degradation of existing features and introduction of new features.

III. CURRENT SCENARIO

There are multiple applications and systems running on traditional relational databases in todays time, where new database technologies like NoSQL exist which are much more efficient than the traditional RDBMS in different scenarios. In this section, we will be discussing certain applications using traditional RDBMS technology:

Greenbills - Bill Management System:

Greenbills is a billing management system which allows vendors as well as sellers to manage their bills and store them locally on their mobile devices or they can store them at a centralised repository.

The system provides two ends one for seller and other one for vendor, first let's see the vendors end; Whenever a purchase is made and a bill is generated the Vendor will login to the web portal and fill in the details of the purchase along with the bill number and additional data related to bill, when the form will be submitted a unique QR code will be generated.Now at the user's end, the user will login to his/her respective account and will scan the QR code generated by the vendor after scanning the bill will be saved into the user's local storage and user will be also provided with the option to upload the bill to the central repository so that later on no additional problems will come while migrating to an another device.Users will be able to see all the bills related to their purchase and will be able to manage them accordingly, the system will also provide the users data regarding their type of purchases and will be able to sort data according to the type of purchase, date of purchase and vendor's name. A counter will be also provided in the system which will calculate the total amount spent by the user till now.

Compiler on Cloud:

The scope of this project consider the user profile customization and file management functions. This also includes the provision of an online programming suite and private cloud infrastructure based compilation and execution of user files. Also for the intent of load balancing, the scope covers the distributed architecture inclusion. The project solves the basic lacking flexibility problem in diverse range of compilers available in the community market for developers. For eg: one can consider the DEV C++ compiler which restricts the use of certain libraries and function namely:

```
#include<conio.h>

clrscr()

getch()
```

While if one consider the C language compiler TURBO C, all above libraries and functions are prominently supported. This shows there are certain constraints on use of grammar and semantics embedded in the execution and compilation scripts of these compilers. This one drawback is addressed by this project by providing a universal and uniform compiling and runtime execution solution. The Operating System requirements is Windows on both client and server side and SQL Server is the most favored option for Databases needs. The Microsoft IIS (Internet Information Services) web server is found to be an optimum choice for Server commitments.

IV. IMPROVEMENTS TO THE CURRENT SYSTEM/ PROOF OF CONCEPT

Greenbills

The technologies used by the developers to build the system are:
Java ME for android application development
PHP and MySQL for backend and server storage

For mobile application Java is used which is the best choice as the developers were targeting only the android devices but if they would have opted for cross-platform development then hybrid app development environments like React Native or Ionic would have been a much better choice.
As the developers were trying to handle a huge repository of data PHP as a server side language is not a good choice because everytime the PHP code runs it gets converted into bytecode and recompiled again which makes it slow.

The database technology developers are using in the System is MySQL which I think is not appropriate at all because of the following reasons.

As the main motto of the system is to store bills and each bill has a different kind of format so it is very complex to store the bill data in a predefined schema because each time the when the bill needs to be stored it has to be converted into a proper format which the schema supports which in turn removes the flexibility of the system.

As there are no update and delete operations getting performed on the database using a storage technology like Column based NoSQL databases would be more efficient as they provide eventual consistency and provide partial retrieval of data which is much more efficient for analysis of data.

In the non-functional requirements the developers are promising a secure environment, but in the first stage itself the data can be breached as the vendor will be providing the user with a QR code which will let user access the content of the purchase he/she has made.

The complete project can be accomplished using the blockchain, as blockchain supports NoSQL column database technology.

Compilers on Cloud

The aforementioned system extensively involves file creating, saving, updating, deleting i.e more of file management functions. For storage purposes the above system employs the SQL server which is a relational database management system (RDBMS) developed by Microsoft. SQL server supports single machine to large Internet facing workload applications with primary query operations done in T - SQL (Transact SQL) which is standardized for set operations in Sybase and Microsoft. Also ANSI SQL is considered for compliance purposes. In SQL , more specifically in RDBMS, the files are stored in the form of BLOB (Binary Large OBjects). This process includes first conversion of files into binary formats and then storing them as BLOBs in database. At the time of retrieval, these BLOBs are first to be reverse converted to finally gain access to the original file. Also there are certain problems associated with the use of BLOBs mainly :

Large BLOBs may slow down your server which does not only host database.
Storing files as BLOBs are hard to maintain especially if the general operation involves periodic updates on original files. Backups of such databases are very difficult and critical in case of sensitive file data.
Migration of data (eg: exporting your database) turns out to be a task of hassle with large BLOBs.

Since it is evident from above summary that the system in consideration involves heavy amount of file storage and updation, using BLOBs is not a proper option. One of the alternatives to storing files as BLOB is storing files in filesystem (project directory structure). This process involves storing files in particular directory of the filesystem and storing the path address of the file in the database.
Although this overcomes few discrepancies in case of BLOBs, but this too may vary down in case of rapid data retrieval and large amount of large files containing huge amount of data.

In this case , the files containing programming codes. Based on above analysis we can affirm the NoSQL alternative for this project/system/web application can be MongoDB. MongoDB is a cross platform document oriented NoSQL database. It is one of the most efficient NoSQL databases existing till date. Applications of MongoDB include file handling, text files, xml files, etc.
Since it is a document oriented database, it stores data in the form of documents. A GitHub based researcher Andre Caetano presented a node.js built application designed to compare SELECT and INSERT performances of MySQL and MongoDB depending upon the count of rows affected.

A snapshot of 10, 100, 1000 and 10000 affected rows is given below:

Save into DB	1 kB(txt)	20 kB(docx)	1 397 kB(docx)	259 227 kB(mkv)
SQL Server	27,9 ms	6,1 ms	26,8 ms	2694,8 ms
MongoDB	7,3 ms	4,9 ms	23,6 ms	4811,1 ms
Load from DB	1 kB(txt)	20 kB(docx)	1 397 kB(docx)	259 227 kB(mkv)
SQL Server	39,4 ms	3,5 ms	17,4 ms	2917,3 ms
MongoDB	39,3 ms	3,1 ms	14,8 ms	2182,8 ms

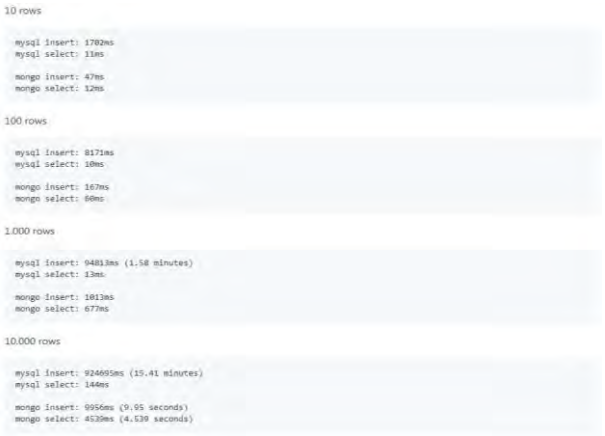


Figure 1 : Snapshot of Comparison

If we consider the maximum number of rows i.e. 10000, it is clearly evident that there is a significant execution time difference between MySQL and MongoDB insert operation (5 seconds is significant if the computation is considered). Even if we consider the minimum number of row i.e 10, one can precisely note the large timing difference (in ms) of MySQL and MongoDB insert operations.

In contrast with the idea of using the SQL for the above system, we can consider the execution times of file insertion in file retrieval operations in MySQL and MongoDB. Barna Burom, a Budapest based TypeScript and Angular developer presented a concise and practical comparison of file saving and loading query operation execution times of SQL Server and MongoDB. The three types of files considered for this test were .txt (text files), .docx (word files) and .mkv (video files). The below table shows the comparison results as provided by the developer himself:

Figure 2 : Mongo - SQL File Operations

From above one can simply conclude that MongoDB is quite an efficient option in case one's application involves file storing and retrieval functions.

V. CONCLUSION

Although from above evidences and research findings, it is implied that mysql is inefficient in multiple operations. NoSQL can be used as at places where MySQL fails in terms of performance, flexibility, data analysis etc. In some cases even NoSQL fails eg: consistency but is a success when it comes to eventual consistency. NoSQL can be used as an alternative to SQL.

VI. REFERENCES

1. Jagdev Bhogal, Imran Choksi, Handling Big Data using NoSQL, IEEE, 2015.
2. Yishan Li, Sathiamoorthy Manoharan, A performance comparison of SQL and NoSQL databases, IEEE, 2013.
3. Rashid Zafar, Eiad Yafi, Megat F. Zuhairi, Hassan Dao, Big Data : The NoSQL and RDBMS review, IEEE, 2016.
4. Surya Narayana Swaminathan, Ramez Elmasri, Quantitative Analysis of Scalable NoSQL Databases, IEEE, 2016.
5. Venkat N Gudivada, Dhana Rao, Vijay V. Raghavan, NoSQL Systems for Big Data Management, IEEE, 2014.
6. Bogdan George Tudorica, Cristian Bucur, A comparison between several NoSQL databases with comments and notes, IEEE, 2011.
7. https://github.com/webcaetano/mongo-mysql
8. https://codingsans.com/blog/nosql-vs-relational-dat abase
9. https://www.microsoft.com/en-us/sql-server/sql-ser ver-2017
10. Ansari Mohd. Arshad, Arshiya Khan, Shaikh Sana,Er. Zainab Mirza, Compilers on Cloud, IJERT, 2013.
11. Er. Zainab Mirza, Greenbills, 2015.

Column-Oriented NoSQL Database: An Optimized Alternative for Analytical Environment

Er. Zainab Mirza (*HOD*), Pradeepkumar Vishwakarma, Arbaaz Tamboli, Wajeed Sharief
Information Technology
M.H.Saboo Siddik College Of Engineering.
Mumbai, India
zainab.mirza@mhssce.ac.in, pradeep090797@gmail.com, arbaazt21@gmail.com, wajeed.sharief54@gmail.com

***Abstract*— Data storage is the slowest component of parallel processing. Even a solid-state drive can't compete with today's memory and processor speeds. With column-oriented database, we can both store as well as analyse big data sets. The column-oriented NoSQL model provides for big data the most suitable model to the data warehouse and the structure of multidimensional data as the OLAP cube and allows it to be deployed in the cloud and a high scalability whilst delivering high performance. This provides far more logical approach to acquiring large sums of data. Column-store method has advantages in reading the records faster as compared to classical row-store in which every row are stored one after another in the disk. This paper mainly focuses on the motivation for column-oriented databases, the optimized improvements to the existing systems. This is an high performance alternate database technology over row-oriented database.**

***Keywords*— Column-oriented database, row-oriented database, NoSQL, RDBMS.**

I. INTRODUCTION

Classic relational database systems support robust data modes, a query language in SQL, referential integrity, and strong consistency. They are designed to handle a few thousand requests over the course of a day and utilize a single database instance on a server. Carefully planned partitioning schemes, replication, and clustering can enable systems to scale out to handle even more load. However, maintaining consistency and availability becomes increasingly difficult as the distributed system continues to grow. Strong consistency across server nodes requires multiphase commit strategies and synchronous replication, which increase the latency of requests and contributes to the potential for failures. While the classical theories of relational database design and normalization are well understood, in practice the rigorous constraints placed on these systems do not always effectively meet the storage requirements of Internet-scale applications. NoSQL on the other hand are a whole new breed of databases that define "logical" aggregate levels to explain the data. Some treat the data as having hierarchical relationship (aggregate being a "node"), while the other treat the data as documents (which is the aggregate level). They do not dictate the physical storage strategy (some may do, but abstracted away from the end user). Also, the whole NoSQL movement is more to do with unstructured data, or rather data sets whose schema cannot be predefined, or in unknown beforehand, and therefore cannot conform to the strict relational model. A column-oriented database is different from traditional row-oriented databases because of how they store data. By storing a whole column together instead of a row, you can minimize disk access when selecting a few columns from a row containing many columns [1]. In row-oriented databases there's no difference if you select just one or all fields from a row. Column-oriented databases are dealing with the way data is physically stored on disk such that each column is stored in its own separate space/file. This allows for two important things:

i. You achieve better compression ratio to the order of 10:1 because you have single data type to deal with.

ii. You achieve better data read performance because you avoid whole row scans and can just pick and choose the columns specified in your SELECT query.

By using numbering system in columnar databases, algorithms can be used to simplify the retrieval of data. Each Columnar database system has highly sophisticated methods of gaining further performance measures.

II. LITERATURE REVIEW

Vandana Bhagat, Dr. Arpita Gopal in their IEEE publication "Comparative Study of Row and Column Oriented Database". In this paper author wants to say that in row-oriented databases, records are easy to read and write which makes it efficient to use in Online Transaction Processing (OLTP). In OLTP, accessing the resources and maintaining data integrity is the challenging thing. And also the users are not satisfied with the hierarchical, network and

relational data models provided by the row-oriented

database. Nowadays size of data are increasing continuously and also RDBMS makes inefficient use of disk to maintain and store large amount data due to which query processing becomes slow. Thus column-oriented database are more efficient for handling records with the minimum use of disk. Almost all the organization maintains record of data for its further use due to which size of the warehouse is continuously increasing. So all this is possible by column-oriented database which makes use of unstructured data storage. The organizations are now moving towards the unstructured data with the help of column-oriented database which is not possible by row oriented database. In this type of data stores, data are stored and retrieve in columns and hence it can only able to read only the relevant data if required. Thus the author concluded that for accessing large amount of data we can compress data using column-oriented database to take the long term decision. For OLAP system row oriented database is not suitable as it takes more time and storage to process data [2].

Daniel J. Abadi, Daniel S. Myers, David J. DeWitt, Samuel R. Madden in their publication "Materialization Strategies in a Column-Oriented DBMS" In this paper, the Authors have focused on the various technique to store data column-wise that has been found very efficient and valuable to use as compared to row-wise storage for 'read-only' databases. ODBC, JDBC, etc are the relational database interface that can benefit from column-wise data storage. There are two methods to store data column-wise: early materialization and late materialization. However, early materialization is not a best strategy for this. Suppose a query consists of three selection operators σ1, σ2, and σ3 over three columns, R.a, R.b, and R.c (all sorted in the same order and stored in separate files), where σ1 is the most selective predicate and σ3 is the least selective. Where late materialization can be effective because it doesn't form tuples until some part is processed as per the plan. C-Store Query Executor has been chosen as the column-oriented database for experimenting and extending it as they were very much familiar with the source code. The column consists of separate file of 64KB blocks which can be optionally used for encoding a variety of compression techniques for storage. A bit-vector encoded file representing a column of size n with k distinct values consists of k bit-strings of length n, one per unique value, stored sequentially. Bit-string k has a 1 in the ith position if the column it represents has the value k in the ith position. C-Store data source will support two operations: reading positions from a column and reading (position, value) pairs from a column. In this paper, there was a huge scope according to our research that CODB (Column-Oriented Database) performance considerations were made with the help of computing algorithms and many performance factors were referred before testing in a JVM environment resulting in efficient execution. Finally, a first approach to a column-oriented database mathematical performance model was created and discussed, on the basis of the results [3].

Daniel Abadi, Peter Boncz and Samuel Madden in their IEEE Publication "The Design and Implementation of Modern Column-Oriented Database Systems"(2013) did

survey research on column-oriented database systems, or column-stores, where each attribute of a table is stored in a separate file or region on storage. Such databases have seen a resurgence in recent years with a rise in interest in analytic queries that perform scans and aggregates over large portions of a few columns of a table. The main advantage of a column-store is that it can access just the columns needed to answer such queries. MonetDB, VectorWise, C-store are the systems that have formed the basis for several well known commercial column-store implementations. We describe their similarities and differences and discuss their specific architectural features for compression, late materialization, join processing, vectorization and adaptive indexing (database cracking).

Some of the features and concepts described above can be applied with some variations to row-store systems as well. In fact, most of these design features have been inspired by earlier research in row-store systems and over the years several notable efforts both in academic and industry tried to achieve similar effects for individual features with add-on designs in traditional row-stores, i.e., designs that would not disturb the fundamental rowstore architecture significantly.

These include compression, vectorization, late materialization, and efficient column-oriented join methods. These ideas have found their way into a several commercial products—both direct descendants of the academic projects (e.g., VectorWise and Vertica) as well as a number of other products (e.g., Aster Data, Greenplum, Infobright, Paraccel, etc). Even row-oriented stalwart Oracle has implemented some column-oriented techniques in their database appliance, Exadata (in particular, they have implemented PAX page layouts and column-oriented compression). These products are reputed to provide one to two orders of magnitude better performance than older-generation row-oriented systems on typical data warehousing and analytics workloads, and have been quite successful commercially (with VectorWise, Vertica, Greenplum, and Aster Data all having been recently acquired). Despite the academic and commercial success of column-oriented systems, there are still several interesting directions for future research. In particular, there is a substantial opportunity for hybrid systems that are partially column-oriented. For example, systems that store groups of frequently accessed columns together could provide better performance than a pure column-store. Additionally, systems that adaptively choose between column and row-oriented layouts for tables, depending on access patterns over time, will likely become important, as requiring users to determine which type of layout for their data is not ideal [4].

III. CURRENT SYSTEMS

Nowadays, almost all systems are working based on row-oriented databases or we can say that relational databases due to which it makes the system inefficient

use of the database. So, in this section we are going to discuss about the current scenario certain systems which are using row-oriented databases or relational databases:

A. Cloud Advisory Services

Cloud Advisory Services is a cloud-based web application in which if the user wants to move their system on the cloud then he can do so by computing total cost of their hardware specifications of their preferred system.

In Front-End of this system, there will be user who will first register and then login through their credentials. After which this cloud-based web application will allow users to specify their hardware requirements based on certain parameters like Compute, Storage, SQL, and APP Engine. Based on users different input for these parameters, the application will generate a total cost for the specified system. In Back-End, a database for the number of users logging in or registering is maintained. Seperate database will also be maintained once the user has computed total cost of their hardware specification.

Whereas in the Back-End, a master page will be maintained by the admin. This master page is responsible for changing dynamic cost of the above mentioned parameters. This changes in the cost of above mentioned parameters will be known to the admin by consulting and analyzing various cloud vendors and their activities about the changing costs. Once the change in the cost is known by admin, then these changes will also be updated in the Front-End which can be available to user also [5].

B. Compilers on Cloud

There are many online program-compiling websites which are been used to compile, run or edit for User's to test their codes on it where they can write code on the Terminal or they can select files from the device where they can select the File Location to import the Program File. Cloud helps this by building a PaaS(Platform-as-a-Service) Platform where User can build and run different programming language codes with providing necessary libraries and a prominent coding UI.

Cloud Computing technology has a huge scope which has help build a platform where User can be connected remotely using a WAN or Internet network to scale a large number of Users where they can write their codes or import files from Cloud Storage to overcome the need to import files from Computer Devices as they can do it using Google Drive and other services. The requirements to use Cloud Compilers are very minimum and scope for future coders is enormous as low storage space is required and libraries are already provided by them as no need to download it from external sources where some of them who are providing unlicensed libraries can affect the compilation and debugging of codes [6].

MinGW 5.3 C/C++ , Microsoft Visual C++ 2017 Family, Microsoft Visual C++ 2015 Professional, Microsoft Visual C++ 2013 Professional are some of the Compilers that are preferred by the MathWorks Cloud and mostly are used as a third-party Compiler. MySQL is a relational database and allows multiple users to number of databases as it runs on server side. The platform Cloud Compilers can support is Windows, Mac and Linux with C/ C#, C++, Java IDE, Python are the leading coding languages they offer [7].

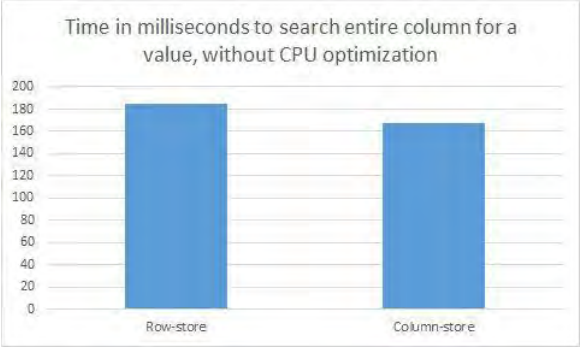
IMPROVEMENTS TO CURRENT SYSTEMS

A. Cloud Advisory Services

The technologies that are used for building the above system are: Visual Studio/C# .NET with ASP.NET for web application and MS SQL SERVER 2005 EXPRESS for Back-End and storage. Visual Studio/C# .NET with ASP.NET have been used for for Front-End which makes system user friendly. But at the same time making use of MS SQL server for Back-End is good if there is small amount of data need to be handled because MS SQL is row-oriented database. But if the large volume of data need to be handled then using of MS SQL Server is not a good choice because as the size and complexity of data base increases, then row oriented database will not be able to provide large storage which can be done using column-oriented database.

Since we have to store users hardware specifications, their cost and account information there is a need a of large database.Also this system requires large number of calculations like SUM, MIN, MAX,AVG, etc. so it is more feasible using column-oriented database ae it stores data in columnar manner and not by row as classic DBMS do. Also this system needs very fast query processing because it might possible that many user search for same hardware specification so in that case getting results quickly will be challenging thing for the system. So for generating results quickly we can use column-oriented database instead of row-oriented database which is normally slow in MS SQL server. Also memory usage is very less in the case of column-oriented database as it uses compression techniques to lower the disk space.

So, in some case this project can be implemented using MS SQL server such as dynamic update of the cost of hardware and if the users are very less and also memory requirements is also very less. But if the users are continuously increasing, memory requirement is also very large and fast query processing is also needed then we can implement this project using column-oriented database.



From the above table we can say that time required to search any value in row-oriented table is more as compared to column oriented because in row-oriented database, values of all the attributes are stored row wise whereas in

column-oriented database, the values of all the attributes are stored in column wise.

TABLE I. MAJOR DIFFERENCES BETWEEN ROW STORE AND COLUMN STORE [9]

Property	***Row Store***	***Column Store***	***Reason***
Memory Usage	Higher[a]	Lower	Compression
Transactions	Faster	Slower	Modifications require updates to multiple columnar tables
Analytics	Slow even if indexed	Faster	Smaller dataset to scan, inherent indexing

From the above table we can say that memory usage is less in column-oriented because of compression technique used by column-oriented database. OLAP is very fast in case of column-oriented store whereas OLTP is fast in case of row-oriented store.

B. Compilers on Cloud

The above-mentioned system requires creation, updation,creation and execution of codes which will be managed by the Cloud. SQL Server is required for that which is developed by Microsoft.

Column-Oriented Database(CODB) can help improve data storing in Cloud. It provides columnar data storage with increased durability and prevents data loss from Cloud. Data Manipulation and Sorting can be done by the Database Users which will result in no-relying on the application. Self-indexing provides effective data querying, processing and analysis. Scalability is increased at a greater extent as the data are stored column-wise resulting is more data space for the Users. It allows data to compress to optimize storage [10]. It allows data to load extremely fast and can store large amount of non-volatile information for long duration. Some of the examples are MonetDB, CrateDB, Apache HBase, Apache Kudu which can help make Compiling efficient on Cloud as they provide faster results as compared to other Databases. CODB can be used for Data Mining and Business Intelligence techniques [11].

The below image shows how storing data in CODB is done compared to Row-Oriented Database(RODB) :

.



Fig. 2. Data Storing in RODB vs CODB [12].

Parallax Inc is a Cloud Compiler Projects which is used to compile Parallax Propeller Spin and C projects. It asks User to install SimpleIDE for Windows and Simple IDE with Propeller GCC Installation Instructions for Linux to configure and compile the Codes.

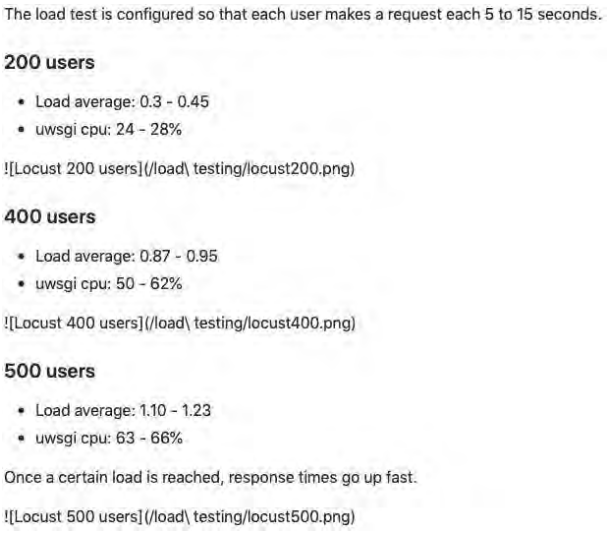


Fig. 3. Load depending on Scalability [13]

The above figure shows the scalability is increasing which results in more load on the database depending on the Users which can help us identifying average time-stamp being taken to compile the codes and maintain the load of the System.

TABLE II. DIFFERENCES BETWEEN SQL AND NOSQL [14]

Property	***SQL***	***NoSQL***
Database Schema	Schema.	Schema-less.
Mapping of data	Relational.	Few or No Relations.
Distribution of data	Done Across Multiple tables.	Merge or Nested in few Collections.
Scaling	Horizontal scaling is not possible whereas Vertical scaling is possible.	Both Horizontal and Vertical scaling is possible.
Read/ Write	Limitations for read and write of queries/ sec.	No Limitations for read and write querying.

The above table shows that NoSQL is better suited for Cloud Computing as it is Schema-less and few relations are needed for storing and querying of data. Tables are merged hence no duplication of data is required. NoSQL gives advantage of both Horizontal and Vertical Scaling and it can process a great amount of read and write queries.

IV. CONCLUSION

Column-oriented databases have a role to play in data that is much larger than available RAM, or where naive aggregates or stream processing is paramount. They're often found for specific purposes in the financial industry, and

there are some absolutely fantastic possible implementations. It is less conclusively a fit if your write load is at all significant, or if the usage of the data is general if not arbitrary. Row-stores and column-stores have very different performance characteristics, and they tend to struggle where the alternative excels [15].

Column-oriented databases need to be even more purposefully used than row-oriented if you hope to have efficient operations. Elastic scaling, handling Big Data, fully managed by DBA, less expensive, flexible data model and many more are the main advantages of it. Columnar databases are designed for large volumes of data, read and write performance, and high availability. Even row-oriented stalwart Oracle has implemented some column-oriented techniques in their database appliance, Exadata. Column datastores also minimizes expensive input and output request called IO to our data storage systems. Billions of rows of data can be queried in less than three seconds. Its application are in content management systems, blogging platforms, systems that maintain counters and services that have expiring usage. Bigtable, Cassandra, HBase, Vertica, Accumulo, Hypertable, etc are some examples of column-store datastores.

Also, Columns are stored sequentially. Sequential access if much faster than random access. Column database can also be compressed more efficiently with similar values stored together. This saves on capacity and values can be read faster. Compression, aggregation queries, scalability and fast to load query are its key benefits [16].

Despite the academic and commercial success of column-oriented systems, there are still several interesting directions for future research. In particular, there is a substantial opportunity for hybrid systems that are partially column-oriented. As workloads get more complex, even hybrid systems will gain importance. The shortcomings of relational database and the strengths of Column-oriented NoSQL databases that motivate migration from relational as well as row-oriented database were discussed. We expect that column-oriented ideas will start to find their way into various systems and even the main purpose of this paper was to provide the motivation for the same.

REFERENCES

1. A Review of Column-oriented Data stores by Zach Pratt at attackofzach.com
2. Vandana Bhagat, Arpita Gopal, *"Comparative Study of Row and Column Oriented Database",* IEEE, 2012, Available at: https://ieeexplore.ieee.org/document/6495251
3. Daniel Abadi, Peter Boncz, Samuel Madden, "*The Design and Implementation of Modern Column Oriented Database Systems",* IEEE, 2013.
4. Daniel J. Abadi, Daniel S. Myers, David J. DeWitt, Samuel R. Madden, *"Materialization Strategies in a Column-Oriented DBMS"*, IEEE, 2007.
5. Hamza Goawala, Hayat Jahagirdar, Ziya Khatib, Er. Zainab Mirza, "Cloud Advisory Services", 2016.
6. Suryawanshi Harshal, Rokade Chakrapani, Ambhore Ajay, Rathod Sharad, *"Compiler as Service over Cloud"*, International Journal of Computer Applications,Volume 70– No.1, May 2013.
7. Uday Chauhan, Mayuri Patole, Yogita Mokashi, Saurabh Padalikar *"Online Cloud Based Compilers System"*, February 2016.
8. DBMS Musings, *"Apache Arrow vs. Parquet and ORC"*, 2017, Available at: http://dbmsmusings.blogspot.com/2017/10/apache-arrow-vs-parquet-a nd-orc-do-we.html
9. TeachMeHANA, *"Difference between Row Store and Column Store"*, May, Available at
10. http://teachmehana.com/row-store-vs-column-store/
11. Ansari Mohd. Arshad, Arshiya Khan, Shaikh Sana, Er. Zainab Mirza, "*Compilers on Cloud",* IJERT, 2013.

12. Pat Research, "*Top 9 Column-Oriented Databases"*, 2019, Available at : https://www.predictiveanalyticstoday.com/top-wide-columnar-store-d atabases/
13. Jon Tong, *"Demystifying Columnar Databases",* April 2012, Available at : https://www.youtube.com/watch?v=wYl_YxqTof4
14. Parallalxinc, *"Cloud-Compiler"*, GitHub, 2015, Available at : https://github.com/parallaxinc/Cloud-Compiler
15. Academind, *"SQL vs NoSQL or MySQL vs MongoDB"*, 2018, Available at: https://www.youtube.com/watch?v=ZS_kXvOeQ5Y&t=1145s
16. Jyoti Kharade, Anil Rama Kale, Dhanaji S. Kharade, "*NOSQL Database: Opportunities and Applications"*, International Journal of Computer Sciences and Engineering, Vol.-6, Issue-5, May2018.
17. Jared Hilam,*"What is a Columnar Database?",* Available at : https://www.youtube.com/watch?v=8KGVFB3kVHQ

Machine Learning: Review of Algorithms

Roshakumar R. Maurya, DhiajYadav, Anand Khandare
Department of Computer Engineering
Thakur College of Engineering and Technology
Mumbai 400101
maurya.roshan1996@gmail.com, erdhiraj1992@gmail.com, anand.khandare @thakureducation.edu

***Abstract*-- Machine learning is a field of computer science that gives computers ability to learn without being explicitly programmed. Machine learning is used in a variety of computational tasks where designing and programming algorithms with good performance is not easy. Using the algorithm we can increase the performance of the software.**
One of the advantages of machine learning is to train computers to utilize data to solve a specified problem. These algorithms are used for various purposes like data mining, image processing, predictive analytics. There are disadvantages of these algorithms on some data.
This paper studied different machine learning algorithms, there application and analyze their performance on different datasets. Experiments are done on these algorithms and summarize their performance with respect to some parameters.

***Keywords*-- Machine Learning, Algorithms, Time Complexity.**

I. INTRODUCTION

Machine learning is used to teach machines how to handle the data more efficiently. Sometimes after viewing the data, we cannot interpret the pattern or extract information from the data. In that case, we apply machine learning. With the abundance of datasets available, the demand for machine learning is in rise. Many industries from medicine to military apply machine learning to extract relevant information. The purpose of machine learning is to learn from the data. Many studies have been done on how to make machines learn by themselves. Many mathematicians and programmers apply several approaches to find the solution of this problem.There are three main types of Machine Learning Algorithm.

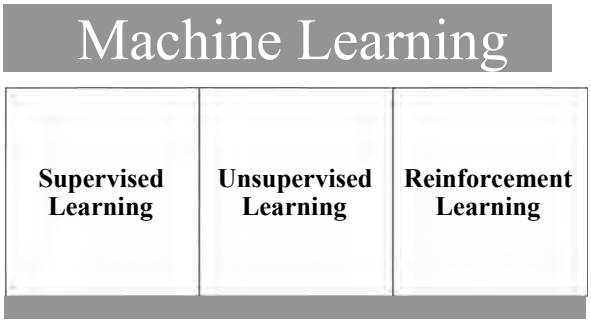


Figure.1. Types of Learning

II. BACKGROUND AND MOTIVATION

A. Background

Machine learning grew out of the quest for artificial intelligence. Already in the early days of AI as an academic discipline, some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what were then termed "neural networks"; these were mostly perceptions and other models that were later found to be reinventions of the generalized linear models of statistics. Probabilistic reasoning was also employed, especially in automated medical diagnosis. However, an increasing emphasis on the logical, knowledge-based approach caused a rift between AI and machine learning.
Probabilistic systems were plagued by theoretical and practical problems of data acquisition and representation. By 1980, expert systems had come to dominate AI, and statistics was out of favor. Work on symbolic/knowledge- based learning did continue within AI, leading to inductive logic programming, but the more statistical line of research was now outside the field of AI proper, in pattern recognition and information retrieval. Neural networks research had been abandoned by AI and computer science around the same time. This line, too, was continued outside the AI/CS field, as "connectionism", by researchers from other disciplines including Hopfield, Rumelhart and Hinton. Their main success came in the mid-1980s with the reinvention of back propagation.
Now, in a last few year every multinational company working on machine learning to create the self leaning machine to increase the performance of the system. They working in different domain such as data mining, chart board, robotics, medical ets.

B. Motivation

Machine Learning is no longer just a niche subfield of computer science but technology giants have been using it for years – Machine learning algorithms power Walmart product recommendations, surge pricing at Uber, fraud detection at top financial institutions, content that Twitter, LinkedIn, Facebook and Instagram display on social media feeds or Google Maps. Machine learning products are being used daily, perhaps without realizing it. The future of machine learning is already here, it's just that machine learning career is exploding now because of smart algorithms being used everywhere from email to mobile apps to marketing campaigns. The studied of Machine
Learning Algorithm we know how to increase the performance using different types of machine learning algorithm.

III SUPERVISED LEARNING

based on their probability of happening. These trees are also known as Bayesian Network.

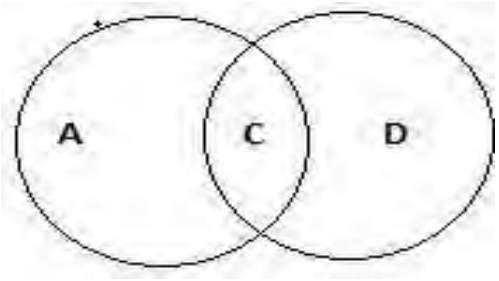
The supervised machine learning algorithms are thosealgorithms which needs external assistance. The inputdataset is divided into train and test dataset. The traindataset has output variable which needs to be predicted orclassified. All algorithms learn some kind of patterns fromthe training dataset and apply them to the test dataset for prediction or classification. The workflow of supervisedmachine learning algorithms is given in

1) Decision Tree*:* Decision trees are those type of treeswhich groups attributes by sorting them based on theirvalues. Decision tree is used mainly for classificationpurpose. Each tree consists of nodes and branches. Eachnode represents attributes in a group that is to be classifiedand each branch represents a value that the node

Formula

$$P(C|F)=\frac{P(F|C)P(C)}{P(F)}=\frac{0.5\times 1/50000}{1/20}=0.0002$$

WHERE



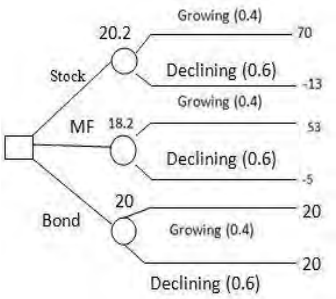


Figure 3. Bayes Theorem

P (A/B) P(A) P (B)

can take. Problem

Alternative	Growing	Declining
Stocks	70	-13
Mutual Funds	53	-5
Bonds	20	20
Probility	0.4	0.6

➔ Decision Node

◯ ➔ Change Node

A AND b ARE Events P (A) is the Priori of A P (A/B) is a Posterior Probability of B. Time

Complexity O (Np)

Problem

A doctor knows that Cold causes fever 50% of the time, Prior probability of any patient having cold is 1/50,000,Prior probability of any patient having fever is 1/20.

Solution

If a patient has fever, what's the probability he/she has cold?

3) Support Vector Machine*:* Another most widely used State-of-the-art machine learning technique is Support Vector Machine (SVM). It is mainly used for classification. SVM works on the principle of margin calculation. It basically, draw margins between the classes. The margins are drawn in such a fashion that the distance between themargin and the classes is maximum and hence, minimizing the classification error.

Stock MF

0.4(70) +0.6(-13) = 20.2

0.4 (53) + 0.6(-5) = 18.2

Now we Investe in Stock

2) Naïve Bayes:Naïve Bayes mainly targets the text classification industry. It is mainly used for clustering and classification purpose . The underlying architecture of Naïve Bayes depends on the conditional probability. It creates trees

Figure 4. Working of Support Vector Machine

INPUT: S, λ, T, k
INITIALIZE: Choose $\mathbf{w}_1$ s.t. $\|\mathbf{w}_1\| \leq 1/\sqrt{\lambda}$
FOR $t = 1, 2, \ldots, T$
Choose $A_t \subseteq S$, where $|A_t| = k$
Set $A_t^+ = \{(\mathbf{x}, y) \in A_t : y \langle \mathbf{w}_t, \mathbf{x} \rangle < 1\}$
Set $\eta_t = \frac{1}{\lambda t}$
Set $\mathbf{w}_{t+\frac{1}{2}} = (1 - \eta_t \lambda)\mathbf{w}_t + \frac{\eta_t}{k} \sum_{(\mathbf{x},y) \in A_t^+} y\mathbf{x}$
Set $\mathbf{w}_{t+1} = \min \left\{ 1, \frac{1/\sqrt{\lambda}}{\|\mathbf{w}_{t+\frac{1}{2}}\|} \right\} \mathbf{w}_{t+\frac{1}{2}}$
OUTPUT: $\mathbf{w}_{T+1}$

Time Complexity O (nSVd).

IV UNSUPERVISED LEARNING

The unsupervised learning algorithms learns few features from the data. When new data is introduced, it uses the previously learned features to recognize the class of the data. It is mainly used for clustering and feature reduction.

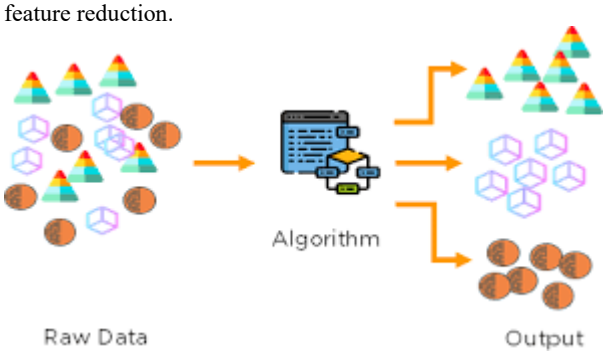


Figure 4. Working of Unsupervised Learning

1) K-Means Clustering**:** Clustering or grouping is a typeof unsupervised learning technique that when initiates,creates groups automatically. The items which possesssimilar characteristics are put in the same cluster. Thisalgorithm is called k-means because it creates k distinctclusters. The mean of the values in a particular cluster is thecentre of that cluster.

Algorithmic steps for k-means clustering

Let X = {x1,x2,x3,……..,xn} be the set of data points and
V = {v1,v2,…….,vc} be the set of centers.

1. Randomly select ‘c’ cluster centers.
2. Calculate the distance between each data point and cluster centers.
3. Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.

$$v = (1/c) \sum_{i=1}^{c_i} x_i$$

Recalculate the new cluster center using:
Recalculate the distance between each data point and new obtained cluster centers.
If no data point was reassigned then stop, otherwise repeat from step 3).

Time Complexity O(min(p^3,n^3))

V. REINFORCEMENT ALGORITHM

Reinforcement learning is a type of learning which makes decisions based on which actions to take such

Time Complexity O (n^2)

2) Hierarchical clustering**:**is an algorithm that groups similar objects into groups called clusters. The endpoint is a set of clusters, where each cluster is distinct from each other cluster, and the objects within each cluster are broadly similar to each other.

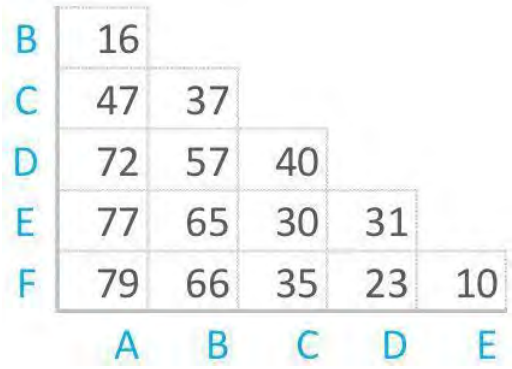


Figure 5. Working of hierarchical Clustering

Time Complexity O(n^3)

3) Principal Component Analysis**:**In Principal Component Analysis or PCA, the dimension of the data is reduced to make the computations faster and easier. To understand how PCA works, let’s take an example of 2D data. When the data is being plot in a graph, it will take up two axes. PCA is applied on the data, the data then will be 1D.

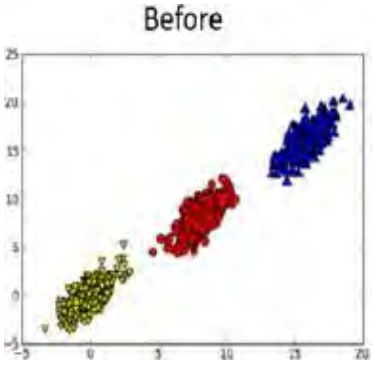


Figure 6. Visualization of data before and after applying PCA

thatthe outcome is more positive. The learner has no knowledge which actions to take until it's been given a situation. The action which is taken by the learner may affect situations and their actions in the future. Reinforcement learning solely depends on two criteria: trial and error search and delayed outcome.

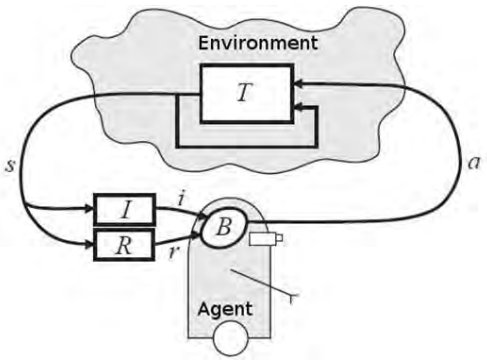


Figure 7. The Reinforcement Learning Model

In the figure, the agent receives an input i, current state s,state transition r and input function I from the environment.Based on these inputs, the agent generates a behavior B andtakes an action a which generates an outcome.

REFERENCES

[1] W. Richert, L. P. Coelho, "Building Machine Learning Systems withPython", Packt Publishing Ltd., ISBN 978-1-78216-140-0

[2] M. Welling, "A First Encounter with Machine Learning"

[3] M. Bowles, "Machine Learning in Python: Essential Techniques forPredictive Analytics", John Wiley & Sons Inc., ISBN: 978-1-118-96174-2

[4] S.B. Kotsiantis, "Supervised Machine Learning: A Review ofClassification Techniques", Informatica 31 (2007) 249-268

[5] L. Rokach, O. Maimon, "Top – Down Induction of Decision TreesClassifiers – A Survey", IEEE Transactions on Systems,

[6] D. Lowd, P. Domingos, "Naïve Bayes Models for Probability Estimation"

[7]https://webdocs.cs.ualberta.ca/~greiner/C651/Homework_2_Fall2008.html

[8] D. Meyer, "Support Vector Machines – The Interface to libsvm inpackage e1071", August 2015

[9] S. S. Shwartz, Y. Singer, N. Srebro, "Pegasos: Primal Estimated sub -Gradient Solver for SVM", Proceedings of the 24th InternationalConference on Machine Learning, Corvallis, OR, 2007

[10] http://www.simplilearn.com/what-is-machine-learning- and-why-itmatters-article

Business audits characterization utilizing sentiment analysis

Kajal Jaiswal
Masters of Computer Engineering (Computer)
Thakur College of Engineering
Mumbai Maharashtra, India
kajalj2222@gmail.com

Jay Pal
Masters of Computer Engineering (Computer)
Thakur College of Engineering
Mumbai Maharashtra, India
jaypal9049@gmail.com

Dr .Sheetal Rathi
Masters of Computer Engineering (Computer)
Thakur College of Engineering line 4:
Mumbai Maharashtra, India
sheetal.rathi@thakureducation.org

Abstract — The examination zone of sentiment analysis, supposition mining, slant mining and feeling extraction has picked up ubiquity in the most recent years. Online surveys are ending up vital criteria in estimating the nature of a business. This paper introduces an assessment investigation way to deal with business surveys grouping utilizing a substantial audits dataset given by Yelp: Yelp Challenge dataset. In this work, we propose a few methodologies for programmed assessment characterization, utilizing two component extraction techniques and four machine learning models. It is outlined a near report on the adequacy of the outfit strategies for audits conclusion order.

Keywords— sentiment analysis; opinion mining; classification; text reviews , voice of customer, assessment investigation

I. INTRODUCTION

Sentiment analysis has turned into an essential research zone for comprehension people's supposition on an issue by examining a lot of data. A huge number of individuals express their musings about different administrations or items utilizing long range interpersonal communication locales, online journals or mainstream audits destinations. The dynamic criticism of the general population is profitable not just for organizations to dissect their customers‟ fulfillment and the observing of contenders, but on the other hand is extremely helpful for buyers who need to inquire about an item or an administration before making a buy.

Cry users‟ surveys express conclusions and estimations about organizations and specialist organizations among a given rating, scaled from 1 to 5, which is utilized as a general measurement audit. Cry test dataset contains data around 61,000 nearby organizations, 1.6 million audits and 366,000 clients in 10 urban communities crosswise over 4 nations on the globe. Howl test dataset is a lot bigger than past discharged Yelp scholastic dataset, which contains 15,585 business and 335,022 user's audits [1].

This paper exhibits the aftereffects of a few machine learning calculations for arranging Yelp surveys utilizing assumption examination and characteristic language handling strategies. The enormous number of client created audits and appraisals for eateries, organizations and specialist co-ops are named either positive or negative as for the star evaluations. We propose a technique to naturally order users‟ notions (positive or negative) utilizing just the business content survey.

This is helpful on the grounds that it enables clients criticism to be communicated without manual intercession. By studding just the rating, it is exceptionally difficult to investigate why the client has evaluated the business as 1 or 5 stars. Be that as it may, the content survey contains a more quantitative incentive for breaking down more than the rating

itself. This paper likewise displays the preprocessing steps required so as to accomplish high exactness in the order assignment.

There is no past research on arranging supposition of content surveys utilizing the most recent audits from Yelp test dataset. Deciding the fundamental supposition of business audit is a mind boggling errand considering a few factors, for example, the implication of a word contingent upon the specific circumstance, language utilized, notion vagueness when utilizing words that don't express a specific assessment or when utilizing mockery. We demonstrate that a slant examination calculation based over machine learning calculations, for example, Naïve Bayes and Linear Support Vector Classification (SVC) has exactness above 90% utilizing 1.6 million business surveys. Moreover, in this paper we present the consequences of our trials and thoughts on the most proficient method to additionally enhance the acquired outcomes.

II. RELATED WORK

Past work in regards to notion examination characterization utilizing machine learning systems in deciding whether the general supposition of an audit is sure or negative utilized motion picture surveys as information. The creators utilize a unigram model and Naïve Bayes, most extreme entropy grouping, and bolster vector machines to play out the supposition arrangement and accomplish 80% precision. They presumed that their outcomes beat the strategy dependent on human labeled highlights [2].

Hu et al. play out the assumption grouping of a record at a sentence level rather than the entire archive and concentrate includes on which feelings have been communicated, recognizing conclusion words by proposing a method that utilizes WordNet lexical database. For each component, the

related sentiment sentence is incorporated into positive or negative classifications and figures a complete check. The highlights are positioned by the recurrence of the appearance in the surveys. The creators give an element based synopsis of surveys of items sold on the web [3].

Past work utilizing estimation investigation and Yelp dataset audits concentrated on foreseeing star rating utilizing the content alone. The creators analyze distinctive machine learning calculations, for example, Naïve Bayes, Perceptron, and Multiclass SVM on an example of 100 000 client audits from Yelp dataset. They use

Bing Liu Opinion Lexicon for highlight determination and some preprocessing procedures, for example, evacuating stop words or stemming (for example lessening the words to their root structure). The best outcome (exactness and review) is acquired utilizing Naïve Bayes and highlight choice with stop words evacuated and stemming [4].

In [5] so as to catch word conclusions the creators utilized a directed learning calculation dependent on similitudes between words which considers the rating of past surveys for catching the portrayal of words vectors. On a dataset of around 20 000 special surveys from Yelp dataset, the exactness detailed is about 70%. Other work on conclusion mining, and specifically in survey mining utilizing Yelp dataset, centers around foreseeing a business‟ rating dependent on its just reviews‟ content for diminishing the inclination of the clients. The creators make a sack of words portrayal of the best incessant words in all content audits or best successive descriptors after Part-of Speech joined with Linear Regression, Support Vector Regression and Decision Tree Regression. Utilizing 35 645 surveys from Yelp Academic dataset, the Root Mean Square Error (RMSE) is around 0.6 [6].

Blair-Goldensohn et al. constructed a framework that can consequently abridge suppositions from a lot of audits for a nearby administration such an eatery or lodging and total the survey conclusion per perspective, (for example, nourishment, administration, stylistic layout, esteem). They actualize a custom manufactured dictionary dependent on WordNet and utilize a classifier at the sentence level [7].

III. SLANTANALYSIS CLASIFICATION

We utilize the Yelp Challenge Dataset [1] which comprises of 1.6 million audits (to around 7 GB of information) from 366 000 clients from 61 000 neighborhood organizations which are put away in JSON group. The dataset contains data about business name, area, classification, clients, surveys, dates, stars. Be that as it may, we have concentrated just on business crude content audits and their star appraisals for feeling order.

So as to utilize regulated learning and train a classifier, we for the most part require a hand marked preparing information, yet considering the extensive scope of organizations and the vast number of audits, it would be hard to physically explain the information to prepare a feeling classifier for surveys.

We accept that the star rating is an exact measure for the notion conclusion of the audit. The normal rating of the Yelp surveys was about 3.7. The star rating of a business audit is a whole number from 1 to 5. We chose to take out elite player appraisals and their comparing survey which are equivalent to 3, and to keep all evaluations over 4 which were considered as "positive" feelings, and furthermore to keep all appraisals beneath 3 which were considered as "negative" estimations. We acquire a dataset containing 1 346 545 surveys.

In the wake of part the separated Yelp test dataset in 80% for preparing and 20% for testing, we use preprocessing methods so as to remove a lot of highlights. This information was utilized to prepare a few classifiers. The characterization tests were directed utilizing 3-overlay cross approval for assessing the precision. Our methodology is to utilize distinctive machine learning classifiers and construct highlight extractors. The machine learning classifiers utilized are Naïve Bayes, Linear Support Vector Classification (SVC), Logistic Regression and Stochastic We utilize the Yelp Challenge Dataset [1] which comprises of 1.6 million audits (to around 7 GB of information) from 366 000 clients from 61 000 neighborhood organizations which are put away in JSON group. The dataset contains data about business name, area, classification, clients, surveys, dates, stars. Be that as it may, we have concentrated just on business crude content audits and their star appraisals for feeling order.

So as to utilize regulated learning and train a classifier, we for the most part require a hand marked preparing information, yet considering the extensive scope of organizations and the vast number of audits, it would be hard to physically explain the information to prepare a feeling classifier for surveys.

We accept that the star rating is an exact measure for the notion conclusion of the audit. The normal rating of the Yelp surveys was about 3.7. The star rating of a business audit is a whole number from 1 to 5. We chose to take out elite player appraisals and their comparing survey which are equivalent to 3, and to keep all evaluations over 4 which were considered as "positive" feelings, and furthermore to keep all appraisals beneath 3 which were considered as "negative" estimations. We acquire a dataset containing 1 346 545 surveys.

In the wake of part the separated Yelp test dataset in 80% for preparing and 20% for testing, we use preprocessing methods so as to remove a lot of highlights. This information was utilized to prepare a few classifiers. The characterization tests were directed utilizing 3-overlay cross approval for assessing the precision. Our methodology is to utilize distinctive machine learning classifiers and construct highlight extractors. The machine learning classifiers utilized are Naïve Bayes, Linear Support Vector Classification (SVC), Logistic Regression and StochasticGradient Descent (SGD) Classifier. For feature extraction we use different pre-processing techniques. Ultimately, we measure the exactness utilizing the testing dataset.

A. Preprocessing

In the preprocessing stage every one of the accentuations and every one of the spaces from the audit content are evacuated. We convert every single capital letter to lowercase so as to lessen excess in highlight determination undertaking.

For highlight extraction we execute two methodologies, one methodology is building a custom word reference from the preparation dataset and the other methodology is performing

lexical investigation of content surveys. We additionally thought about utilizing a current supposition vocabulary, however the rundown of highlights to be removed would have a high predisposition. A few highlights explicit to Yelp dataset would not be mulled over and end up unessential regarding the notion mining. Despite the fact that the highlights set would contain just words scored for extremity (utilizing an incorporated rundown of negative and positive sentiment words) some vital highlights that are not found in the dataset would have been prohibited.

B. Feature extraction

After preprocessing steps, for the first approach in features selection algorithm we loop over the entire training dataset to tokenize each word. We implement the following steps:

Negations are linguistic developments that change the significance of a sentence to the inverse. A methodology for dealing with refutations is to attach an alternate character, for example, "!" to each word among nullification and the accompanying accentuation [8]. We chose to modify only the word immediately before and after the negation term, rather than all the words until the end of the sentence. This progression is done so as to express more data of the sentence setting.

- Expel stop words (regular words in English, however with no opinion data) from the content surveys utilizing Porter corpus of prevent words from Natural Language Toolkit (NLTK)
- Expel words that have under three letters.
- Apply stemming (decreasing a word to its root structure or base structure) utilizing Porter calculation which is executed in NLTK. This progression is improved the situation expelling dreary highlights and it is additionally valuable since it diminishes the span of the vocabulary by around 30 percent.
- Further, in the last advance of the primary methodology for building the element extraction calculation, to frame the element vectors we utilized a pack of words portrayal for content audits of all business utilizing unigrams. We manufacture a custom lexicon for mapping the recurrence of event of the words in the preparation set – every token event is treated as a component and the vector of all the token frequencies speaks to a multivariate example. We expel the words with recurrence lower than one for decreasing the sparsity of the preparation lattice.

For the second methodology of the element choice calculation we likewise apply the preprocessing steps recently characterized. Further, we expel stop words and words with under three letters and apply the stemming venture as depicted previously. Next we implement the following steps:

- We circle over the whole preparing dataset to tokenize each sentence and relegate a POS (grammatical feature tag).
- Because of the way that each word can have numerous faculties, we apply Lesk's word sense disambiguation (WSD) calculation based over WordNet which is actualized in NLTK. The calculation restores an equivalent word set (Synset) that has the most elevated number of covering words between the setting sentence and diverse definitions from every WordNet equivalent word set.
- In the wake of finding the nearest importance with regards to the word utilizing WSD, we utilize a slant conclusion dictionary SentiWordNet [9] for registering the positive and negatives scores of each word in each sentence of the content audit. SentiWordNet relegates to every WordNet Synset three slant scores: energy, antagonism and objectivity (for every Synset the entirety of the three scores is 1, and each score esteem is going from 0 to 1). For example, for "moderate", there are 3 faculties of the action word "moderate", 3 faculties of the modifier "moderate", and 2 faculties of the intensifier "moderate" in WordNet. In SentiWordNet as per its POS for one feeling of the descriptive word "moderate", the negative score is 1, yet for another sense the negative score is 0.
- For each tokenized sentence of the audit we utilize all scores of the words in a sentence and register the all out positive and all out negative score of the content survey. We utilize all out positive and negative scores as indicators.

C. Learning Methods

We adapt and apply four learning models: Multinomial Naïve Bayes, Support vector machines: Linear Support Vector Classification, Logistic regression and Stochastic Gradient Descent Classifier for sentiment analysis classification using Yelp challenge dataset. All calculations are actualized in the scikit-learn machine learning library written in Python [10].

Naive Bayes is a customary methodology for content arrangement.In Multinomial Naïve Bayes calculation, we speak to a business content audit utilizing an element vector removed utilizing the proposed highlight extraction calculations. The length of the element vector is equivalent to the quantity of words in the extricated lexicon. We utilize a Boolean component vector structure portrayal, accepting that the event of each word could really compare to the recurrence. We utilize a twofold weighting capacity and allot 1 if the word is available in the lexicon and 0 generally. Laplace smoothing is utilized for maintaining a strategic distance from over fitting.

We construct the component extraction vector utilizing TF-IDF (term recurrence reverse report recurrence) change technique to reflect how imperative a word is to a business content audit and apply the accompanying classifiers: Linear Support Vector Classification (SVC), Stochastic Gradient Descent Classifier and Logistic relapse classifier. In our "pack of words" approach, each word is weighted utilizing its recurrence in the lexicon made with the proposed highlight extraction calculations portrayed in past area. We accept that the word recurrence is identified with the marker of the assumption. For example if "incredible" is rehashed in a content audit: "extraordinary administration and extraordinary view! The nourishment was extraordinary additionally!", at that point we accept that the feeling audit would enroll as positive.

IV. EVALUATION AND RESULTS

For surveying the execution of the proposed feature extraction estimations and a couple of classifiers we use exactness - which addresses the dimension of test tests that are gathered precisely from all test tests.

We develop a structure system for exploring differing philosophies in feature extraction procedures in mix with different classifiers as depicted in fragment III. In our first examinations, we used a case of 10 000 business reviews dataset removed from the filtered Yelp Challenge Dataset. We moreover part this dataset in 80% for planning and 20% for testing.

As seemed Table I, the best precision score on test dataset was cultivated by the structure (92.6%) using the primary procedure displayed in zone III – B, while preprocessing steps are associated before the segment decision arrange (emptying emphases) and ousting stop words, applying stemming, managing invalidations and using unigrams in molding the component vectors in sack of words depiction. The result is gotten using Stochastic Gradient Descent (SGD) Classifier with the going with parameters: the squared Euclidean standard L2 for discipline added to the adversity work and an estimation of 0.0001 of the enduringto build the regularization term.

I. CLASSIFIER ACCURACY

TABLE

Features	Naïve Bayes	Linear SVC	Logistic Regression	SGD
Unigrams + Stop words + without removing punctuations + Without handling Negations	**0.881**	**0.917**	**0.893**	**0.914**
Unigrams + Stop words + without removing punctuations + Handling Negation	**0.883**	**0.919**	**0.898**	**0.911**
Unigrams + Stop words + removing punctuations + Handling Negations	**0.897**	**0.923**	**0.900**	**0.926**
Unigrams + POS + WSD	**0.787**	**0.786**	**0.787**	**0.786**

Utilizing a similar component determination calculation and applying Linear SVC we get a shut score (92.3%), however while applying Naïve Bayes for arranging positive and negative suppositions, the framework performed somewhat more regrettable (89.7%). We get a comparable exactness score (90%) utilizing Logistic Regression classifier (MaxEnt) with L2 standard for regularization and a resilience of 0.0001 for the halting criteria.

Without applying the progression of taking care of nullifications in the element choice calculation the system's exactness, paying little heed to the classifier utilized, diminishes with 1%. Our conviction that refutations taking care of is an imperative advance in opinion investigation is affirmed. We demonstrate that without expelling accentuations from the content survey in the preprocessing stage, however applying the dealing with nullifications, the exactness marginally diminishes other than. We assess (Table II) the viability of the learning calculations utilizing the main methodology of the component extraction calculation dependent on accuracy and review, the first for positive

opinion recognizable proof, the other for negative notion distinguishing proof and f1 measure - the symphonious mean of accuracy and review (1.0 is the best esteem). Accuracy speaks to the proportion tp/(tp + fp) where tp is the quantity of genuine positives and fp the quantity of false positives. Review is as the proportion tp/(tp + fn) where tp is the quantity of genuine positives and fn the quantity of false negatives. The outcomes are acquired on the test dataset.

TABLE II. **PRECISION, RECALL AND F-MEASURE OF CLASSIFIERS**

Classifier	Positive			Negative		
	Precision	F1 Meas.	Recall	Precision	F1 Meas.	Recall
Naïve Bayes	**0.874**	**0.714**	**0.604**	**0.900**	**0.937**	**0.976**
Linear SVC	**0.895**	**0.800**	**0.723**	**0.928**	**0.952**	**0.977**
Logistic Regression	**0.952**	**0.705**	**0.559**	**0.892**	**0.939**	**0.992**
SGD	**0.911**	**0.806**	**0.723**	**0.928**	**0.954**	**0.980**

We found that the second methodology of highlight extraction, utilizing POS (grammatical form tag), applying WSD (word sense disambiguation) and processing a positive and a negative score utilizing the feeling conclusion dictionary SentiWordNet was not valuable. The exactness of the framework for all classifiers diminished with 14% when contrasted with the principal approach results. We additionally discovered that the execution diminished amid cross-approval, and in this way these highlights were excluded in the following tests. Despite the fact that this methodology doesn't require preparing contrasted and the principal approach, we trust that the significance of the entire content survey - which can be expansive - is too wide relying upon the unique circumstance. Estimation equivocalness can happen, for example a sentence of a content audit containing a positive or negative word doesn't essentially express any feeling (e.g.: "you prescribe a decent eatery near this organization").

Next, for assessing the precision of the proposed strategies we utilized the test business surveys dataset separated from the sifted Yelp Challenge Dataset which contains about 1.34 million content audits. We apply the preprocessing steps, evacuate stop words, apply stemming, taking care of nullifications and use unigrams for highlight vectors portrayal as depicted in past area (in the wake of part the expansive Yelp Challenge dataset in 80% for preparing and 20% for testing). The grouping tests were led utilizing 3-overlap cross approval for assessing the frameworks exactness.

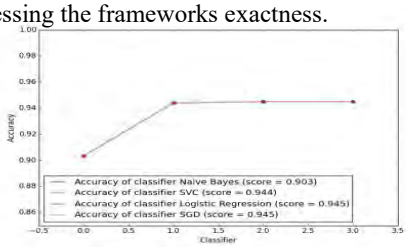


Fig. 1. Grouping precision of business surveys

An examination of precision on the test dataset utilizing distinctive learning calculations is appeared in Fig 1. The best score of 94.4 % was gotten utilizing the primary methodology on highlight extraction calculation as portrayed in area III joined with Linear SVC on the test dataset. A similar exactness was acquired when utilizing Logistic relapse for arranging the positive and negative audits.The system's accuracy imperceptibly reduces by 5% while using the Naïve Bayes classifier

V. CONCLUSIONS AND FUTURE WORK

This paper investigates the use of a few component extraction strategies and classifiers for arranging business content surveys utilizing a vast dataset: Yelp test dataset containing more than 1.6 million audits [1]. Our best classifiers Linear SVC and SGD have acquired a precision of 94.4% utilizing the main methodology proposed in the component extraction calculation. As far as execution, Naïve Bayes and Logistic Regression classifiers will in general have marginally most exceedingly terrible outcomes.

Slants articulation, sentiments, feelings and conclusions are a troublesome undertaking in humans‟ investigation. Wilson et al. discovered 82% assention by two people in the task of expression level notion extremity to consequently recognize abstract articulation [11]. Assumption examination strategies demonstrated to perform well to group slants of Yelp business surveys by considering the star rating given by the clients. We feel that the systems‟ exactness could in any case be enhanced by investigating the utilization of bigrams or trigrams, word lumps, or even grammatical form (POS) as highlights so as to recognize a similar word includes that are utilized as various POS.

REFERENCES

[1] [Online]. Available: https://www.yelp.com/dataset_challenge/dataset

[2] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," Proceedings of theACL-02 conference on Empirical methods in natural language processing, vol. 10, pp. 79-86, 2002

[3] H. Minqing and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the tenth ACM SIGKDD international conference onKnowledge discovery and data mining, 2004

[4] X. Yun, X. Wu and Q. Wang, "Sentiment Analysis of Yelp„s Ratings Based on Text Reviews", pp. 1-5, 2014. [Online]. Available: http://cs229.stanford.edu/projects2014.html

[5] J. Jong, "Predicting Rating with Sentiment Analysis," pp. 1-5, 2011. [Online]. Available: http://cs229.stanford.edu/proj2011/Jong-PredictingRatingwithSentimentAnalysis.pdf

[6] F. Mingming and M. Khademi, "Predicting a business star in Yelp from its reviews text alone", arXiv preprint arXiv:1401.0864, 2014

[7] S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. Reis, and J. Reynar, "Building a Sentiment Summarizer for Local Service Reviews", Proc. of the WWW Workshop on NLP Challenges in theInformation Explosion Era, NLPIX, 2008

[8] D. Sanjiv, M. Chen, "Yahoo! for Amazon: Sentiment extraction from small talk on the web", Management Science, pp. 1375-1388, 2007.

[9] A. Esuli and F. Sebastiani, "SentiWordNet:A Publicly Available Lexical Resource for Opinion Mining", Proceedings fromInternational Conf. on Language Resources and Evaluation (LR),2006, pp. 417-422

[10] "scikit-learn" [Online]. Available: http://scikit-learn.org/

[11] T. Wilson, J. Wiebe and P. Hoffmann "Recognizing contextual polarity in phrase-levelsentiment analysis", Proc. of the conference onHuman Language Technology and Empirical Methods in NLP, 2005

Forecast of Foreign Exchange Prices with techniques of Regression

Yadnesh Mohan Zagade, Shivani Gupta, Siddhesh Vinay Vartak, Dr. Sheetal Rathi
Thakur College of Engineering Mumbai, Maharashtra India

zagade.yadnesh9@gmail.com, shivanigupta3005@gmail.com, vsid55@yahoo.co.in, sheetal.rathi@thakureducation.org

***Abstract*— This paper investigates and contrasts relapse procedure and outfit relapse methods in perspective on two group getting the hang of: Bagging and Boosting for two Foreign Exchange (FX) information to be specific INR/USD and INR/EUR. The similar outcomes demonstrate that relapse outfit with Least Square Boost performs superior to others with MAPE =0.6338 for one day ahead expectation at the testing stage. It has likewise been seen that MAPE esteem increments with expanding estimation of N if there should be an occurrence of N-Days ahead forecast of FX rate.**

Keywords*— Regression, Foreign Exchange (FX), Ensemble Techniques.

I. INTRODUCTION

Money related time arrangement information is confused, uproarious, fluctuating and nonlinear as various occasions have happened in different timeframes. In this way, it is trying for specialists to build up the precise prescient model. Expectation for Foreign Exchange (FX) rate (Galeshchuk, 2016) is additionally an exceptionally critical errand for N days ahead forecast due to unstable nature of FX Data. Factual procedures are not ready to productively anticipate the FX rate. Consequently, unique machine learning strategies have been utilized by numerous analysts for exact forecast.

A. In this examination work, we investigate two procedures for FX rate expectation: Regression and Ensemble Regression. Nonetheless, relapse method is broadly utilized for direct information forecast, yet not all that well known for nonlinear information expectation. Then again, gathering relapse method may enhance the execution of foreseeing model. These two strategies are connected to two diverse FX information, i.e., INR/USD and INR/EUR to discover N days ahead forecast. Models were group (Opitz et al., 1999) utilizing the gathering learning procedures: Bagging and Boosting (Freund et al., 1999) and delivers progressively exact outcomes when contrasted with the individual model. A kind of boosting as Least square boosting (LSBoost) is utilized in this examination work for the analysis. The first FX rate information has just a single component which will most likely be unable to foresee following day FX rate all the more precisely. In this manner, highlights are extricated from the first FX rate information utilizing five specialized pointers (Handa et al., 2015) proposed by numerous creators. Trials were finished utilizing K-overlap cross approval system through self-composed MATLAB code.

II. LITERATURE WORK

Monetary time arrangement information determining through machine learning procedures are drawing in analysts in the previous multi decade. Measurable strategies, Data mining, and delicate registering systems are the most prevalent methods for this. Nonetheless, as of late information mining systems like choice tree strategies and Artificial Neural Network (ANN) and its varieties as delicate processing methods are more dominant procedures than others. These are organized in Table 1 as beneath:

Table 1: Summary of Literature Review

YEAR	REFERENCES	JOURNAL
2017	Hota et al.	International Journal of Computational Intelligence Research
2016	Galeshchuk	Neurocomputing
2015	Simidjievski et al.	Expert system with applications
2015	Handa et al.	International Journal for Research in Applied Science & Engineering Technology
2014	Rehman et al.	ScienceDirect
2011	Pacelli et al.	Journal of Intelligent Learning Systems and Applications
2000	Dietterich	Machine Learning
1999	Schapire	Proceeding of sixteenth International Joint conference on Artificial Intelligence

III. INFORMATIONAL COLLECTION AND METHODOLOGY

Two FX rate information are gathered from www.fx.sauder.ubc.ca web source beginning from November ninth, 2016 to July 31st, 2017. An aggregate of 189 examples of informational index was standardized by separating each example with the most noteworthy estimation of the example in FX information. This preprocessing system is utilized to change the crude information into some important information. Standardization is ordinarily utilized preprocessing system for information smoothing. Condition 1 is utilized for information standardization, which scales information in the scope of [0 1]:

Xnew=XXmax (1) Where x is day by day conversion scale of FX information, Xmax is most elevated estimation of FX information and Xnew is acquired standardize information. So as to produce new applicable and important highlights, include extraction method is connected on standardized FX information dependent on 5 specialized pointers, i.e., SMA (Simple Moving Average), EMA (Exponential Moving Average), WMA (Weighted Moving Average), Variance and Standard Deviation as appeared Table 2.

The accompanying two procedures were utilized to manufacture learning models:

Relapse: Regression is the information mining procedure (Han et al., 2012) used to anticipate numbers as it were. Relapse utilizes the information where the objective esteem

is known. It evaluates the connection between one ward variable and at least two free factors.

Troupe Model: For expanding the forecast exactness of any prescient model, gathering models have been created. A gathering model is a blend of prepared classifiers (Optiz et al., 1999) like the neural system or choice tree. For building up a novel model, the expectations originating from individual models are joined to deliver a more exact outcome than any single model. The outfit is usually utilized for machine learning undertakings (Simidjievski et al., 2015) like relapse and arrangement. In Regression group two famous strategies are utilized: Bagging and Boosting. Sacking (Bootstrap Aggregation) is utilized to diminish the change for expanding exactness in expectation (Han et al., 2012) utilizing distinctive preparing sets. In sacking new preparing information is produced utilizing diverse preparing set where some preparation sets are utilized more than multiple times or some preparation sets may not be incorporated.

The sacked model (Breiman, 1996) gives preferable execution over the single model which delivers the last outcome dependent on the normal anticipated esteem. Then again, in boosting weight is relegated to each preparation information (Han et al., 2012). The preparation tuples might be chosen more than one time which depends on its weight.

For boosting approach, the dependable guidelines are connected utilizing strategies or calculations for discoveries. This calculation of boosting is designated "frail" learning calculation. The loads of powerless students get expanded and these loads go for next round to create preparing tests to support the exactness of the model. This procedure is rehashed each time with various subset of the preparation tests. Each time in all adjusts the single expectation rule is created by joining all these powerless principles by boosting calculation. Boosting centers around the loads of frail students in the past round. This is the principle thought behind boosting.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

IV.RESULT ANALYSIS

Trial work is finished by utilizing a self-composed MATLAB produced code for Regression and Ensemble Regression (Bagging and LSBoost). Relapse and outfit relapse strategies have been utilized for exact forecast of FX rate with new component space. FX information acquired after component extraction were powerfully apportioned into preparing and testing informational indexes utilizing k-crease cross approval. Cross approval strategy (Hota et al., 2017) is superior to the static technique for apportioning the information. Static apportioning of information with a settled level of preparing and testing information may predisposition the learning models. Then again, dynamic parcel of information like k-crease segments information as preparing and testing progressively in which each overlap partakes in both preparing and testing stages. For investigation k=10 crease cross approval was utilized in this exploration work. Preparing and testing information were displayed to relapse and outfit relapse for N days ahead forecast as 1-Day, 3-Days, and 5-Days. The anticipated estimation of every day is thought about against the relating genuine estimation of following day FX rate for both FX information: INR/USD and INR/EUR. The exactness of the prescient model is examined by ascertaining Mean Absolute Percentage Error (MAPE) as appeared in Equation 2 and displayed in Table 3 and Table 4 separately for INR/USD and INR/EUR.

$$MAPE=(\Sigma|Ya,i-Yp,i|)/n$$

Table 2: Comparative results of Regression and ensemble Regression in terms of MAPE for INR/USD at testing stage.

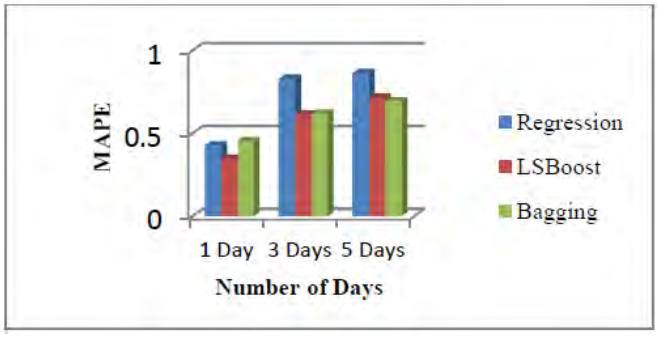
N-Day	Regression		Regression Ensemble
	LSBoost	Bagging	
1-Day	0.429	0.350	0.455
3-Day	0.829	0.617	0.620
5-Day	0.863	0.715	0.695

Table 3: Comparative results of Regression and ensemble Regression in terms of MAPE for INR/EUR at testing stage.

N-Day	Regression		Ensemble Regression
	LSBoost	Bagging	
1-Day	0.839	0.633	0.806
3-Day	1.239	1.097	1.022
5-Day	1.297	1.193	1.161

These tables demonstrate the outcomes according to desire, out of two procedures, relapse group is delivering better outcomes and reliably expanding estimation of MAPE while expanding the estimation of N. Then again, out of two outfit systems LSBoost is creating better expectation results (MAPE= 0.3502 for 1-Day ahead forecast and MAPE= 0.6178 for 2-Days ahead forecast) while Bagging is delivering marginally better outcomes (MAPE= 0.6954 for 5-Days ahead forecast) for INR/USD FX uncommon information. It has additionally been seen that there is less change in INR/EUR FX rate when contrasted with INR/USD FX rate for the picked period.

Results are represented with the graphical representation.



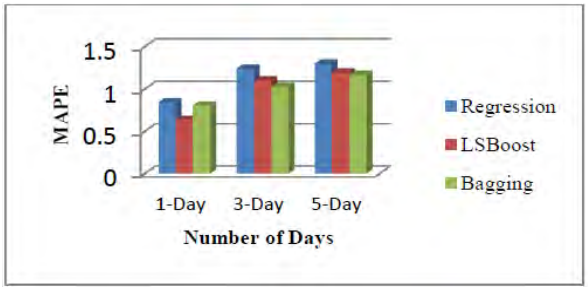


Figure: Near MAPE of Regression, LSBoost and Bagging for (an) INR/USD FX information and

(b) INR/EUR FX information

V. V.CONCLUSION

FX rate expectation utilizing direct model is somewhat dreary because of nonlinear nature of information. Relapse system, be that as it may, isn't as prominent as machine learning procedures particularly for time arrangement information expectation which endeavors to discover best fit direct line regarding one ward variable and at least one free factors. Then again, gathering relapse joins more than one systems together and enhances the outcome if there should be an occurrence of nonlinear information up to some degree.

This paper investigates two procedures for expectation of two FX rate information in particular INR/USD and INR/EUR with five separated highlights utilizing specialized markers as recommended by creators. Results demonstrate that troupe system with both sacking and boosting is performing superior to relapse procedure for N-Days ahead forecast. Likewise, both the troupe methods are performing better with slight varieties as far as MAPE

REFRENCES

[1]. Pacelli, V., Bevilacqua, V. and Azzollini, M. (2011). An Artificial Neural Network Model to Forecast Exchange Rates, *Journal of Intelligent Learning Systems and Applications*, 3, 57-69.

[2]. Rehman, M., Khan, G. M. and Mahmud, S.A. (2014). Foreign Currency Exchange Rates Prediction using CGP and Recurrent Neural Network, *IERI Procedia,* 10, 239-244. Hota, H.S., Handa, R. and Shrivas, A.K. (2017). Time Series Data Prediction Using Sliding Window Based RBF Neural Network, *International Journal of Computational Intelligence Research*, 13(5), 1145-1156.

[3]. Han, J., Kamber, M. and Pei, J. (2012). Data Mining Concepts and Techniques, *Morgan Kaufmann*.

[4]. Opitz, D. and Maclin, R. (1999). Popular Ensemble Methods: An Empirical Study, *Journal of Artificial Intelligence Research,* 11, 169-189.

[5]. Hota, H.S. and Tandan, S.R. (2015). Stock Market Prediction with Various Technical Indicators Using Neural Network Techniques, *International Journal for Research in Applied Science & Engineering Technology*, 3(6), 604-608

[6]. Breiman, L. (1996). Bagging Predictors, *Machine Learning*, 24, 123-140.

[7]. Dietterich, T.G. (2000). An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization, *Machine Learning*, 40, 139–157.

[8]. Galeshchuk, S. (2016). Neural networks performance in exchange rate prediction*, Neurocomputing*, 172, 446-452.

Integration of C# with CUDA for performance improvement of Standalone applications

Abhinav Watve, Dr. Vinayak Bharadi *HOD,* Siddhesh Salvi, Sarvesh Horambe
Information Technology Finolex Academy of Management and Technology Ratnagiri
watveabhinav@gmail.com, vinayak.bharadi@famt.ac.in, siddheshsalvi525@gmail.com, sarveshshorambe@gmail.com

***Abstract*—Standalone applications which require heavy computational power take much time for processing. By using CUDA for execution of these heavy processes, we can reduce the time required for processing.This integration of C# with CUDA can be done using mainly 3 methods. These 3 methods involve exporting CUDA C code to DLL file, PTX file and writing C# code and convert it to CUDA code using Hybridizer.All of these methods are valid for integration. But, the performance may vary from method to method.**

Keywords—CUDA, C#, Nvidia GPU

I. INTRODUCTION

This paper includes different methods to integrate a standalone application developed using C# with CUDA C program for performance improvement of these applications. There are mainly three techniques by which this integration can be achieved. First technique includes exporting the function to call from C# code to a DLL file and importing that DLL file in your project. Second is exporting the function to a PTX file using NVCC commands and importing that your project using ManagedCUDA package from GitHub. Third is writing the code to run parallely in C# and then using Altimesh Hybridizer to Hybridize the code to run on Nvidia GPU.[1,2,3,4]

II. NEED OF INTEGRATION

Performance Imporvement
The general purpose CPUs usually contain 4 cores for a desktop computer. This means that 4 threads can be running at the same time. Also, these CPUs are developed to be used for general purposes. This means that they execute all types of instructions like memory read, write, peripheral device instructions, Ram management, etc. Due to this build of the CPU, they are not as efficient as the GPUs for mathematical computations.[1]
Nvidia GPUs contain thousands of GPU cores which can run the same instructions parallely, which helps to improve the performance. Also, these GPUs are specifically developed to efficiently run mathematical instructions.[1]

Simplicity of Coding
In typical programs which run on CPU, to execute every statement multiple times loops are used. But in the case of using CUDA, the programmer only needs to write the instructions that need to be executed parallely. While calling the function, we can specify the constraint of how many times the instruction needs to be executed.[1]

III. METHODOLOGY

In the following listed three methods, the CUDA code is written in C and it will be called from the C# application. Even though the methods to call the functions differ, the results will be same.

Using DLL file
To export the CUDA C++ file to DLL file, you can use the setting provided by Visual Studio to export to DLL, file but it also needs to add some extra lines of code, as follows
Kernel.h
#pragma once #ifdef KERNEL_H #define KERNEL_H

#include "cuda_runtime.h"
#include "device_launch_parameters.h"

#ifdef cplusplus extern "C"
#endif //

void declspec(dllexport) FunctionName(ParameterList);

#ifdef cplusplus #endif // cplusplus
#endif

This file defines the method in the DLL file, through which it can be called from the C# project.

Project.cs

```
void declspec ( dllimport ) FunctionName(ParameterList);
```

This way the file is imported into the C# project. After this, the function can be directly called as if it is function in header file.
This completes the coding part for DLL file. Now to build it, Go to the Project Project properties Configuration Properties General Configuration Type, set this property to Dynamic Link Library(.DLL). after this build the project, a .dll file be generated in the project directory[1]

Using PTX file
This technique does not require to add or change CUDA code but it needs to add ManagedCUDA package from GitHub
Install-Package ManagedCuda-100 -Version 10.0.31
Above shown command can be used inside Visual studio to add the latest ManagedCUDA version to the C# project. This takes care of the package required to import the PTX file to C# project. Now, to convert CUDA C++ file to .ptx file, a simple command is provided by yhe NVCC compiler which generates a .ptx from a .cu file which is as follows
>nvcc –ptx kernel.cu
This command should be given after changing the current working directory to where the kernel.cu file is present. This completes the requirement of all the resources required for integration. In the C# project we need to write the code to implement the methods.[2]

Using Hybridizer

This is very for the programmer to implement. In this method the multithreaded code is written in C#. Then, Altimesh Hybridizer is used to directly hybridixe the code to run on the CUDA framework.

```
public static void VectorAdd(double[] a, double[] b, int N)
{
Parallel.For(0, N, (i) => { a[i] += b[i] });
```

Above shown is a simple example of how vector addition is done using multithreading in C#. This can be directly hybridized using Altimesh Hybridizer to run on CUDA framework.[4]

IV. ACKNOWLEDGMENT

This paper was developed from the research of the three authors and the guide who guided us for the paper. We would like to acknowledge all the efforts of the people who guided us, the teachers as well as our colleagues.

REFERENCES

The template will number citations consecutively within brackets

[1]. The sentence punctuation follows the bracket

[2]. Refer simply to the reference number, as in [3]—do not use "Ref. or "reference [3]" except at the beginning of a sentence: "Reference [4] was the first ..."

[3]. Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

[4]. Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished"

[5]. Papers that have been accepted for publication should be cited as "in press"

[6]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

[7]. For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation

[8]. Ayman M. Wahba, Fatma Zaky Sakr, Mohammed Taher, "High Performance Iris Recognition System On GPU",29 Nov.-1 Dec. 2011,The 2011 International Conference on Computer Engineering & SystemsJ.

[9]. https://stackoverflow.com/questions/21956127/creating-cuda-dll-and- using-it-on-vc-project

[10]. http://kunzmi.github.io/managedCuda/

Cloud Based Automated Malware Analysis for Advanced Persistent Threats for Government organization

Nikhil Pandey, Amit Tiwari, Ekta Tiwari
Dept. of Computer Engineering TCET
pandeynikhil804@gmail.com,tiwariamit.0308@gmail.com,ektatiwari43@gmail.com
Mr. Vikas Singh (Guide)

***Abstract*- Goal of this project is to make your web application more difficult to hack. Web Application or any software for that matter will never be completely secure and free for detect. It is only matter of time. Before a determined attacker will find some vulnerability or misconfiguration to exploit and compromise either your site or one of its user.**
You should take moment to come to terms with this truth before processing. Many people wrongly assumed that hiring smart developer or deploying commercial security product will magically make their site hacker proof. So, we are making a virtual image of system implemented by top security features.

I. INTRODUCTION

The Project will be Self configurable Virtual Machine Image which can be used by an enterprise to protect its own application by putting VM image in front of Application. Enterprise need to develop application which can demonstrate correct functionality. Non Functional Requirement of the application like Scalability and Security will be handle by Self configurable Virtual Machine.

Here VM will protect Enterprise wide application though various Hierarchical mechanism. VMs will allow Access level Protection along with Predictability of Attack on Application**.**

The outcome of the Project will be Self configurable Virtual Machine Image which can be used by an enterprise to protect its own application by putting VM image in front of Application. Enterprise need to develop application which can demonstrate correct functionality. Non-Functional Requirement of the application like Scalability and Security will be handle by Self configurable Virtual Machine. Here VM will protect Enterprise wide application though various Hierarchical mechanism.

II. THEORY

Now a day's malicious attack keep occurring. One of the main reasons is the poor quality of the software used in systems and application software. So, we are came up with the innovative idea for making network secure from malicious file entering network. We are making a product which will be act as intermediate between the client and users (legitimated users / hackers). This product will fully responsible for detecting malicious files that enter the organization network

Network will never be completely secure and free for detect. It is only matter of time before a determined attacker will find some vulnerability or misconfiguration to exploit and compromise either your site or one of its user. You should take moment to come to terms with this truth before processing. Many people wrongly assumed that hiring smart developer or deploying commercial security product will magically make their network hacker proof.

Making a product with highly configured security core rules. Initially we are developing a sandbox for static and dynamic analysis of a file in a network. The existing system include direct access of client and server, where in people are just relying on where they are hosting there website, or a platform where they are deploying there website which is completely dependent on the users concern.

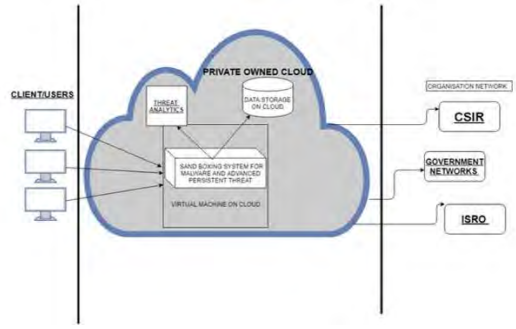
We need to analyze the need of the clients for our project also the demand in the market. Everything depends on demand and supply policy. If the demand is more and use of our product will be more. We need to analyze how our system is better compared to other systems this can be done by identifying gaps and also providing features that our others are not providing and make our product more useful.

This project will majorly attract the population of the conservative mindset with proper use of technology to lover the exhaustion of cloud based knowledge and startups.

III. PROJECT WORK

Our Project idea is to provide Cloud based Security against signature/anomaly based attacks i.e. sophisticated malware and zero-day threats to government network and Websites without sharing any data to any public cloud (AWS, Azure) vendors. Zero day threats include those undetected threats that can be occurred due to system inaccuracy and they are undiscovered till they occur. Our system with the help of behavior analysis can thwart this threat by signaling any misconduct in the system.This whole process can be done either on the commodity hardware and also cloud services can be used.
Cloud will however be a better option if we keep automation, simplicity and organizational cost in mind.

Many government organizations use third party cloud which is not a full proof safe idea considering the importance of the data. Our idea is to have a government owned private cloud that will protect the government network against the malicious attack by implementing a sandbox system inside a virtual machine thereby analyzing the malicious file and giving a comprehensive report regarding the file by checking both pattern analysis and behavior analysis.



System Architecture

IV. CONCLUSIONS

This cloud based system can be used to protect the government network by blocking any malicious activity by its sandbox (efficiency of sandbox depends on its implementation). The detailed reporting of the malicious files provided can also be used for the security auditing of the organization. The proposed use of government owned private cloud adds to the security and also to the overall performance of the system. Deployment of the system on cloud will provide much flexibility to its use. The cloud will act as an intermediary between the end user and the government network thereby securing the government network. It will also provide the protection from zero day attacks which cannot be achieved by the products of current leading vendors.

V. FUTURE SCOPES

This project will majorly attract the people those running his /her business online and want to grow his business with full boost. In initial state it may not stand with the today's security service provider like GoDaddy and CloudFlare and many more. But it will be best product in future. In market there are many security service provider as competitor.

VI. ACKNOWLEDGEMENT

We sincerely thank our guide Mr. Vikas Singh for her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

We thank the Head of Dept. Dr. Sheetal Rathi, Dean Academic Dr. R.R Sedamkar, the Principal Dr. B. K. Mishra and the college management for their support.

REFERENCES

[1] Ryan C. Barnett, Web
[2] ApplicationDefender's

[3] Cookbook, Battling Hackers and Protecting Users, December 2012.

[4] Ivan Ristic, Modsecurity Handbook, 2010.

[5] OWASP top ten security core rule set for latest security in web application https://www.owasp.org/index.php/Category:OWASP_ Top_Ten_Project (Accessed on 12 October, 2016)

[6] OWASP Appsensor Project
[7] https://www.owasp.org/index .php/OWASP_AppSensor_P roject (Accessed on 15 September, 2017)

Design of Efficient URL based Web Page Classification using NLP

Karan Risbud, Rohit Singh, Vivek Vishwakarma, Dr. R. R. Sedamkar
Computer Engineering, TCET University of Mumbai
risbudkaran@gmail.com, rohits130198@gmail.com, vishwakarmavivek83@gmail.com, rrsedamkar27@gmail.com

***Abstract* – Exponential increase in the number of web pages in the World Wide Web poses a great challenge in information filtering and also makes topic focused crawling a time consuming process in searching for relevant information. We propose an URL based web page classification method that does not need either the web page content or its link structure. In the proposed approach, character n-gram based features**

Are extracted from URLs alone and classification is done by Support Vector Machines and Maximum Entropy Classifiers. The performance of the system was evaluated on two bench mark datasets viz., ODP with 2 million URLs and WebKB with 4K URLs. We experimental used Fl results as a showed performance an metric improvement and of our 20.5% increaseon WebKB dataset and 4.7% increase on ODP dataset.

I. INTRODUCTION

Web page classification is the process of assigning a web page to one of the predefined categories. In the existing approaches, the classification task is performed by using contents of the web pages or contents of sibling pages and HTML tags. Using the contents of a web page slows down the classification speed since the web age needs to be downloaded for extracting features. URL based Web page classification systems have gained popularity, as it improves classification speed by using features extracted from the URLs alone without the need for downloading the page. URL based features can be combined with content based features, to improve the classification accuracy further. In this paper, we show that applying only URL features can give good performance when suitable features are derived from URLs.

Many organizations and educational institutions provide the Internet facility, but want to block some unnecessary web pages or to restrict heavy download (like movie download). So the classification should be done before fetching that web page and seeing the contents of it. Topic focused crawlers need to fetch the relevant topic-specific information by

avoiding unnecessary downloads that waste the bandwidth. Such applications need an automated URL based web page classification system, since the existing approaches are not desirable when the bandwidth is limited, content is not available or to block unnecessary websites before downloading that page.

Web page classification based on URL is a challenging task, since URL is a small fraction of a web page and contains compound words (e.g. http://www.realestatechennai.com to represent Real Estate in Chennai), abbreviations (e.g. http://www.bbc.co.uk, to denote British

Broadcasting Corporation) or non-meaningful words derived from the original words in English (e.g. http://www.espncricinfo.com/ for cricket information from Entertainment and Sports Programming Network). Some URLs may not contain any related information about a web page posing more challenge on classification (e.g. http://www.dvk.com/).

In text classification, dictionary based approach is followed to select a subset of document terms for classification. Using statistical properties of terms in the training dataset, dictionary is constructed with important terms that are used as features for classification. Baykan et al. [5] applied such statistical dictionary based approach for URL based web page classification. But dictionary based methods containing URL tokens (eg. espncricinfo) fail to c1assity a URL when such tokens are seen only in the test set. So character n-gram based approach is proposed in this paper that overcomes these challenges and still performs web page classification by using the features derived from URLs alone.

Character n-grams are character sequences of length n (e.g. 3-grams in TEXT: TEX, EXT). For URL classification task, 3-grams can be used as features instead of tokens as 3-grams capture the abbreviations, compound words, shortened words etc. better than tokens. Even though 3-gram dictionary based approaches perform better than token dictionary, it greatly depends on 3-grams chosen as dictionary terms. Since it involves more manual work or depends on some ranking methods to construct dictionary, we propose a novel feature representation by using all possible 3-letter combinations (among 26 alphabets) as features, thereby avoiding the need for constructing the dictionary. In our approach, the term frequencies of 3-grams present in an URL form the feature vector in 26^3 (17576) dimensional feature space. In traditional bag-of-words approach, the term frequency of all the 3-grams dictionary (important 3-grams in all the training URLs) form the feature vector. In both cases, the sparsity (average number of non-zero terms) of feature vector is inevitable.

In the proposed method, the features are dataset independent. The advantage of using dataset independent features is that, the dimensionality of feature vector is independent of any dataset with fixed number of features (17576). This makes the proposed method to be applicable for any public dataset as well as personal dataset containing some collection of URLs.

We perform binary classification by creating individual classifier models that are trained with

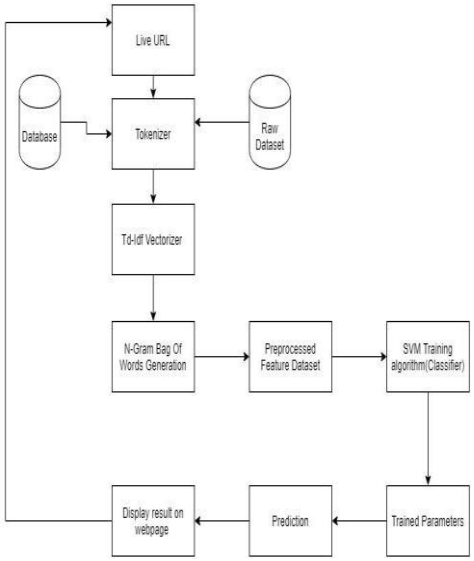
3-gram features derived from corresponding category URLs. By this approach, more categories can be added or removed depending on the requirement without the need of retraining the existing classifier.

The proposed approach was evaluated on two bench mark datasets, a large dataset Open Directory Project (ODP) with 14 categories containing about 2 million URLs and a small dataset WebKB with

4 categories containing 8K URLs. We used two different machine learning algorithms that can handle very high dimensional feature space and sparse feature vector, Support Vector Machines(SVM) and Maximum Entropy(ME) Classifier for the classification task with the 3-gram features. Our experimental results showed an improvement of 20.5% in F-measure value on WebKB dataset and 4.72% increase in F-measure value on ODP dataset when compared with existing methods.

Proposed Work:

System Architecture:

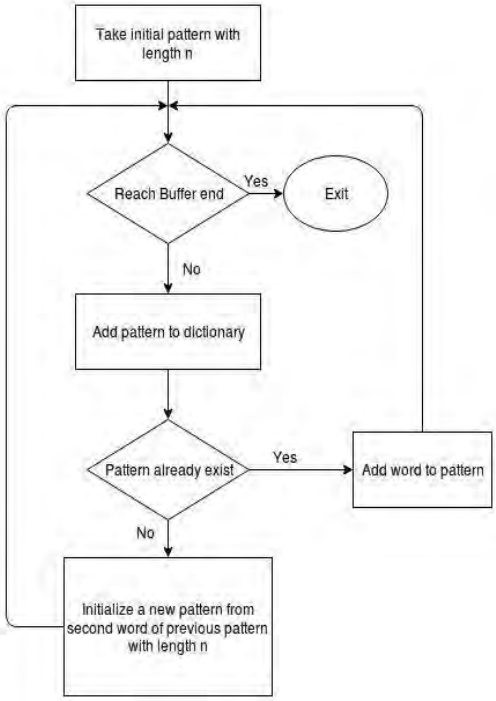


Algorithm/Implementation:

Implementation of N-Gram in Python:

```
def getNGrams(wordlist, n):
ngrams = [] for i in range(len(wordlist)-(n-1)):
ngrams.append(wordlist[i:i+n]) return ngrams
```

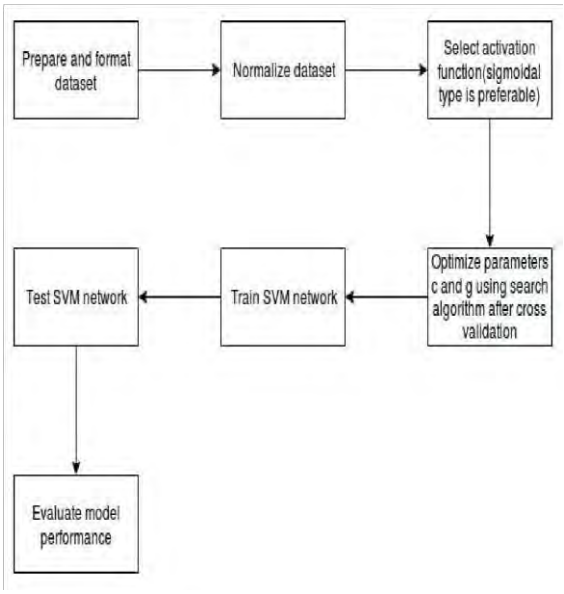
Flowchart(N-Gram):



Implementation of SVM in Python:

```
#Import Library from sklearn import svm
#Assumed you have, X (predictor) and Y (target) for training data set and x_test(predictor) of test_dataset
# Create SVM classification object
model = svm.svc(kernel='linear', c=1, gamma=1)
# there is various option associated with it, like changing kernel, gamma and C value.
model.fit(X, y) model.score(X, y) #Predict Output
predicted= model.predict(x_test)
```

Flowchart (Support Vector Machine):



Literature Survey:

The existing automated classification systems classify web pages based on HTML content or by structure of the document or by generating summaries to decide the category of the web page under consideration. Categorized the web pages using not only the contents of it, but also with class information from the neighboring pages. They used class of neighboring pages to determine the topic of a web page.

As the content based approaches are not suitable when the web page contains only images.

Since content based classification is inefficient for pages with images so URL based classification is used. There were many existing approaches for categorizing URL through words or texts or sentences or phrases. After reading few papers we came to know about few approaches are as follows:

1. All gram with Naive Bayes Classifier
2. N-Gram with Naïve Bayes Classifier
3. N-gram with SVM Classifier

Feasibilty Analysis:

II. ECONOMICAL FEASIBILITY

The EFS is composed of two required forms:

- **Business Case**

The Business Case provides an analysis of the business environment including-

- **Cost Benefit Analysis:**

The Cost Benefit Analysis summarizes the revenues and costs involved with the proposed project. As the proposed system will be used for the benefits of the public, no additional cost will be paid by them. No hardware system is included in our project, so the hardware cost gets minimized. Only minimal amount will be required for hosting the Django framework. Hence, our system is Cost Efficient.

III. TECHNICAL FEASIBILITY :

Technology used at front end: CSS, bootstrap, HTML,Javascript
Technology used at Back end: Programming language – Python , Framework - Django
Resources Required: Manpower, Programmers, testers, debuggers
Software required: Testing Tools (to perform black box and white box testing), Python Debugger(PDB)
Editor: Visual Studio Code

IV. MANAGERIAL FEASIBILITY :

Management support, public involvement, and commitment are key elements required to gauge managerial feasibility in the proposed project. The success of the project partly depend on managerial competence of the major ingredients of the proposed project which are the users i.e. companies & institutes and the benefiters i.e. general public. The capability of the infrastructure of a process is to achieve and sustain the properties of isolation, atomicity, durability, consistency etc in the matter of the data of the users and benefiters.

V. OPERATIONAL FEASIBILITY

The proposed system is concerned with the end users who are the general public. At initial phase the scope of the system will be at local level which will be helpful for the locals. The website will be directly used by the end users for the classification of web pages and for detection of a malicious website. The end users of the system will be the institutes(at local level)or the companies(at global level).The important stakeholder of the system will be the public, who will be directly using the system. Detection of malicious website is the key process of the system which will be easily handle by an automated part of the system. If the usage of the system is very beneficial and helpful to locals then the system can be expand at the global level also. At the end the system will contribute for protection of data of the end users.

VI. CONCLUSION

We have presented a malicious url detector which can detect percentage danger of a url based on the tokens present in the url. The technique used is efficient than all gram model and the decision boundary of SVM classifier is more accurate for new data than decision boundary of Logistic Regression.

REFERENCES

[I] R.Rajalakshmi and Chandrabose Aravindan,” Page Classification using n-gram based URL Features”2013 Fifth International Conference on Avanced Computing(ICoAC).

[2] S. Meshkizadeh and A. Masoud-Rahmani, "Webpage classification based on compound of using HTML Features and URL features and

features of sibling pages," Int. J. Adv. Compo Techn., vol. 2(4, pp. 36-46, 2010.

[3] R. Rajalakshmi and C. Aravindan, "Naive bayes approach for website classification," in Information Technology and Mobile Communication. Springer Berlin Heidelberg, 20 I I, pp. 323 - 326.

[4] M.-Y. Kan and H. O. N. Thi, "Fast webpage classification using URL features," in Proceedings of the 14th ACM international conference on Information and knowledge management,ser. CIKM '05. New York, NY, USA: ACM, 2005, pp. 325-326.

[5] E. Baykan, M. Henzinger, L. Marian, and 1. Weber, "Purely URL-based topic classification," in Proceedings of the 18th international conference on World wide web, ser. WWW '09. New York, NY, USA: ACM, 2009, pp. 1109-1110. [6] L. Wern Han and S. M. A1hashmi, "Joint web-feature (JFEAT): A Novel web page classification framework," Communications of IBMA, vol. Article 1D 73408, 2010.

[7] C . H . H . Rung Ching Chen, "Web page classification based on a support vector machine using a weighted vote schema," Expert Syst. Appl., vol. 31 (2), pp. 427-435, 2006.

Development of a Knowledge Management Tool for Calculating Attainments of Course and Program Outcome

Md. Awes Shaikh, Ashutosh Singh, Saurabh Singh, Jigarkumar Soni
Computer Engineering, TCET, University of Mumbai
skawes1997@gmail.com, atsingh19997@gmail.com, singhsaurav727@gmail.com, jigarsoni181@gmail.com

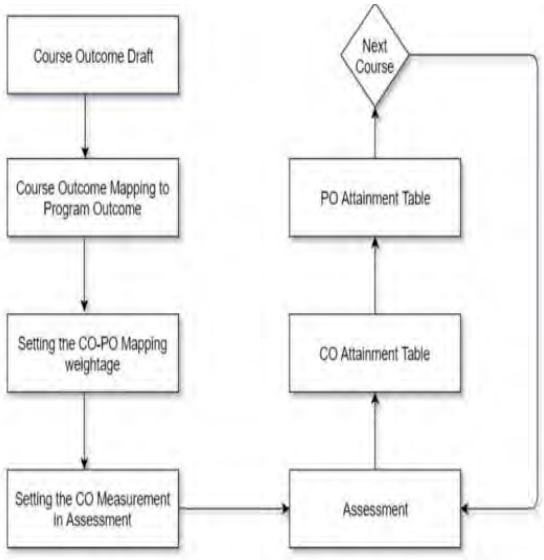
***Abstract* - National Board of Accreditation (NBA) has set out a few conventions to assess the Outcome of a program. The conventions for an instructive establishment is excessively monotonous and time expending for an instructive organization to perform physically. To defeat this issue we aim to build up a device that can create the assessments consequently as indicated by standard rules given by National Board of Accreditation (NBA). Outcome based instruction gives a shot for a general realizing where a building graduate gets prepared in all the program Outcome laid by the National Board of Accreditation, India. This guarantees an alumni not just have a sound information in the particular program they additionally can have a worldwide versatility and acknowledgment. With the end goal to accomplish the program outcome the educational modules for the program is planned in a way that the understudies get prepared in all the 12 Program Outcomes. The courses are instructed to the understudies by different encouraging systems like classroom educating, research center investigations, workshop and ventures. Each course is composed with particular Course Outcome and each Course Outcome is mapped to the Program Outcome. The accomplishment obviously outcome is evaluated by surveying understudies. The evaluation philosophy includes nitty gritty inquiry paper enunciation wherein each inquiry is mapped to a Course Outcome and it's relating System Outcome.**

I. INTRODUCTION

Graduate engineer is considered to be a global citizen possessing necessary professional skills suitable for the global scenario. Here with we require a method of quantification of their skills. Each institute has its own vision and mission which is abided by the individual programs through their vision and mission. The National Board of Accreditation (NBA) in India has indicated a very well defined set of Program outcomes. These program outcomes can be oriented towards the specific discipline to direct the framing of course outcomes [1, 2]. The undergraduate curriculum is designed in manner that the students will be taught fundamental courses which prepares them to take advanced courses in the program in the form of elective or higher level advance courses [4]. National Board of Accreditation (NBA) has laid down some protocols to evaluate the outcome of a program. The protocols for an educational institute is too tedious and time consuming for an educational institute to perform manually. To overcome this issue we intent to develop a tool that can generate the evaluations automatically according to standard guidelines given by National Board of Accreditation (NBA). Although the project is beneficial only for colleges under National Board of Accreditation (NBA) it will reduce the tedious work of faculties by automating the process of calculating Course and Program Attainment. There will be a functionality which will allow the faculties to directly import the data of the students via excel sheet so there is no need of entering the data twice. The project aims to help the faculties of the colleges under National Board of Accreditation (NBA) and there is a pretty good requirement of the project. The project targets the faculties and the end product will help them to attain the Course Outcomes and Program Outcomes. There are many future scopes of the project which are not foreseen currently. The end product will have the ability to map Course Outcomes and Program Outcomes. Also it will help faculties to see the data of students & decide whether a course has attained its outcomes. Outcome based education ensures that a graduate not only possess a sound knowledge in specific domain but they also can have a global mobility and acceptance.

II. PROPOSED WORK

System Architecture:



Algorithm/Implementation:

This area proposes a technique for evaluating the relationship between the Course and the Program outcome:

Map each question with corresponding Course Outcomes and learning levels

Map each CO with different learning levels

CO341	CO342	CO343	CO344	CO345	CO346
U	R A E	R A AN	R A AN C	AN	AN

Count the number of students in each level and find the attainment of each LO using **Experiment − wise attainment** = 3 ∗ No of students in Level 3 + 2 ∗ No of students in Level 2 + 1 ∗ No of students in Level 1 ∑ Number of students on level i i=3 i=1 E.g. CO1 (EXP 1) Attainment = 3∗5+2∗0+1∗ 0 5 = 3

Literature Survey:

The course outcomes are defined by certain brainstorming sessions in a team comprising of various faculty members, module and course coordinators and the program coordinator. The course Outcomes in our area of expertise are characterized by certain meetings to generate new ideas in a group including different employees, module and course facilitators and the program organizer [3, 5]. The course outcomes are defined using the following keywords:

Yr.	No.	Remember	Understand	Apply	Analyze	Evaluate	Create
FE	1	Label	Compare	Change	Conclude	Choose	Arrange
	2	List	Explain	Use	Deduce	Test	Collect
	3	Select	Illustrate	Show	Question	Revise	Modify
	4	Name	Classify	Complete	Illustrate	Evaluate	Rewrite
	5	State	Derive	Calculate	Outline	Determine	Create
	6	Write	Differentiate	Classify	Identify	Contrast	Construct
	7	Read	Discuss	Illustrate			Revise
SE	1	Outline	Convert	Compute	Differentiate	Explain	Solve
	2	Indicate	Give examples	Relate	Summarize	Relate	Design
	3	Describe	Express	Solve	Diagram	Select	Develop
	4	Define	Predict	Choose	Determine	Estimate	Draw
	5	Draw		Interpret	Categorize		Explain
TE	1	Recall	Distinguish	Review	Select	Compare	Specify
	2	Relate	Describe	Sketch	Experiment	Justify	Integrate
	3	Reproduce	Comprehend	Apply	Analyze	Describe	
	4	Find		Examine	Relate	Assess	
	5	Characterize		Schedule	Distinguish		
BE	1	Cite	Summarize	All the above	Compare,	Conclude	Synthesize
	2	Match	Estimate		Contrast	Measure	
	3	Tabulate	Contrast		Investigate	Summarize	

Map each question with corresponding Course Outcomes and learning levels

	1a	1b	1c	1d	1e	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b
CO Mapping	CO1	CO1	CO5	CO6	CO4	CO2	CO2	CO2	CO3	CO4	CO4	CO3	CO4	CO4	CO3
Marks	5	5	5	5	5	10	10	10	10	10	10	10	10	10	10
BT/LO Levels	U	U	AN	AN	C	A	R	E	R	A	C	A	AN	R	AN

Consider the table shown above, we have the inputs for end semester exams,1a represents question no.1's a part and further similarly. There are five learning levels Read, Understand, Apply, Evaluate, Create and Analyze which represents the nature of question. Marks represent the weightage of the questions. Weightage allotment depends on subject and weight allotted by university.

Map each PO/PSO with COs

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CO1	H	M		M	H	M		L	M	L		H		H	M	L
CO2			M		M		M				M		H		M	
CO3		H	L	H		H		M		M	L		H	L		
CO4	H	M		M	M	M		M	M					H	H	H
CO5			H		L				L				L		M	
CO6		M	M	H				H		M				L		

The above table represents the mapping of Cos and Pos with three levels High, Medium and low, if 60% or above students score more than 70% of marks level is High, between 50% and 60% the level is Medium and less the 50% is Low.

As shown above CO1 with PO1, H (High) is obtained.

The formula for attainment is shown below:

Learning level Attainment = 3 ∗ No of students in Level 3 + 2 ∗ No of students in Level 2 + 1 ∗ No of students in Level 1 ∑ Number of students on level i i=3 i=1 E.g. CO341 (U) Attainment = 3∗2+2∗2+1∗ 1 5

= 2.2

Here the number of students who attainted level 3 is calculated and multiplied by 3. Similarly number of students who attained level 2 and 1 are calculated and multiplied by 2 and 1 respectively and their average is calculated.

The value obtained like we got 2.2 above is the CO value of particular subject.

PO/PSO attainment w.r.t Cos

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
CO 1	2.2	2.2		2.2	2.2	2.2		2.2	2.2	2.2		2.2
CO 2			2.34		2.34		2.34		2.34		2.34	
CO 3		1.86	1.86	1.86		1.86		1.86		1.86	1.86	
CO 4	1.84	1.84		1.84	1.84	1.84		1.84	1.84	1.84		1.84
CO 5			2.05		2.05		2.05		2.05		2.05	
CO 6		2.27	2.27	2.27		2.27		2.27		2.27	2.27	
AVG	2.02	2.04	2.13	2.04	2.11	2.04	2.2	2.01	2.11	2.04	2.13	2.02

	PSO 1	PSO 2	PSO 3	PSO 4
CO 1		2.2	2.2	2.2
CO 2	2.34		2.34	
CO 3	1.86	1.86		
CO 4		1.84	1.84	1.84
CO 5	2.05		2.05	
CO 6	2.27	2.27		
AVG	2.13	2.04	2.11	2.02

III. CONCLUSION

With the help of Knowledge Management Tool for Calculating Attainments of Course and Program Outcome the institutes under National board of Accreditation (NBA) will be able to not only calculate the Course and Program Attainments but also to store the academic records of the students. Manual work load of the faculties of using Excel sheets and calculating values at each stage will be automated. The attainment values can be obtained for each node i.e. Institute level attainment, Department level attainment or Subject level attainment. This will help the institutes to analyze results easily within a few seconds rather than calculating the values at each level. This will save time of the faculties and the monotonous and tedious work will be reduced.

REFERENCES

[1] NBA India:
http://www.nbaind.org/En/1033-washington-accord.aspx
[2] NBA India:
https://www.nbaind.org/files/co-po-mapping-and-attainment-of-cos-and-pos.pdf
[3] Attainment of Course Outcomes: A Sample Empirical Case Study for incorporating
Outcome based education, Journal of Engineering Education Transformations,
Volume 30, No. 3, January 2017, ISSN 2349-2473, eISSN 2394-1707.
[4] Method for estimation of Attainment of Program outcome through Course outcome
for Outcome based Education, 2014 IEEE International Conference on MOOC, Innovation and Technology in Education (MITE).
[5] Karen M. Bursic, “Improving Student Attainment of ABET Outcomes Using Model-
Eliciting Activities (MEAs)”, American Society for Engineering Education, 2011.

Dietary Management System

Chinmay Pethe, Anagha Kiran Mundale, Priyanka Prakash Parab, Prof. Deven Ketkar
Information Technology
Finolex Academy of Management and Technology
Ratnagiri, India
chinmay.pethe123@gmail.com, anagha262@gmail.com, parab.priyanka18@gmail.com, deven.ketkar@famt.ac.in

***Abstract*— Many people suffer from three chronic diseases (diabetes, hypertension, cholesterol), and they often use search engine to collect related information. However, most of dietary information on the networks is not convenient for users to collect about the diet recommendations. In this paper, a diet recommendation system is suggested which can recommend a rational diet for users. The system uses BMI (Body Mass Index) as a guide prior to inference. According to the patient's health information, the system infers daily calories requirement.**

***Keywords*— BMI (Body Mass Index), chronic diseases, rational diet.**

I. INTRODUCTION

Diet is one of the supporting factors of life, plays a key role in maintaining health and prevention of diseases. In Modern Science, the balanced diet is one which includes a variety of foods in such quantities and proportions that the need for energy, proteins, vitamins, minerals, fats, carbohydrates and other nutrients for maintaining health. In Ayurveda the "balanced diet", which nourishes both the mental and physical built [1].

Individual's food choices depend on energy needs, nutrient needs and enjoyment. Family, friends and personal beliefs, including cultural and environmental considerations, also play a major role in people's food selection. Many genetic, environmental, behavioral and cultural factors can affect a person's health. Understanding family history of disease or risk factors, such as body weight and fat distribution, blood pressure and blood cholesterol, can help people make more informed decisions about how to improve health. Making good food choices is among the most pleasurable and effective ways of improving health.

II . ACKNOWLEDGEMENT

The author would like to thank HOD and IT department for their support and help for research and publishing the paper.

III. CONCLUSION

Balanced diet should help users to prevent eating disorder such as anorexia nervosa, diabetes, and obesity. This study shows this system can be used to represent recommended energy and nutrient intake adequately. Here we are utilizing BMI(Body Mass Index) to suggest best diet for the user. This system will be helpful for dieticians and general physicians for suggesting a suitable diet for patients using automatic approach.

REFERENCES

[1] Neeru Nathani, "An appraisal of the concept of Diet and Dietetics Ayurved", *Asian Journal of Modern and Ayurvedic Medical Science*, vol. 2, pp. 3-9, January 2013.

[2] Kurozumi, "FML-based Japanese diet assessment system. in Fuzzy Systems (FUZZ)", 2013 IEEE International Conference 2013.

Look Based Media Player

Vandana Munde, Atulkumar Mishra, Abhijit Chaudhary, Rahul Vishwakarma
Information Technology, Thakur College of Engineering & Technology
Mumbai, India

***Abstract*: In this project we are developing an advanced media player which plays and pauses the video by detecting the users face looking at screen or not. System monitors whether the user is looking at the screen or not using a web camera. If yes then doesn't interrupts the video and allows it to play. In case if the user is not looking at the or say the system couldn't detect the users face then it immediately stops the video. This system can also be used as a prototype for developing other more powerful systems. Moreover this system also has multiple other features such as controlling other aspects of a media player like volume up and volume down, play, pause and stops the video using hand gestures. We are trying to integrate both the systems together and make it one system as a whole.**

Keywords Media Player; Hand Gestures; Haar Cascade; Sensors.

I. INTRODUCTION

Usually when users are watching a video and someone calls, the user has to look somewhere else or go away from pc for some time so the user misses some part of the video. Later the user needs to drag back the video from where the user saw it. Well here is a solution to this problem. A look based media player that pauses itself when user is not looking at it. The player starts running again as soon as the user looks at it again. This is done using the camera or web camera on top of the computer. As long as the camera detects the users face looking at it, the media is played. The player pauses as soon as users face is not completely seen Even this system can be used as a prototype for developing other systems which can be used for security purposes. This system also provides the feature of controlling other functions of media players such as volume up, volume down, forward and backward using hand gestures.

II. AIM AND GOAL OF THE SYSTEM

The goal of our project is to create an advanced media player based face detection and hand gestures recognition. We have set the following objectives for our media player to achieve the goal:

1. The user interface of media player should efficient and user friendly.
2. The media player should be accurate in terms of result.
3. The media player pause the video as soon as the user face is not detected.
4. The hand gestures should be captured accurately and actions associated to them should performed perfectly.
5. We are trying to make this project with an aim for 10.

III. CONCEPT AND EVALUATION

A. Existing System:

Mostly existing systems use eye recognition. Due to which results aren't accurate. One of the previous systems tried to induce both face detection and hand gesture recognition
system in a single media player but the results weren't published.

B. Proposed System:

In this project we are using face detection and hand gestures recognition for controlling media player. Face detection is used for pausing and playing. Various hand gestures are used for controlling other functions of media player.

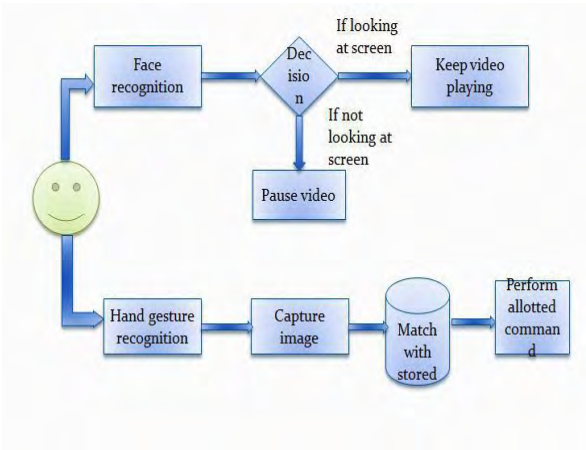


Fig 1: Work flow of proposed system

IV. IMPLEMENTATION METHOD

The proposed system uses Haar Cascade Classifier algorithm [1]
HAAR CASCADE CLASSIFIERS:

In the Viola–Jones object detection framework, the Haar-like features are therefore organized in something called a classifier cascade to form a strong learner or classifier. The key advantage of a Haar-like feature over most other features is its calculation speed. Haar-like features are digital image features used in object recognition. They owe their name to their intuitive similarity with Haar wavelets and were used in first real-time face detector. In the detection phase of the Viola–Jones object detection framework, a window of the target size is moved over the input image, and for each subsection of the image the Haar-like feature is calculated. This difference is then compared to a learned threshold that separates non-objects from objects. Because such

a Haar-like feature is only a weak learner or classifier (its detection quality is slightly better than random guessing) a large number of Haar-like features are necessary to describe an object with sufficient accuracy. In the Viola– Jones object detection framework, the Haar-like features are therefore organized in something called a classifier cascade to form a strong learner or classifier.

The key advantage of a Haar-like feature over most other features is its calculation speed. Due to the use of integral images, a Haar-like feature of any size can be calculated in constant time (approximately 60 microprocessor instructions for a 2-rectangle feature).

Open CV's algorithm is currently using the following Haar-like features which are the input to the basic classifiers:

1. Feature = w1 x RecSum(r1) + w2 x RecSum(r2)
2. Weights can be positive or negative .
3. Weights are directly proportional to the area.
4. Calculated at every point and scale.

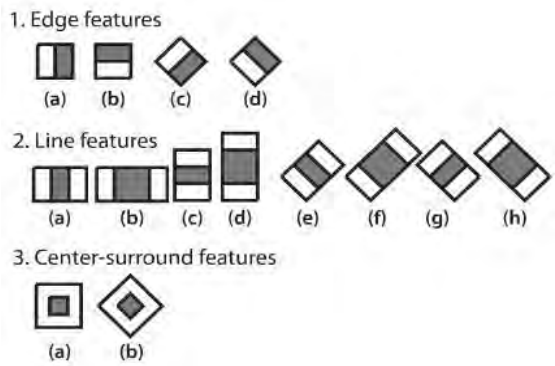


Fig 2: Haar Cascade Classifier

[2] CONTOUR EXTRACTION:

Contour tracing is a technique that is applied to digital images in order to extract their boundary. A digital image is a group of pixels on a square tessellation each having a certain value. This method will be dealing with bi-level images i.e. each pixel can have one of 2 possible values namely: 1, in which case we'll consider it a "black" pixel and it will be part of the pattern, OR 0, in which case we'll consider it a "white" pixel and it will be part of the background. Contour tracing is a technique that is applied to digital images in order to extract their boundary.

A digital image is a group of pixels on a square tessellation each having a certain value. This method will be dealing with bi-level images i.e. each pixel can have one of 2 possible values namely: 1, in which case we'll consider it a "black" pixel and it will be part of the pattern, OR 0, in which case we'll consider it a "white" pixel and it will be part of the background.

Well, the contour pixels are generally a small subset of the total number of pixels representing a pattern. Therefore, the amount of computation is greatly reduced when we run feature extracting algorithms on the contour instead of on the whole pattern. Since the contour shares a lot of features with the original pattern, the feature extraction process becomes much more efficient when performed on the contour rather on the original pattern.



Fig 3: Contour Extraction

V. RESULT

1. Accuracy: - The degree to which the result of a measurement, calculation, or specification conforms to the correct value or a standard for image processing.

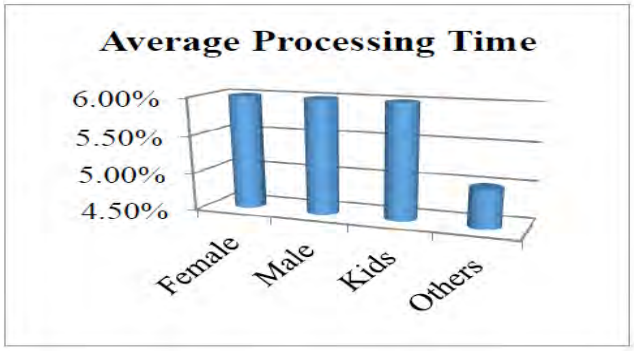
2. Processing Time:- The period during which one or more inputs are transformed into a finished output

AVG. PROCESSING TIME:

Table 1: Average Processing Time

INPUT	**AVG. PROCESSING TIME**
Female	6s
Male	6s
Kids	5s
Others	6s
Total Average	5s

BAR GRAPH:



ACCURACY:

(For Face Detection)

Formulae: - (TP+TN)/ (TN+TP+FN+FP)

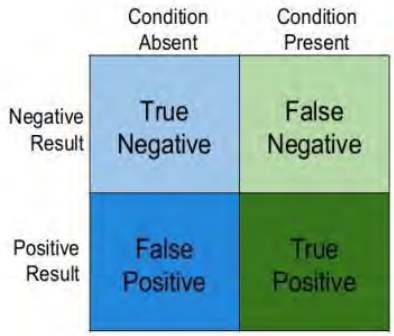


Table 2: Accuracy for Face Detection

INPUT	TP	TN	FP	FN
Female	10	0	0	10
Male	10	0	0	10
Kids	10	1	0	10
Others	10	0	0	10
Avg.	10	0.25	0	10

Total Accuracy: - (TP+TN)/ (TN+TP+FN+FP)

= (10+0.25)/ (10+0.25+0+10)

= 98%

ACCURACY:

(For Hand Gesture)

Table 2: Accuracy for Hand Gesture

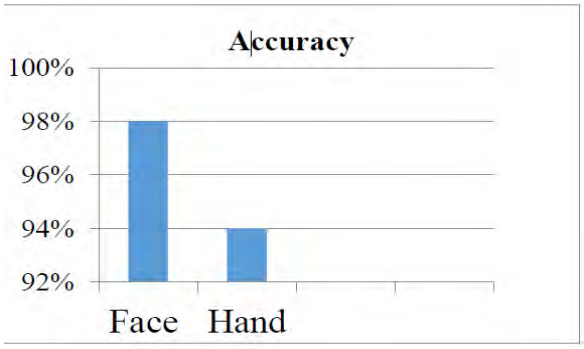
INPUT	TP	TN	FP	FN
Volume UP	9	10	0	0
Volume Down	10	9	1	0
Play	9	9	0	0
Pause	10	9	0	0
Avg.	9.5	9.25	1	0

Total Accuracy: - (TP+TN)/ (TN+TP+FN+FP)

= (9.5+9.25)/ (9.5+9.25+1+0)

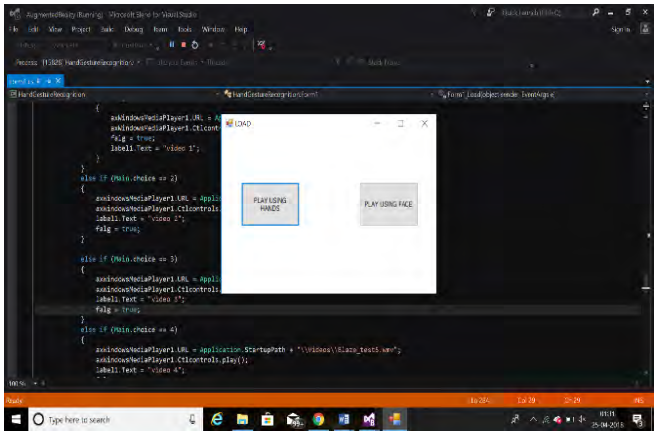
= 94%

BAR GRAPH:



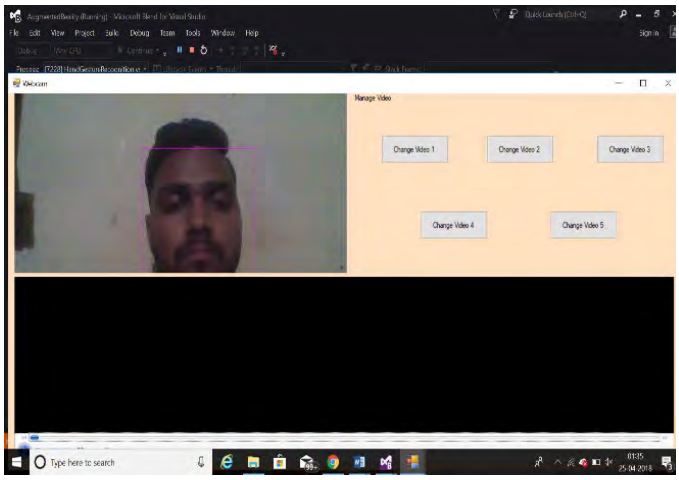
Main Graphical user interface:

This is main GUI of the developed system which involves the applications main interface through which the media player can be accessed and features like hand gesture and face recognition can be used. It involves “Play using face” and “Play using hands” option to use these two features in media player while watching video.



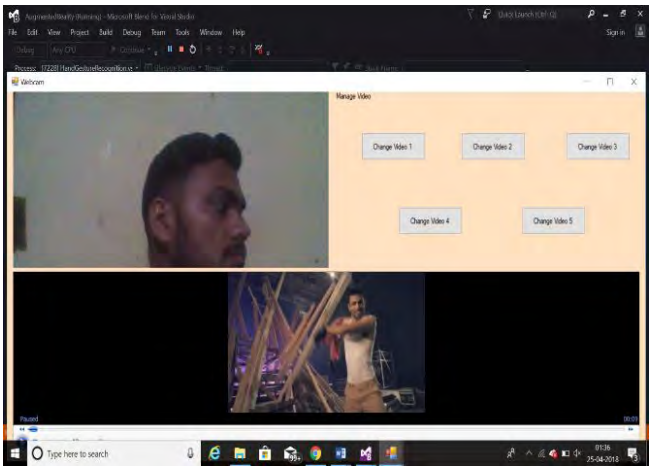
Playing the video:

Till the time the face is detected from camera then the video will be played. As the user is watching the video then with the help of harr cascade algorithm the face is detected and video is played. As we can see a violet box appearing around the face of the user is the sign of part of body being detected.



Pausing the video:

As we can see that the user has moved his face away from the camera and now the users face is not detected by the camera therefore the video is paused. We can also see that the violet line is not appearing around the face part of user that indicates that the face is not detected and the video is paused.



VI. Conclusion

The main concern of this project is to help the user get best experience of using a media player. We have tried to achieve this goal by automating the media player in a wide extent. We are doing this by using face recognition and hand gestures for controlling varied features of the media player such as pausing and starting the video again when the user isn't looking at the screen(for which face recognition is used), and controlling functions as forwarding , back-warding, volume up and volume down(for which hand gestures are used).

REFERENCES

[1]. [1]Sidharth Rautaray et Anupam Agrawal, "A Vision

[2]. Based Hand Gesture Interface for Controlling VLC Media Player", Journ. International Journal Of Computer Applications(0975-8887), vol. 10-No.7, November 2010.

[3]. [2]Harsha Jadhav, Sabiha Pathan, Neha Rokade et UmaAnnamalai"ControllingMultimedia

[4]. Applications Using Hand Gesture Recognition", Journ, International Research Journal Of Engineering and Technology (IRJET), vol. 02, PP 1200-1203, August-2014.

[5]. [3]Jyoti Rani et Kanwal Garg, "Emotion Detection Using Facial Expression", Journ. International Journal Of Advanced Research & Software Engineering, Vol. 4, PP-465467, April 2014

[6]. [4] Xiaoming Liu et Tsuhan Chen "Video-Based Face Recognition using Adaptive Hidden Markov Model",

[7]. Electrical and Computer Engineering, Caregie Mellon University, Pittsburgh, PA, 15213, U.S.A.

[8]. [5] D Jadhav et Prof L.M.R.J Lobo "Hand Gesture

[9]. Recognition System To Control Slide Show Navigation", Journ. International Journal Of Applications or Innovation in Engineering & Management (IJAIEM)", vol. 3, PP-283-286, January 2014.

[10]. [6]N. Krishna Chaitanya et R. Janardan Rao

[11]. "Controlling OF Windows Media Player Using Hand Recognition System", Journ. The International Journal Of Engineering And Science (IJES), vol. 3, PP 01-04, 2014.

[12]. [7]Viola and Jones, "Rapid object detection using a boosted cascade of simple features", Computer Vision and Pattern Recognition, 2001.

Cognitive Prediction and Estimation of Neurological Disease Using Machine Learning

Ganesh Mishra, Smit Supeda, Abhishal Yadav, Kaushal Yadav, Gayatri Dantal, Sunil Yadav

Dept of Information Technology, Shree L.R. Tiwari College Of Engineering(Mumbai University) Mumbai, India
ganeshmishra671@gmail.com, smitsupeda@gmail.com, abhishalyadav24@gmail.com,
kaushalyadav312@gmail.com, gayatridantal@gmail.com, sunil.yadav@slrtce.in

***Abstract*—According to the UN World Health Organization, nearly one in every six of the world's population is affected by a neurological disorder. Neurological disorders which involves both CNS and PNS are being visualized through patient CT or MRI scans; however, this study focused on the shared SNPs in different neurological disorders generated from pre-existing data. Mining discriminative patterns from graph data has attracted great interest in recent years. It has a wide variety of applications in disease diagnosis, neuroimaging, etc. Children with Global development delay (GDD) can also have Gradable Tumors and Neurological disease symptoms. Clinical patients who are diagnosed for EEG can have different neurological disease and also children with Different Development Rate get affected by this neurological disease. Prediction of Survival rate in Neurological disease such as Glioblastoma using various prediction algorithms such as linear regression, Non-Linear Regression, etc. We classified three classes and also conducted classification of neurological disorders for different combination.**

Keywords— MRI, Glioblastoma, GDD, Neurological Disease, ML.

I. INTRODUCTION

Number of patients who gets diagnosed with Glioblastoma(GBM) is increasing every year. Very few people tackle GBM for longer period. Although GBM is considered as incurable disease but with the help of medical treatment life expectancy can be increased. Estimating the size and growth rate of GBM is a very complicated task and requires a lot of years of experience. Once a person is diagnosed with GBM patients life reduces to few years to months. If medical treatment is provided to patient the number of living days increases by months to years depending upon the situation of patients. Even if treatment is fully completed it doesn't mean that person will not have to face GBM again. It may regenerate. So as we can see that it is almost incurable disease, but life expectancy of patient can be increased.

In this project we aim to provide a tool/software that helps the medical practitioners to easily estimate the growth rate and number of days a person can live.

GBM is a grade- IV level of astrocytoma. It is a most common type of brain cancer all over the world. Survival rate of GBM patients is very low. So, how we can increase the survival rate of GBM patients? Its is major concern for medical practitioner as well as medical students.

Our aim to increase the survival rate of GBM patient. To do so we first have to identify the cancer at early stages itself only. After identifying the cancer we need to find out what is the growth rate of GBM. At initial stages the growth rate is very rapid and as the cell size of cancerous cell increases growth rate decreases

.

II. LITERATURE REVIEW

As we all know cancer is becoming a very dangerous disease for human mankind. There is various type of cancer. Here we are going to talk about the brain cancer. In India every year 5 to 10 per 100,000 people get diagnosed with brain cancer. GBM is the most common and the deadliest. Almost 35% of cancer cases are GBM. GBM survival cases are very less. We cannot cure it but we can prolong the life period of patients suffering from GBM. GBM is a last grade of its phase. Survival rate is very low in case of GBM. The question here is that how can we increase the survivability of patients. In India every year 5 to 10 per 100,000 people gets diagnosed with CNS (Central Nervous System) Cancer. GBM is the most common among them. 59.5% are cases of GBM. GBM is the most dangerous and deadliest brain cancer. At, initial stages the growth rate of GBM cells is very rapid. As the cell size of GBM increases the growth rate is decreased. Surviving ratio is very low in case of GBM

III. DATASETS

The Dataset collected from The Cancer Imaging Archive (TCIA) which consist of MRI images of 40 Patients having diagnosed with Glioblastoma Multiforme(GBM) as shown in Table 1. This MRI images provided also estimates the average survival rate of each distinct patients.

TABLE I. DATASET DISCRIPTION

No of Images	Patient Cases
150	40

IV. METHODOLOGY

The proposed method consist of four crucial steps as described in the figure 1. Each MRI image given as input is applied to Gaussian filter to remove noise and distortions. Otsu segmentation extracts the affected area of Glioblastorma, the final step for determining the survivability rate the mathematical formulae is used.
The steps for determining survivability rate is given below.

A. Image processing.

Images produced by MRI machine have different sizes. In order to achieve better segmentation results we resize all the images to a fixed resolution of 560x560 pixels. For resizing the images we use bicubic interpolation, which uses 4x4 neighborhood pixel of known values.

B. Filtration

For better noise reduction and image cleansing we have used Gaussian filter method which will helps to cleanse the distorted method and provides clean image of 560x560 pixels.

Mathematically Gaussian filter for one dimensional impulse form can be represented as,

$$g(x) = \sqrt{\frac{a}{\pi}} \cdot e^{-a \cdot x^2}$$

C. Segmentation

This step involves image segmentation which is done by using Otsu segmentation method.Otsu is used to automatically perform clustering-based image thresholding, or, the reduction of a graylevel image to a binary image.

The main steps of Otsu's algorithm are described as follow:

- In the first step of Otsu's the pixels are divided into two clusters. Compute mean value for every pixels cluster.
- When mean value for every cluster is determined, find the square of difference for each cluster respectively.
- Number of pixels in one cluster multiplied by number of pixels in other cluster.

D. Prediction and Estimation

This is one of the major step for he proposed method, in this step we use an activation function to determine the survivability rate. In this step the data extracted from MRI image and the trained data is used to determine the survivability rate.

REFERENCES

[1] Eman M, Mohammed E, Rashid A. Brain Tumor Segmentation based on a hybrid clustering technique. Egyptian Informatics Journal 2015 Production and hosting by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University; 71-81.

[2] Suman S. Image Denoising Using New Adaptive Based Median Filter. Signal & Image processing: An International Journal (SIPIJ) Vol.5, No.4, August 2014

[3] .[3] Geethu M, M. Monica S. MRI Based medical image analysis: Survey on brain tumor grade classification. Published by Elsevier Biomedical Signal Processing and Control 39 (2018) 139–161.

[4] Mohammadreza S, Guang Y, Tryphon L, Nigel A, Timothy J, Thommas B, Franklyn H Xujiong Y. Supervised Learning based Multimodal MRI Brain Tumour Segmentation using Texture Features from Supervoxels. Accepted Manual. To appear in: Computer Methods and Programs in Biomedicine.

[5] Sanjeev K, Chetna D, Sunila G. Classification of Brain MRI Tumor Images: A Hybrid Approach. Published by Elsevier B.V. Peer-review under responsibility of the scientific committee of the 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.

[6] Javeria A, Muhammad S, Mussarat Y, Steven F, A Distinctive Approach to Brain Tumor Detection and Classification using MRI, Pattern Recognition Letters (2017), doi: 10.1016/j.patrec.2017.10.036.

[7] Heba M, El-Sayed A. E, El-Sayed M. E, Abdel-Badeeh

[8] M. S. Classification using deep learning neural networks for brain tumors. Future Computing and Informatics Journal xx (2017) 1-4.

[9] Ali I, Cem D, Melike S. Review of MRI-based brain tumor image segmentation using deep learning methods. 12th International Conference on Application of Fuzzy Systems and References Detection and Classification of Brain Tumor Using Artificial Intelligence Soft Computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria. Published by Elsevier B.V.

[10] Rajesh C, Dr. Kosalani R. Tumor Detection in Brain using Genetic Algorithm. 7th International Conference on Communication, Computing and Virtualization 2016. Published by Elsevier B.V.

[11] Ivan C, Iker G. MRI Segmentation Fusion for Brain Tumor Detection, Information Fusion (2016), doi: 10.1016/j.inffus.2016.10.003.

[12] Swapnil T, Amit P, Ankita K. Detection of Brain Tumor from MRI Images by using Segmentation and SVM. 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (WCFTR'16).

[13] Sudipta R, Debnath B, Samir K, Tai-Hoon KIM. An

Improved brain MR image binarization method as a preprocessing for abnormality detection and features extraction. Research Article. Accepted April 1, 2016.

[14] Shaswati R, Pradipta M. A Simple Skull Stripping Algorithm for Brain MRI. 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)

[15] BOOKS

[16] Navigating Life with a Brain Tumor. Lynne P. Taylor, By Alyx B. Porter Umphrey and With Diane Richard.

[17] Brain Tumors an Encyclopedia Approach 3rd Edition. By Andrew H. Kaye, Edward R. Laws. Jr.

Estimation and Simulation of Distribution of Liability Insurance

Parvaneh Rafieisangari Chongqi Zhang

School of Economics and Statistics
Guangzhou University
Guangzhou, China
parvaneh.rafiei@yahoo.com cqzhang@gzhu.edu.cn

***Abstract*— Liability insurance, is part of the public insurancesystem, which directly affects the social relationships of individuals, recognizes the rights and responsibilities of individualstowards each other and is critical to the security of businesses and professional activities.In this paper we use a non-parametric method to the estimation of distribution of liability insurance based on Maximum Entropy (ME) and then simulation of this distribution. Sampling the ME distribution is vital. Given the difficulties in carrying out exact simulation, for the approximate simulation of the ME distribution we use Adaptive Importance Sampling (AIS) algorithm.**

***Keywords*—liability insurance,maximum entropy,adaptive importance sampling.**

I. INTRODUCTION

The principle of maximum entropy was first expounded by [1] in information theory. Based on this principle, among all probability distribution functions that applies to certain constraints, the distribution must be selected that have maximum entropy. This principle taking into account the possible conditions, actually leads to the most uniform to distribution.

Exact simulation of the ME distribution is not explicit. For approximate simulation of the ME density we use Adaptive Importance Sampling (AIS). The AIS-based simulation method is fully compatible with ME, because simulation can only be done by knowing the density function. AIS algorithm based on mixtures of normal densities was first introduced by [2]. In this paper, we use AIS method proposed by [3] is based on mixtures of lognormal density.

Liability insurance, is part of the public insurance system, which directly affects the social relationships of individuals, recognizes the rights and responsibilities of individuals towards each other and is critical to the security of businesses and professional activities. Although liability insurance is provided to third parties in relation to the risks arising from the activity and, but the insurer will receive a calm and assured reassurance at the time of its activity, in the light of compensation from the insurer (insurance company). The distribution of liability insurance is not usually belonging to any parametric family; therefore, ME method can be very useful.

In this paper, section 2 reviews the ME method and AIS algorithm. Section 3 we using real data of the liability insurance from one of the insurance companies of Iran, which was collected during four years (2008-2012), will estimate the distribution of data and will simulate this distribution. Finally, Section 4 concludes.

II. ME METHOD AND AIS ALGORITHM

A. ME Method

Suppose the problem is determined h (density) under the constraints

$$\begin{cases} h(x) \geq 0 & \forall x \in S \\ \int_S h(x)dx = 1 & \\ \int_S h(x) f_i(x)dx = \upsilon_i \quad i = 1, \dots, m & \end{cases}$$

so that the function $G(x)$ is maximized below

$$\max_f \left\{ G(x) = \int_S -h(x)\, log\, h(x)\, dx \right\} (1)$$

where h is a density, G is the Shannon entropy relevant to h, S is support of h, m is the number of constraints and f_is are the characterizing moments that in most cases are

$$f_i(x) = x^i$$

or

$$f_i(x) = (\log x)^i$$

and for the i-th characterizing moment υ is the sample counterpart.

The ME density, takes the form

$$h(x) = exp\left\{ -\sum_{i=1}^{m} \gamma_i f_i(x) \right\} (2)$$

Equation)1(cannot be solved analytically for $m \geq 2$, [4] has developed a generalization of the Newton-Raphson algorithm for find a solution, the constraints from the lowest to the highest moment at a time impose and update the ME density. For choice the optimal value of m there are following ways.

(a(The maximum log likelihood is equal to

$$-n \sum_{i=0}^{m} \hat{\gamma}_i^{(m)} \hat{\upsilon}_i$$

where the $\hat{\gamma}_i^{(m)}$s are the parameters of ME(m) density and n is the sample size. For

$$H_0: m = k \quad (k = 1,2, \dots)$$
$$H_a: m = k + 1$$

the test statistics is given by (log-likelihood ratio)

$$llr = -2n \left(\sum_{i=0}^{k+1} \hat{\gamma}_i^{(k+1)} \hat{\upsilon}_i - \sum_{i=0}^{k} \hat{\gamma}_i^{(k)} \hat{\upsilon}_i \right)$$

its asymptotic distribution is χ_1^2. For selection model:

1. estimate the ME(k) density with $k = 1,2, \dots$;
2. do the llr test for each value of k;

3. stop test at the first value of s (say k_0) such that the hypothesis $k = k_0$ cannot be rejected and deduce that the optimal value of m (say m^*) is equal to k_0.

(b) The previous solution does not apply to calculating the estimation of a model with a large number of parameters. The solution is computing the Bayesian (AIC) or Akaike (BIC) Information Criterion. It stop at the value m^* such that one of the following conditions to be established:

1. the llr test cannot reject the hypothesis $m = m^*$;
2. the numerical value of $BIC(m^*)$ [or $AIC(m^*)$] is smaller than the numerical value of $BIC(m^* + 1)$ [or $AIC(m^* + 1)$].

In general, density (2) cannot be integrated analytically, so the MEdistribution function $H(x)$cannot be written as a closed form.Due to these problems, we use Adaptive Importance Sampling (AIS) method that developed by [3].

B. Algorithm AIS

- For $t = 1$:
  - *Let* $f^{(1)} = \sum_{d=1}^{D} \pi_d f(x; \mu_d, \sigma_d^2)$ where $f(x; \mu_d, \sigma_d^2) \sim Logn(\mu_d, \sigma_d^2)$;
  - Simulate $X_1^{(1)} = \left(x_1^{(1)}, \dots, x_n^{(1)}\right)'$ independently from $f^{(1)}$;
  - Compute $w_1^{(1)}, \dots, w_n^{(1)}$ (the importance weights), where
$$w_j^{(1)} = h\left(x_j^{(1)}\right) / f^{(1)}\left(x_j^{(1)}\right) = h\left(x_j^{(1)}\right) \Big/ \sum_{d=1}^{D} \pi_d^{(1)} f\left(x_j^{(1)}; \mu_d^{(1)}, \sigma_d^{2(1)}\right) \quad (j = 1, \dots, n).$$
- For $t > 1$:
  - Update the importance function to $f^{(t+1)}$ according to the minimum Cross-Entropy criterion, using of the previous weighted sample $\left(x_1^{(t)}, w_1^{(t)}\right), \dots, \left(x_n^{(t)}, w_n^{(t)}\right)$,
$$w_j^{(t)} = \frac{h\left(x_j^{(t)}\right)}{\sum_{d=1}^{D} \pi_d^{(t)} f\left(x_j^{(t)}; \mu_d^{(t)}, \sigma_d^{2(t)}\right)} \quad (j = 1, \dots, n)$$
and the normalized weights are
$$\bar{w}_j^{(t)} = w_j^{(t)} / \sum_{j=1}^{n} w_j^{(t)};$$
  - Simulate $X_1^{(t+1)} = \left(x_1^{(t+1)}, \dots, x_n^{(t+1)}\right)'$ independently from $f^{(t+1)}$;
  - Compute $w_1^{(t+1)}, \dots, w_n^{(t+1)}$ (the importance weights).

According to [2], for any $i = 1, \dots, D$

$$\pi_i^{(t+1)} = \sum_{j=1}^{n} \bar{w}_j^{(t)} \tau_i\left(y_j^{(t)}; \pi^{(t)}; \mu^{(t)}; \sigma^{2(t)}\right)$$

$$\mu_i^{(t+1)} = \frac{\sum_{j=1}^{n} \bar{w}_j^{(t)} y_j^{(t)} \tau_i\left(y_j^{(t)}; \pi^{(t)}; \mu^{(t)}; \sigma^{2(t)}\right)}{\pi_i^{(t+1)}}$$

$$\sigma_i^{2(t+1)} = \frac{\sum_{j=1}^{n} \bar{w}_j^{(t)} \left(y_j^{(t)} - \mu_i^{(t+1)}\right)\left(y_j^{(t)} - \mu_i^{(t+1)}\right)' \tau_i\left(y_j^{(t)}; \pi^{(t)}; \mu^{(t)}; \sigma^{2(t)}\right)}{\pi_i^{(t+1)}}$$

where

$$\tau_i(y; \pi; \mu; \sigma^2) = \frac{\pi_i f\left(x; \mu_i, \sigma_i^2\right)}{\sum_{d=1}^{D} \pi_d f(x; \mu_d, \sigma_d^2)}$$

and $y = log(x)$.

According to [5], a practical measure of the convergence of the algorithm is the normalized perplexity $perp_n = exp\{G_n\}$, where $G_n = -\sum_{j=1}^{n} \bar{w}_j \log \bar{w}_j$ is the Shannon entropy of the normalized weights $\bar{w}_j$.So the stop criterion is$perp_n = exp\{G_n\}$.

On average, the normalized perplexity is incremental in every repetition, so it is logical that the stopping of the algorithm occurs when this criterion cannot be increased (hence the entropy cannot be further reduced). Therefore, when $perp_n$ does not change significantly, or when the algorithm reaches this number for the first time by taking a fixed number for this criterion, the algorithm stops, which is a guarantee for the desired level.

When simulating, the goal is answer to the question of whether simulated data can be produced by the ME density. To answer this question, the difference between the simulation data and the probability distribution function of the ME density should be computed by the Kolmogorov-Smirnov test.

III. Estimate and Simulate the Distribution of Liability Insurance Dataset

In this part, we use a liability insurance dataset from one of the insurance companies of Iran which has collected 608 items during the four years (2008-2012).

The histogram of liability insurance dataset is drawn in Figure1. In this figure, a large number of small losses and a small number of large losses has clearly caused the skewness.

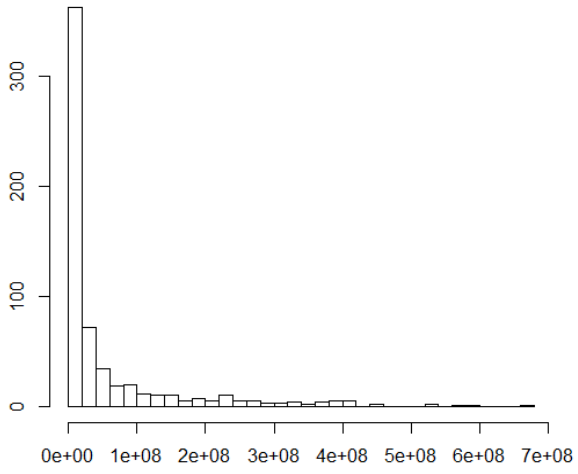


Fig. 1 The histogram of liability insurance dataset.

The analysis consists: (1) Estimate the ME(m) density with $m = 1,2, \dots$, (2) Using the AIS algorithm, simulate the estimated ME density , (3) Compute the Kolmogorov–Smirnov (KS) distance.

A. Estimate the ME density

In Table1, we can see Averages of estimated parameters of the ME(m) density with $m = 1,2, \dots ,6$.

Table 1.Averages of estimated parameters of the ME(m) density.

	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$
λ_0	2.60	2.498	2.392	2.362	2.365	2.348
λ_1	0.861	0.332	0.257	0.279	0.124	0.126
λ_2		0.122	0.203	0.219	0.291	0.301
λ_3			-0.011	-0.021	0.003	-0.001

λ_4	-0.001	-0.012	-0.012
λ_5		0.001	0.001
λ_6			-0.000

According to Table 2, we can see that the results are not the same: according to the numerical value of the llr test, m = 2 is accepted, and the BIC and AIC criteria both indicate m = 3.According to the selection steps of the model in Section 2, we deduce that $m^* = 3$.

Table 2.Values of the tests for the Liability Insurance dataset.

	llr	AIC	BIC
m=1		5871.159	5879.979
m=2	5597.343	5603.343	5616.574
m=3	9.336	5596.007	5613.648
m=4	0.415	5597.592	5619.642
m=5	5.19	5594.402	5620.863
m=6	0.026	5596.376	5627247

In Figure 2, for the data smaller than 100, can be seen the estimated ME(2), ME(3), ME(4) and ME(5) densities superimposed on the normalized histogram of the data. According to Figure 3, the difference in the right tail of the densities are very small.

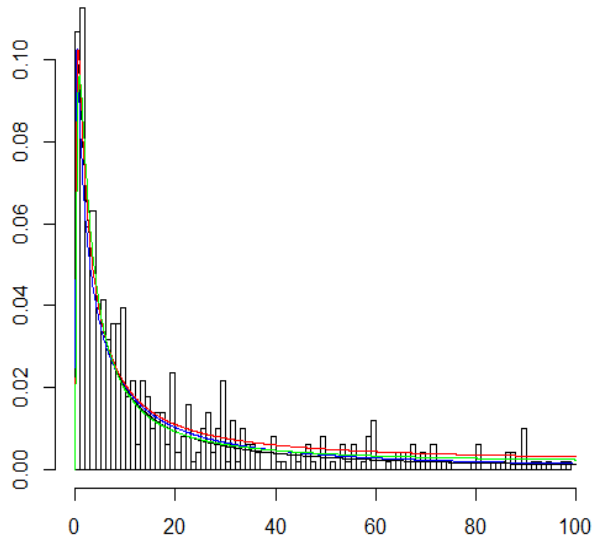


Fig. 2.The histogram of the Liability Insurance datasmaller than 100 with the estimated ME(2) (Blue), ME(3) (Black), ME(4) (Red) and ME(5) (Green) densities.

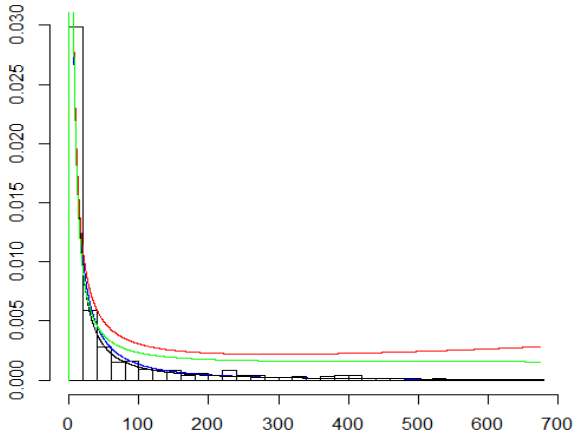


Fig. 3.The histogram of the Liability Insurance datawith the estimated ME(2) (Blue), ME(3) (Black), ME(4) (Red) and ME(5) (Green) densities.

B. Simulate the estimated ME(3) density using AISAlgorithm

We implement the algorithm with D = 6 and the following values:

$$\pi = \begin{pmatrix} 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \end{pmatrix} \quad \mu = \begin{pmatrix} 6.5 \\ 5 \\ 1 \\ 4 \\ 2 \\ 7 \end{pmatrix} \sigma^2 = \begin{pmatrix} 0.2 \\ 0.15 \\ 1 \\ 0.5 \\ 1 \\ 0.3 \end{pmatrix}$$

[7] showed that this algorithm even with different initialization, ever converges to the same parameter values.

We simulate $n = 10{,}000$ observations. The normalized perplexity at the 14-th iteration is 0.9984, which is a large value and close to 1, that guarantees a high level of precision.

Figure 4 and 5 shows that the estimated ME(3) density superimposed on the normalized histogram of the simulated observations is fit excellent.

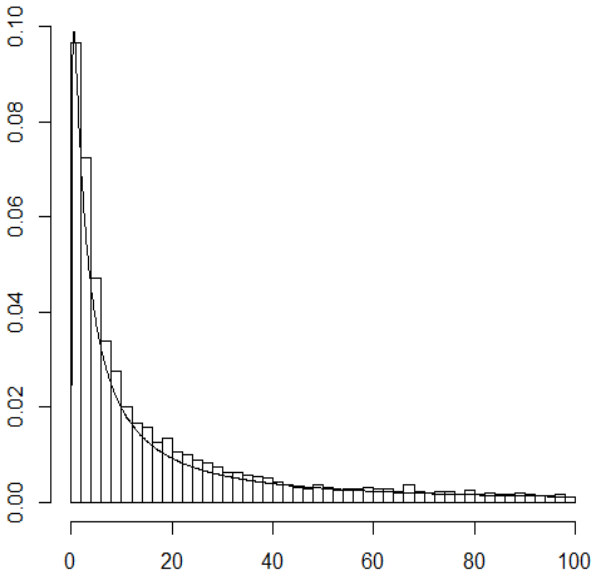


Fig. 4.The estimated ME(3) density superimposed on the normalized histogram of the simulated observations smaller than 100.

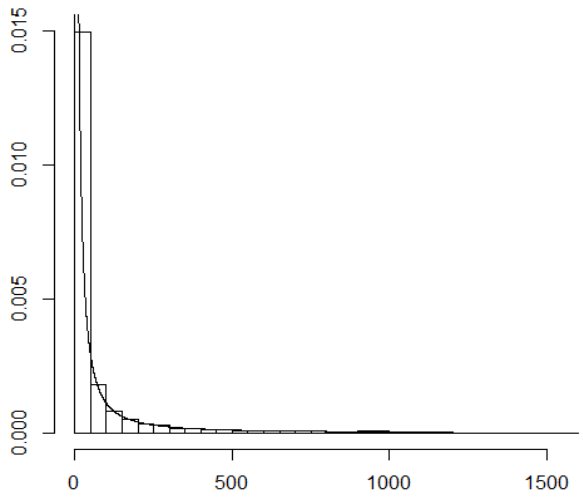


Fig. 5. The estimated ME(3) density superimposed on the normalized histogram of thesimulated observations.

C. *Compute the KS Test*

The Kolmogorov–Smirnov test is

$$KS = \max|H_n(x) - H(x)|$$

where H_n is the empirical probability distribution function of the simulated data and $H(x)$ is the theoretical probability distribution function of the estimated ME(3) density.

$KS = 0.0122$, and p-value is 0.215.

Therefore, it can be concluded that the process of producing data from simulated observations is the estimated ME(3) density.

IV. Conclusions

In this paper using the proposed method by [5] estimation and Simulation of the distribution of liability insurance. Commercial liability is an important segment for the insurance industry, liability insurance is 10% of global non-life premiums.Liability insurance is one of the insurance industries that directly affects the social relationships of individuals, recognizes individuals' rights and responsibilities, and ensures the professional security of businesses and activities.

This method obtains the most appropriate distribution for data, especially when there is no theoretical reason to adopt a parametric distribution. This is also a test for choice between two or more parametric models.

References

[1] Jaynes, E.T. Information theory and statistical mechanics. Phys. Rev. 1957, 106, pp.620–630.

[2] Cappé O.; Guillin, A.; Marin, J.; Robert, C. Population Monte Carlo. J. Comput. Graph. Stat. 2004, 13, pp.907–929.

[3] Bee, M. A maximum entropy approach to loss distribution analysis. J.Entropy. 2013, 15(3), pp.1100–1117.

[4] Wu, X. Calculation of maximum entropy densities with application to income distribution. J. Econom. 2003, 115, 347–354.

[5] Cappé O.; Douc, R.; Guillin, A.; Robert, C. Adaptive importance sampling in general mixture classes. Stat. Comput. 2008, 18, 447–459.

[6] Stutzer, M. A simple nonparametric approach to derivative security valuation. J. Financ. 1996, 51, 1633–1652.

[7] McLachlan, G.; Krishnan, T. The EM Algorithm and Extensions, 2nd ed.; Wiley: New York, USA, 2008.

Capstone Project-Big Data

Anurag Singh, Nishant Singh, PranaySurana,
Thakur College of Engineering &Technology,Mumbai,India

***Abstract*- Big data is concept of analyzing large number of data. Build an ecosystem using big data for data sets available online to analyze and do research analysis on it. Simulate big data i.e. generated from large number of users. Use tools like Gephi to analyze data and display it graphs and node-edge diagrams. Use mongo dB as database for unstructured data. Use Splunk searching, monitoring, and analyzing machine-generated big data. Practice these tools and build an ecosystem for Big Datasets. Collect different datasets initially like twitter database and any news databases online and analyze it follow different analyzing steps for acquiring, exploring, preparing, analyzing, and reporting.**

Keywords: Hadoop, Datasets, Spelunk, KNIME etc

I. INTRODUCTION

In this culminating project, we will build a big data ecosystem using tools and methods form the earlier courses in this specialization. We will analyze a data set simulating big data generated from a large number of users who are playing our imaginary game "Catch the Pink Flamingo". During the Capstone Project, we will walk through the typical big data science steps for acquiring, exploring, preparing, analyzing, and reporting. In the first two weeks, we will introduce to the data set and guide through some exploratory analysis using tools such as Spelunk and Open Office. Then we will move into more challenging big data problems requiring the more advanced tools you have learned including KNIME, Spark's ML Lib and Gephi. Finally, we will show how to bring it all together to create engaging and compelling reports and slide presentations. As a result of our collaboration with Spelunk, a software company focus on analyzing machine-generated big data.

II. LITERATURE SURVEY

Spelunk: Useful to show Analysis of Data and thus gain Business Intelligence.
KNIME: To integrates various components for data mining through its modular data pipelining concept.
Spark: To access diverse data sources including HDFS, Cassandra, H Base, and S3
OS: Cent OS Hardware:
Minimum 250GB hard disk Pentium IV or higher
"Report of dataset analysis" this project is mainly designed to reduce manual work. In the Analysis report of user generates the report according to user's profile, score and other details.
The result analysis is generating the graph according to user's dataset.

Fast access to database User friendly interface Scope of the Project
Integration – Integrating of various proposed modules for analyzing data
Table 1: Survey with gap Analysis

III. RELATED WORK

The systematic flow or the sequential approach to the flow of the project is given by the following flowchart.

PHASE 1: (Planning, Analysis, Design, Coding)
Planning – Conceptualizing and designing the data model for the game filter tools of Big Data to be used.
Analysis – Critical study, analysis and review of feasibility for proposed solution. Choosing the right type of components such Splunk Analysis tool .
Design – Design an ecosystem for data set of game.
Coding – Creating various modules of an IT based solution product for analyzing the data of the guests.
PHASE 2: (Integration, Testing, Deployment)

Testing –Exhaustive Testing using test cases to check the integration and fixing bugs for proposed solution. Perform alpha testing after completion of prototype.
Deployment – Give the completed ecosystem for dataset and beta testing.

IV. METHODOLOGY

It is inclined towards automating the whole storage system and giving it touch of big data. It will result in reduced data spoilage and data managers will be able to analyze data properly Research of Big Data tools like KNIME, Gephi will provide comfort for companies like graphics companies, entertainment group where huge amount of data is generated.
Companies will be likely to adopt this manually build ecosystem.

Algorithm:

Input	Set of data	Bus, Car, bus, car, train, car, bus, car, train, bus, TRAIN,BUS, buS, caR, CAR, car, BUS, TRAIN
Output	Convert into another set of data (Key,Value)	(Bus,1), (Car,1), (bus,1), (car,1), (train,1), (car,1), (bus,1), (car,1), (train,1), (bus,1), (TRAIN,1),(BUS,1) , (buS,1), (caR,1), (CAR,1), (car,1), (BUS,1), (TRAIN,1)

V. TOOLS & TECHNOLOGIES

Splunk: Useful to show Analysis of Data and thus gain Business Intelligence.

KNIME: To integrates various components for datamining through its modular data pipelining concept.
Spark: To access diverse data sources including HDFS, Cassandra, HBase, and S3
OS: Cent OS Hardware:
Minimum 250GB hard disk Pentium IV or higher
“Report of dataset analysis” this project is mainly designed to reduce manual work.In the Analysis report of user generates the report according to user’s profile, score and other details.
The result analysis is generating the graph according to user’s dataset.
Fast access to database User friendly interface Scope of the Project

VI. CONCLUSION

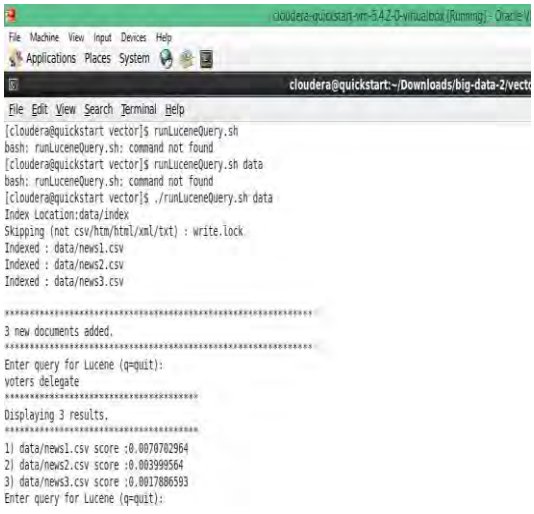
In terms of performance Big Data offers highly reliable data collection even in huge storage, Big Data technology can provide new capabilities as well as an efficient method to collect, manage, disseminate, store, and analyze dat.

- In terms of data management, this method collects all user entries who plays the game.

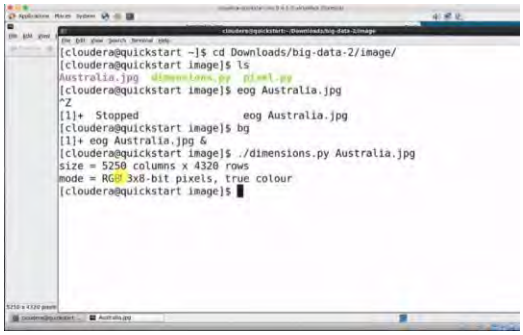
- Manual entry work will be automated, so time, cost & man power will be reduced.

•Analysis will be performed on data acquired so as to manage the flow of users in different profiles effectively, reduce the waiting time and serve the visitors in a more convenient manner. Thus, the proposed solution will be implemented

VII. RESULTS







REFERENCES

[1]. C. Romero S. Ventura P. G. Espejo and C.Hervás "Data mining algorithms to classify students ".In EDM 2008 pp.8-17. Y. Zhang S. Oussena T. Clark and H. Kim "Use data mining to improve student retention in higher education- A case study " In ICEIS (1) 2010 pp. 190- 197.

[2]. B.Namratha, Niteesha Sharma “Educational Data Mining – Applications and Techniques”.In IJLTET Vol 7 issue 2 July 2016.

[3]. Z. Erdogan and M. Timor "A Data Mining Application in a Student Database " in Journal of Aeronautics and Space Technologies 2005 vol. 2No. 2 2005 pp. 53-57

[4]. Li Ma, Lei Gu, Jin Wang, “Research and Development of Mobile Application forAndroid Platform” , International Journal of Multimedia and Ubiquitous Engineering Volume 9, No.4(2014)

[5]. Ganesh R. Vishwakarma , “Android College Management System”

[6]. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5, Issue 4, April2016.

[7]. https://www.splunk.com/web_assets/pdfs/ sec ure/Splunk_and_MapReduce.pdf

[8]. https://www.splunk.com/pdfs/soluti on- guides/splunk-guide-to-operational- intelligence.pdf

Study of Google Analytics and Open Web Analytics on Analyzing the Web Traffic

Mrs. Veena Kulkarni , Mrs. Lydia Suganya
Department of Computer Engineering Thakur College of Engineering and Technology Mumbai,India
veena.kulkarni@thakureducation.org , lydia.jacob@gmail.com

***Abstract*—Now-a–days the common gate way to any business for any customer is the website of the business/company. The more the usability of the website, the more business comes to the company. A poorly developed and less visited website can pose a problem to the corporation. The web traffic to any website of the corporation can be analyzed, thereby helps the developer to improve the pages in the website. This helps the corporations to place their marketing campaigns strategically on their web pages.**

The data about the most visited pages and the path through which customers frequently traverse in the website can be easily obtained using different web analytics tools. Web analytics is gathering the information of the web traffic. This helps the businesses to analyze the customer behavior while they browse their website. Thus it can help improve the website

***Keywords*—Web Analytics, behavior, website, web traffic, usability**

I. INTRODUCTION

Web analytics is a method to gather the data about user's online behavior whenever a user browses any website. After gathering the web server logs, the useful information can be collected like number of visitors on the site, day wise, number of sessions. Number of resources visited etc.

Why web analytics is required?

The website of a company depicts the entire company in today's era. The customers have access to the competitor website which is just a click away. It becomes necessary to increase the stickiness of the website in order to keep the customer glued as well as provide maximum information related to discounts and promotions. The more the stickiness likely are the chances that the customer becomes a buyer from the website/company. Also there are some decisions that have to take into consideration the behavior of existing customers.

Web Analytics Process:

Web analytics is not only a technology for reporting, but a recurring process of website optimization which, among other things, measures costs, identifies the most profitable user behavior and optimizes a website by improving performance and profitability. Waisberg & Kaushik [1] identified the following steps in a web analytics process:

- Objective (goal) determination
- KPIs definition
- Data collection
- Data analysis
- Change implementation

The goal of measuring the effectiveness of application or website is to support business goals. Targeting is a fundamental component of any analytical measurement plan. Having accurately configured goals allows Web Analytics to provide important information to better understand if users have completed the actions and want them to accomplish. It is not possible to evaluate the effectiveness of online businesses without this information. Providing this information is important, as they lead to selecting the appropriate metric to achieve a certain return on investment (ROI) for a website. Measuring the success of objectives is accomplished by defining key performance indicators (KPIs) that show the particular progress or detect lagging in achieving goals. Entrepreneurs must choose the right KPIs. Each KPI is expected to correlate with a specific action and match the criteria/attributes of timelessness, simplicity, relevance and usefulness. The example [1] of one qualitatively defined KPI that covers all of these attributes is the bounce rate defined as the percentage of visitors who leave a website after the first page visit. This metric is simple, i.e. easy to understand and explain. It is relevant because it identifies sites with a lower content, technically or in terms of quality. This KPI is also timely since is readily available in all web analytics tools. Finally, the observed KPI offers instant usefulness given that a decision maker can act immediately and focus on pages with a high bounce rate. The next activity is accurately collecting data and saving it in a database for subsequent data analyses. Data analysis includes observing and transforming previously collected data in order to discover useful information that supports future decisions. During this activity, numerous associations, patterns or trends that exist in the data set are revealed. Implementing change is the final activity in the process. In other words, collected and analyzed data provide new information and support change implementation. Otherwise, all the previous activities are unsupported. The scheme of the web analytics process is shown in Fig 1.



Fig. 1. Activities in the web analytics process, Waisberg and Kaushik [1]

Web Analytic Tools

A variety of web analytics tools have been developed and are available on the market that aim to obtain quantitative and qualitative data as a basis for the decision-making process, The author [2] has classified web analytics tools into five categories:

1. Traditional web analytics tools that mostly relied on clickstream data obtained by the visitors themselves, competitors and data from the company's internal sources. Clickstream data generally corresponds to the question "what happens on websites" or "visitor behavior while browsing the website" and "how many conversions have been achieved on the website".
2. Web analytics tools that track performance on social networks
3. Web analytics tools for gathering visitor feedback aim to answer the question "reasons for the visitor behaving or not behaving in a certain way". 4. Web analytics tools for mobile websites with a growing importance in line with an increase in website turnover caused by the use of mobile devices. These tools provide insight into visitor behavior on websites accessed via mobile devices similar to traditional web analytics tools, and are necessary for achieving compatibility with mobile devices.
5. Web analytics tools for experimenting, testing and find optimal technical or design solutions that should improve visitor satisfaction.

How Web Analytic tool works?

There are two methods by which the Web Analytic tool gathers information

1. Web server log file analysis – In this method the log files are accessed and the data is gathered
2. Page tagging- In this method a java code / Snippet is inserted in the page and the data is collected.

Web Metrics

Web Metrics is a generic term for the many types of measurements that can be made about a website and its visitors [6]. Some of the key we metrics are listed below:

- Visitors: Number of people that visited your site during a specific time frame (e.g., yesterday, last week, last month).
- Referrals:Referrals track users as they click on links in search engines, on other blogs, and other websites to your web site.
- Bounce Rate: A "bounce" is when someone visits your site and immediately clicks the back button or closes their browser tab.
- Exit Pages
- Conversion Rate
- Top 10 pages

II.GOOGLE ANALYTICS

Google Analytics is an open source tool that generate analysis of the various activities on the website like no of visitors on the page, the most visited page, the number of new visitors to your website, the stickiness of the page etc.
It can also gather data from mobile devices from the iOS and Android Apps. For a Search engine it analyses the amount of visitors to the site for a particular keyword search.

Google Analytics is the mostly used for site traffic tracking and reporting. It provides information that helps in decision making about how to improve the online business. It gives the information to the business owner about their customers and what they want. Google Analytics studies the site traffic, provides real-time statistics, and analyzes user interactions with websites.
Google analytics provides data that is generally designed for advertisement, marketing as well as the type of the traffic the website receives and usefulness of the owners marketing campaigns. It tracks the customers from all other sites and the number of customers converted into real buying customers or members from each page. Google Analytics can provide feedback for subsequent marketing campaigns.
Google analytics needs the user to have a Gmail account or to register with Google and give the name of website to be tracked .It provides a code snippet, which is to be placed on every page of the web application to be tracked. When the website is loaded, this code is activated and captures information of the visitors like browser type, language setting, country, visit time, city, and operating system.
Once the code is inserted the data can be collected and this can be useful to make certain decision regarding the current customers as well as to attract more customers.

III. OPEN WEB ANALYTICS (OWA)

It is ranked as the best alternative to Google analytic tool. The UI is most similar to GA, it also comes as an add-on in word press. There by useful to track sites inside the site as well as from outside of the sites using JavaScript, PHP or REST based APIs. OWA is written in PHP and uses a MySQL database, which makes it compatible for running with an AMP solution stack on various web servers.OWA is comparable to Google Analytics, though OWA is server software anyone can install and run on their own host, while Google Analytics is a software service offered by Google. OWA supports tracking with WordPress and MediaWiki, two popular web site frameworks. This application helps you keep track of and observe the influx of views on your website. The program also tracks your competitors and their company's growth compared to yours.OWA allows website owners to view detailed site visitor reports as graphs, charts, and other diagrammatic displays.
OWA can be configured in a variety of ways to meet the needs of your application(s).

- **Configuration File**

owa-config.php is OWA's main configuration file. This file can be created manually prior to an installation by simply copying owa-config-dist.php to owa-config.php in the same directory.
This configuration file is where OWA gets information about the database that it should connect to as well as a few other global configuration settings.

- **Configuration Constants**

The following convenience constants are included in your config file and should be used to setup certain aspects of your OWA instance.
_OWA_PUBLIC_URL-This is the web URL of OWA's base directory. This value must always end with a "/" (e.g. http://yourdomain/path/to/owa/).
OWA_ERROR_HANDLER This setting controls which of OWA's error handlers should be used. This value defaults to "production". Set this to "development" if you are trying to Debugging OWA.

OWA_LOG_PHP_ERRORS This will log all of PHP's errors/notices to OWA's errors.txt log file. This is handy for Debugging.
OWA_CACHE_OBJECTS This settings turns on and off OWA's object caching system.

- **Changing other Settings**

Instance wide configuration values can be set via the configuration file. To set a value use the $this->set() method and pass it the following arguments:

- name of the module that houses the setting your wish to change
- the key (or name) of the setting
- the value you wish to change it to.

For example:
$this->set('base, 'mailer-from', 'owa@myinstance.com');

- **Base Module Configuration Settings**

The default values for the base modules configuration settings are located in the following file: ...modules/base/classes/settings.php

- **Settings Management Screen**

Many of OWA's settings can be controlled via the Settings Management user interface. This screen is accessed from within the OWA user interface and allows Administrators to customize many of OWA's default settings and behavior.

- **Changing Settings from within PHP Code**

All of OWA's settings can be accessed and changed using the APImethods owa_coreAPI::getSetting() and owa_coreAPI::se tSetting() methods. owa_coreAPI::setSetting('modulename', 'somesetting', 'somevalue');

If you need to persist a setting to the database use the owa_coreAPI::persistSetting() method like so:
owa_coreAPI::setSetting('modulename', 'somesetting', 'somevaluetobestored');

IV. CONCLUSION

Web analytics is a technique of collection of web traffic data and its analysis. It helps in devising the right website and the correct marketing strategies. Google analytics and OWA have become important tools to help management of web apps and online business analysis

REFERENCES

[1] Waisberg, D. and Kaushik, A. (2009). Web Analytics 2.0: Empowering Customer Centricity, SEMJ.org 2, No. 2.

[2] Teixeira, J. (2011). Get Involved: 5 Types of Web Analytics tools to start using today! http://www.morevisibility.com/blogs/analytics/get-involved-5-typesof-web-analytics-tools-to-start-using-today.html [Accessed on 03 June 2014].

[3] *Clifton, B. (2012). Advanced Web Metrics with Google Analytics (3rd ed.). Indianapolis, IN: John Wiley & Sons.*

[4] Ganapathi, A., & Zhang, S. (2011). Web Analytics and the Art of Data Summarization. In Managing Large-scale Systems via the Analysis of System Logs and the Application of Machine Learning Techniques (pp. 6:1–6:9). New York, NY, USA: ACM.

[5] Rapoza, J. (2010, December 2). Web Analytics: A New View. Information Week. Retrieved from http:// www.informationweek.com/ web-analytics-a-new-view/d/d-id/1094560

[6] Ivan B. and Daniela G. *Web analytics tools and web metrics tools: An overview and comparative analysis,* Croatian Operational Research Review CRORR 6(2015), 373– 386

Guest Tagging System Analysis for Surveillance using IoT

Prachi Jain, Drishti Kadam, Dr. Deven Shah

Thakur College of Engineering and Technology, University of Mumbai, India

***Abstract* - Govardhan Eco-village needs to find an eco-friendly cost-effective solution to tag the different categories of guests arriving and granting them various permission access to different zones in the village. Every type of guest needs to be served in a more convenient manner so as to maximize the level of user experience at the village which is an attraction place for hundreds of devotees. After solving the above problem, we realized Crowd management is a very complex, challenging and costly exercise. If not managed or controlled effectively, it can result into crowd disasters aka stampede like situations. The aim of our proposed innovation is to provide a crowd management framework application which predicts decisions on how to distribute the crowd evenly in such spatially constrained areas of the village and also provide maximum user experience in such scenario.**

Keywords - Artificial intelligence, Business Intelligence, Crowd Risk, Crowd Disaster, IoT Analytics, Mass Gatherings, Neural Networks, Prepared Plan Evacuation, Precaution Plan Evacuation, Rescue Plan Evacuation, Simulation, Stampede, Splunk, Supervised Learning

I. INTRODUCTION

Govardhan Ecovillage home to a Wellness retreat and sustainable farming community is situated near Mumbai. Considering the present scenario at Govardhan Ecovillage, it has a manual way of authorizing the visitors by providing them simple Identity cards for authentication purposes. Being an 85- acre farming community, Govardhan Ecovillage (GEV), does not has an adequate way of controlling the access of the visitors and requires a system which can automatically manage the visiting crowd. Around a few hundred guests visit daily.

While finding the solution we realized, large accumulation of crowds at a specific area can prove to be dangerous. If the crowd is big enough, it can exert enough pressure to bend steel. Crowd management is essential to weaken this "Crowd Risk".

Crowd Disasters are tragically common in various parts of the world, including India, where heavy population density and a high tolerance for crowded spaces create gnarly conditions that can easily turn deadly.

Some Aspects of Crowd Management are:

1. Planning– Spatial, temporal and event related dimensions have to be considered in the planning process. The space which is going to host the crowd has to be studied first. Additional planning is required if any sudden surge of activity is anticipated at any point of time. Besides, special attention has to be given if there is a specific event happening in that space-time context

Sudden unforeseen emergencies (like a fire) can also trigger incidents like a human choking, stampede. Even when such incidents cannot be predicted, they should always be anticipated and prepared for

Enlarging or creating extra space for the event, counters, extending the time for the event, planning for large crowds, and anticipating whether any particular event can trigger sudden or unusual movement are the possible solutions.

II. PROBLEM DEFINITION

The objective of the problem definition is to find an eco-friendly cost-effective solution to tag the different categories of guests that come to Govardhan Eco Village, namely daily visitors, residential guests, unregistered day visitors, villagers and VIPs and predict analytics useful for the enterprise such as number of visitors at particular locations, their preferences etc.

Every type of guest needs to be served in a more convenient manner so as to maximize the level of user experience at the village which is an attraction place for hundreds of devotees.

III. PROPOSED SOLUTION

A. Design of Proposed Solution

The proposed solution uses a RFID tag and Long-Range UHF Passive reader to record the current location and timestamp of the guests (visitors). RFID Reader with a range of about 25 m will be present at a center of each prominent location of the Govardhan Eco-Village, thus covering a diameter of about
50 m. Each RFID Tag will be mapped with the details of the Visitor to whom it is assigned.

RFID reader will be used to record the RFID Tag's identity and current timestamp. RFID Reader will send this data along with its deployed location to a centralized database (Server) so that any kind of user can be tracked and their access can be controlled to different zones of the area. Long range communication will be implanted by line of sight communication whereas short range by using the radio frequency communication in walky-talkies which is already used in the village.

Real Time Analytics will be performed on data acquired so as to manage the flow of users in different zones of area effectively, reduce the waiting time of guests and serve the visitors in a more convenient manner.

For e.g., if the number of users is more in the Prasad Ghar of the village than its holding capacity we can move them to another area for other services.

Currently the entire scenario is simulated as a smart city model on Cup Carbon U-One 3.8, a Smart City and IoT Wireless Sensor Network Simulator Tool.

Also, Data Visualization in the form of charts and graphs will be generated automatically to provide knowledge which in turn will help in making more intelligent, optimized decisions.

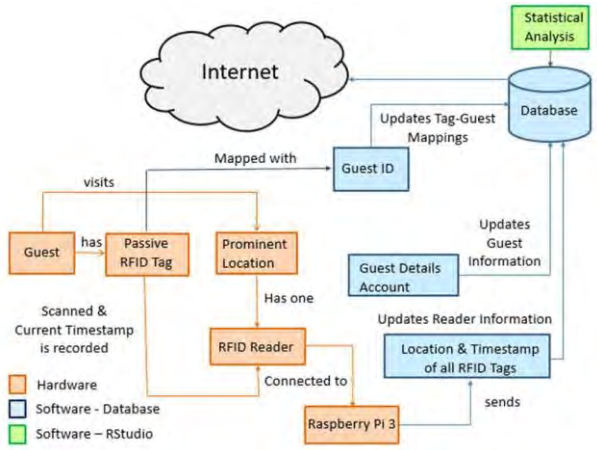


Fig. 1. Workflow of Guest Tagging System Design for Surveillance using IoT. A Guest visits a prominent location in village. Each guest is provided with the Passive RFID tag which is mapped with guest details. Passive RFID reader operated via a Raspberry Pi 3 at every prominent location of the village reads each tag and sends the tag's identity, current timestamp recorded and location to the database. All data is collected from the database and provided to RStudio for Statistical Data Processing and

B. Importance of the Proposed Solution

Proposed solution incorporates the following features:

1. Crowd Management and Control

   - Crowd management refers to situations where the evacuations places for the crowd is relatively more and therefore the crowd gets various options to ease itself.
   - Crowd Control refers to scenarios where evacuations places for the crowd is relatively less and therefore the crowd has few or one option(s) to ease itself. Hence, crowd is forced to follow a particular plan.

2. Prevention, Avoidance and Detection of Crowd Disasters (Stampede like phenomenon).

3. Real time with Archived Data Analytics

   The proposed innovation provides real time analysis and predictions useful to generate faster response time and effectively manage and control the crowd. Also. archived data is used for analytics purpose to generate more accurate predictions.

4. Track or Monitor Individual User Movements

   At the granular level, with the support of high computational power we even can able to track individual person movements as well.

5. High Scalability and Interoperability

   The proposed application can be integrated to any RFID System which track its users automatically or needs to swipe its RFID card such as Railway Stations which uses NFC cards, Attendee Tracking in any mass- gathering event etc. The proposed application can also be integrated to location aware mobile system, Wi-Fi membership etc.

6. Predicts and recognizes various Knowledge Patterns of a crowded environment visually on the dashboard such as:

   a. Peak time of crowd gatherings.

   b. Average waiting time of the user in a queue.

   c. Average time spent by the user in an area.

   d. Most popular and populated places under the area being monitored.

   e. Preferences of the crowd for the services offered under any RFID System (or other means like location aware mobile system, Wi-Fi membership) etc.

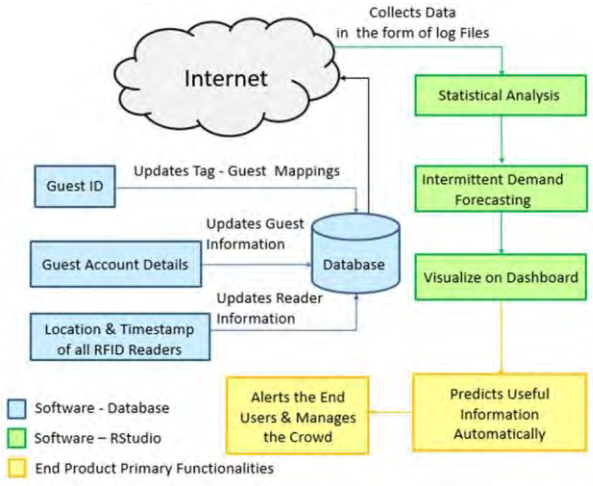


Fig. 2. Workflow of Guest Tagging System Analysis for Surveillance using IoT. All the collected data in the form of log files is forwarded to RStudio which uses Intermittent demand forecasting- Croston's Method for simple exponentially smoothening of data and forecasting intermittent demands (Stampedes here). Data Visualization's such as average waiting time of visitors, popular locations of visitors based on their preferences etc. is illustrated. Intermittent Demand forecasts helps administrators distribute the crowd adequately and control crowd disasters.

Alerts help the users avoid long waiting queues and prevent stampede like disasters.

IV. Methodology

The aim of our proposed innovation is to provide a crowd management framework application which predicts decisions on how to distribute the crowd evenly in spatially constrained areas and also provide maximum user experience in such scenario in order to prevent, detect and avoid Crowd Disasters (stampede like situation).

Statistical Analytics will be performed on data acquired so as to manage the flow of users in different prominent locations of the village effectively. For e.g., if the number of users is more in the Prasad Ghar of the village than its holding capacity we can move them to another area for other services. Analysis tool used is RStudio, an open-source software.

For Statistical Analytics, ARIMA's Model is used and Intermittent Demand forecasting for Stampede like phenomenon is done using Croston's method as the dataset is

- univariate,
- homogenous and
- with few attributes in nature.

ARIMA models are a popular and flexible class of forecasting model that utilize historical information to make predictions. ARIMA stands for auto-regressive integrated moving average and is specified by these three order parameters: (p, d, q). The process of fitting an ARIMA model is sometimes referred to as the Box-Jenkins method.

An auto regressive (AR(p)) component is referring to the use of past values in the regression equation for the series Y. The auto-regressive parameter p specifies the number of lags used in the model. For example, AR(2) or, equivalently, ARIMA(2,0,0), is represented as

$$Y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$$

where φ1, φ2 are parameters for the model.

The d represents the degree of differencing in the integrated (I(d)) component. Differencing a series involves simply subtracting its current and previous values d times. Often, differencing is used to stabilize the series when the stationarity assumption is not met, which we will discuss below.

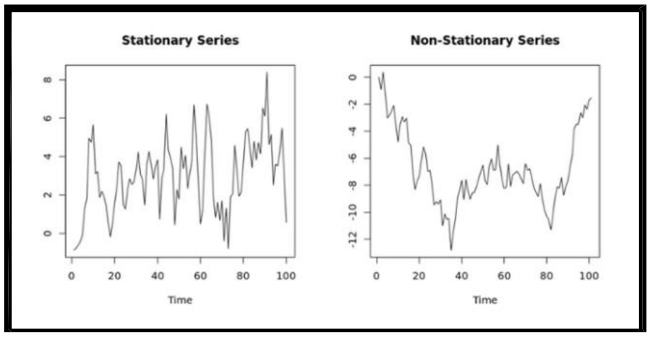
A moving average (MA(q)) component represents the error of the model as a combination of previous error terms e_t. The order q determines the number of terms to include in the model.

STEPS:

1. Plot data, examine examine its patterns & irregularities, prepare series for modelling
   - Clean up any outliers or missing values if needed
   - tsclean() is a convenient method for outlier removal and inputting missing values
   - Take a logarithm of a series to help stabilize a strong growth trend

2. Extract the seasonality component from the time series
Does the series appear to have trends or seasonality?
Use decompose() or stl() to examine and possibly remove components of the series

3. Stationarity
   - A series is said to be stationary when its mean, variance, and autocovariance are time invariant. Is the series stationary?
   - Use adf.test(), ACF, PACF plots to determine order of differencing needed.



4. Autocorrelations and choosing model order
   - Choose order of the ARIMA by examining ACF and PACF plots

5. Fit an ARIMA model

6. Evaluate and iterate
   - Check residuals, which should have no patterns and be normally distributed. If there are visible patterns or bias, plot ACF/PACF. Are any additional order parameters needed?
   - Refit model if needed. Compare model errors and fit criteria such as AIC or BIC.
   - Calculate forecast using the chosen model

Croston's Method is a modification of exponential smoothing for sporadic demand product time series suggested by Croston in 1972. The core value of this method is not only the estimation of average demand volume, but also estimation of time interval length between two non-zero demands, term called as intermittent demand.

The Croston method works in two steps,

1. Separate exponential smoothing estimates are made of the average size of a demand (number of Visitors).
2. Intermittent demands are calculated. This is then used in a form of the constant model to predict the future demand

- *Results and Discussion*



Fig. 4. Simulation of 500 routes & 5 Readers to design the Smart City for Guest Tagging System on CupCarbon One 3.8



Fig. 5. Simulation of 500 Mobile devices to design the Smart City for Guest Tagging System on Cup Carbon One 3.8 for a Simulation Time of around 100 Seconds

1	ARUN KHARA	AKOLA	M	54
2	ASHOK MARU	MUMBAI	M	49
3	RAMESH PARDESI	KHANDWA	M	31
4	SANJAY SHAH	KOPARGAON	M	15
5	AYUSH DOSHI	DHULE	M	9
6	MIHIR SAWANT	HARDA	M	22
7	OMKAR JANGAM	MADRAS	M	33
8	SHALINI AYYAR	KOPARGAON	F	29
9	NAMITA JANI	HYDRABAD	F	9
10	PRIYA MADNANI	MANGLORE	F	22
11	SHREYA PATKAR	AGRA	F	50
12	DAMINI ROHTE	INDORE	F	61
13	JANAK BHIMANI	DEWAS	M	34
14	NARAY ANAND	UJJAIN	M	51
15	SANJEEV BHASIN	NIMACH	M	40
16	RAJAK KHAN	RATLAM	M	16
17	MILAN DESPANDEY	NAGDA	M	42
18	NAMAN GOUR	SURAT	M	49
19	AJAY JAIN	BARODA	M	11
20	RAJESH BALAI	BARUCH	M	13
21	NITIN NANE	PALI	M	25
22	SUDHIR MANIK	JAMSEDPUR	M	27
23	PRAKASH REDKAR	BHILWARA	M	58
24	AJAY JAIN	BHIWANDI	M	46
25	RASHMI KAPDIYA	PALGHAR	F	24
26	RAMILA BALI	JALGAON	F	30
27	PUSHPA MANI	BHUSAWAL	F	20

IoT_Data

Fig. 6. 500 Guest Details Namely Name, Address, Gender, Age mapped with 500 RFID identies in comma separated format.



Fig. 7. Log file Generated from READER 1

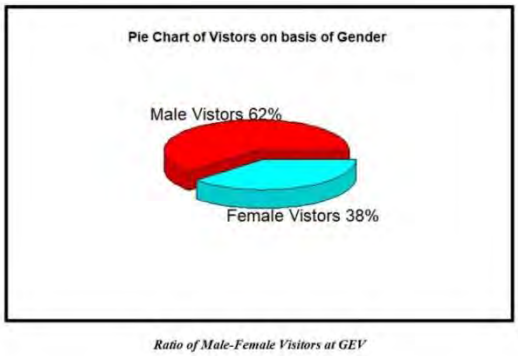


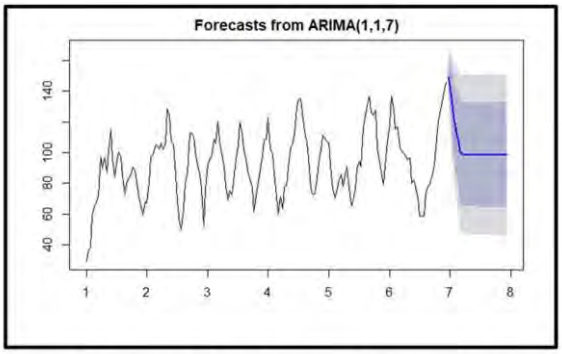
Table 1. Log Files and its attributes generated while simulation of Smart City Environment from Readers

ID	*Position x*	*Position y*	*Locations (A, B, C, D, E)*	*Current Timestamp*
1	-4.45	48.39	A	37.094064

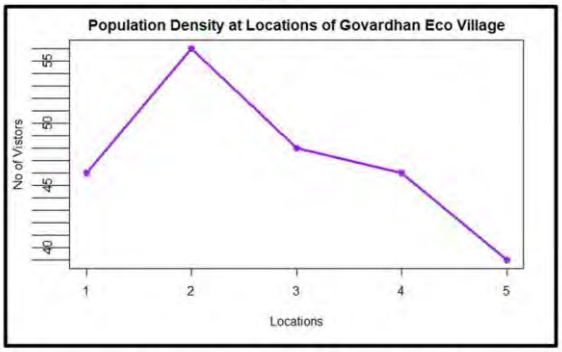
From the log files generated we can infer various types of analysis such as:

1.Track people preferences in the smart city.

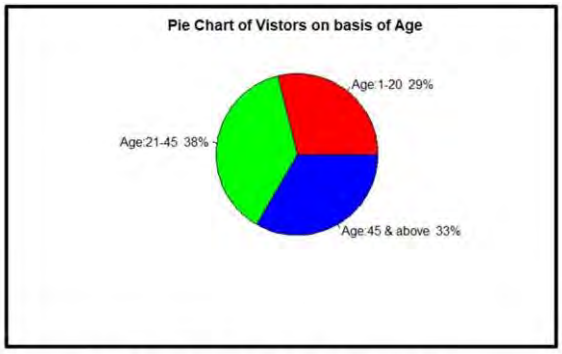
2. Track total number of people at particular locations.



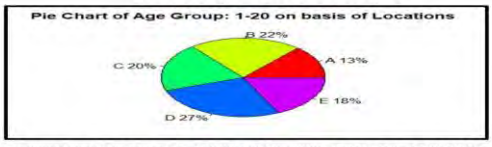
ARIMA Forecasts for the next Guests Ids going to visit Location A in next second after the simulation is completed.



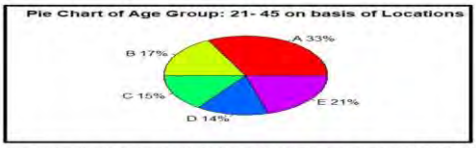
Population densities at Prominent Locations A, B, C, D and E inferred from simulation.



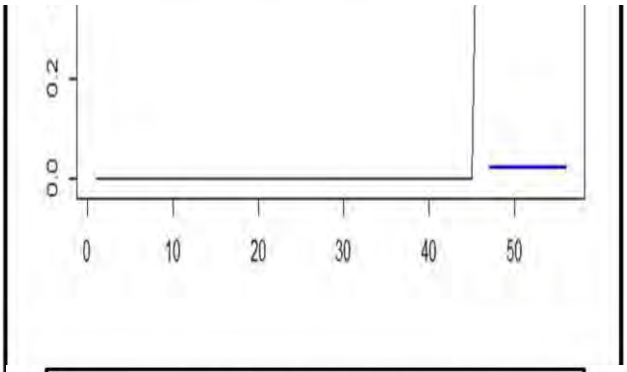
Ratio of Age Groups – 1-20, 21-40 and 45 + Visitors at GEV

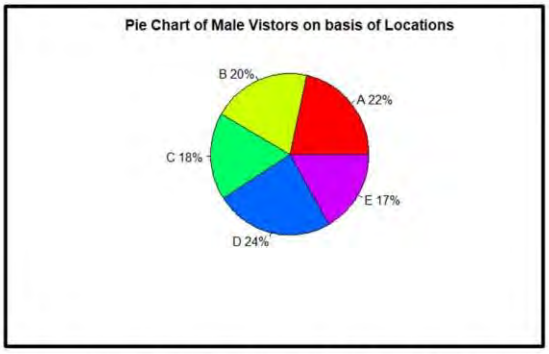


Proportion of Age group: 1-20 visitors at Locations A, B, C, D and E of GEV

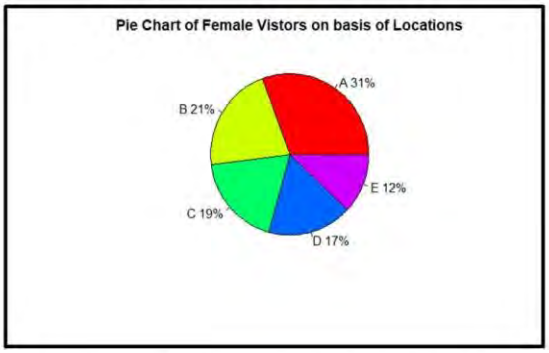


Proportion of Age group: 21-45 visitors at Locations A, B, C, D and E of GEV





Proportion of Male visitors at Locations A, B, C, D and E of GEV



Proportion of Female visitors at Locations A, B, C, D and E of GEV

In our proposed solution to manage crowd and control crowd disasters we have implemented ARIMA model and Croston's method for Intermittent Demand forecasting of Stampedes respectively.

We have implemented ARIMA models for all the 5 Locations – A, B, C, D and E in Govardhan Eco Village to forecasts the Guests Ids incoming at those locations after the simulation is completed. This helps us predict the number of incoming visitors at Govardhan Eco Village and thus provide useful insight in managing crowd. Croston's method for Intermittent Demand forecasting is performed to find out the probability of occurrence of stampede at Locations A, B, C, D and E with the threshold set to 45 – maximum number of people assumed to be the capacity of each location. Highest Probability is for Location B in the Village.

For a sample of 500 data assumed, around 62% are found to be male and remaining 38% are female. In males, the population is distributed as 22%, 20%, 18%, 24% and 17% at Locations A, B, C, D and E respectively. In females, the population is distributed as 31%, 21%, 19%, 17% and 12% at Locations A, B, C, D and E respectively. Thus, D and A are the most popular preferences of males and female visitors.

For a sample of 500 data assumed, around 29% are found to be of 1-20 years age, 33% are 21-40 years age and remaining 38% are 45+ years in age. In young age, the population is distributed as 13%, 22%, 20%, 27% and 18% at Locations A, B, C, D and E respectively. In middle aged, the population is distributed as 33%, 17%, 15%, 14% and 21% at Locations A, B, C, D and E respectively. In old age, the population is distributed as 22%, 20%, 18%, 24% and 17% at Locations A, B, C, D and E respectively. Thus D, A and D are the most popular preferences of young, middle and old age visitors.

VII. CONCLUSION AND FUTURE WORK

Crowd Monitoring and Management in large scale gathering are complex and multifactorial phenomenon which are not so easy to solve. We need automated systems as companions to help solve these issues in the most optimized manner. And save the great havoc- loss or damage to human life.

- The proposed solution tries to achieve both by providing assistance to the people serving and managing the crowd with accurate statistical knowledge and most optimized decision.
- In terms of performance neural networks provide accurate results with proper initialization of thresholds on the nodes, effective and accurate study of conditions constrained on the neural network.
- In terms of security, this method collects all guest entries who visit the village on remote server, so any theft or stranger will be detected.
- Manual entry work will be automated, so time, cost & man power will be reduced. All guest's entries will be forwarded online through internet. Previously entry work was in an offline mode where simple guest ids were given and noted in a register.
- Analysis will be performed on data acquired so as to manage the flow of guests in different zones effectively, reduce the waiting time and serve the visitors in a more convenient manner.

Thus, the proposed solution will be implemented to maximize the user experience during crowd management at large scale events and prevent stampede like phenomenon.

The proposed solution is more inclined towards the social and functional requirements of crowd management strategic approaches- preventing, detecting and avoiding stampede like phenomenon which creates a great havoc – loss of human life in large scale gathering events / areas. Also, it would be an added feature to maximize the user experience during crowd management at large scale events

The system can be scaled to various Crowd Management phenomenon like Stampede which can occur in any densely crowded areas such as Mumbai Local Stations, Pilgrimages like Hajj, Kumbh Mela, Festivals like Ganpati, Durga Puja, Mass Gatherings, Live Shows etc.

We know it's not that easy to manage a crowd of say 1 million, but when it comes to saving lives, some measures need to be taken.

VIII. ACKNOWLEDGMENT

Special thanks to Govardhan Eco-Village for providing us the problem definition and allowing us to visit your place for requirement gathering and pursue our innovative idea.

REFERENCES

[1]. RFID Handbook: Fundamentals And Applications In Contactless Smart Cards, Radio Frequency Identification And Near-Field Communication, Third Edition WILEY Publications

[2]. Designing the Internet of Things: Adrian McEwen, Hakim Cassimally WILEY Publications

[3]. Identification and tracking of persons using RFID-tagged items: United States Patent Application 20020165758 Kind Code A1 Hind, John R. ; et al. November 7, 2002

[4]. http://www.spychips.com/documents/ATT00075.pdf

[5]. The Use of RFID for Human Identification: A DRAFT REPORT from DHS Emerging Applications and Technology Subcommittee to the Full Data Privacy and Integrity Advisory Committee Version 1.0

[6]. https://www.dhs.gov/xlibrary/assets/privacy/privacy_advcom_rpt_rfid_draft.pdf

[7]. https://www.splunk.com/web_assets/pdfs/secure/Splunk_and_MapReduce.pdf

[8]. https://www.splunk.com/pdfs/solution-guides/splunk- guide-to-operational-intelligence.pdf

[9]. http://www.corerfid.com/Files/White%20Papers/033%20Health%20Fact%20Sheet.pdf

[10]. https://www.ecovillage.org.in
[11]. https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html
[12]. https://www.researchgate.net/post/Neural_Network_V_s_Fuzzy_Logic_V_s_Genetic_Algorithm
[13]. https://www.quora.com/Crowd-Control-Is-it-possible-to- prevent-large-human-stampedes
[14]. http://blog.kkaggarwal.com/2013/12/crowd-management-and-stampede-prevention/
[15]. http://decibelmanagement.com/crowd-dynamics-how-to-prevent-and-survive-a-human-stampede-or-crowd crush/
[16]. https://spectrum.ieee.org/tech-talk/consumer-electronics/portable-devices/hajj-pilgrims-to-get- electronic-racelets-in-wake-of-deadly-disaster
[17]. http://epmonthly.com/article/human-stampede-an-unexamined-threat/

To Study and Implement IEEE 802.11g Security Aspects Using Index Policy Method

Yamisha Savalia, Ashwini Nakil, Jasmina Ambani, Dr. Rajesh Bansode
Department Of Information Technology Thakur College of Engineering and Technology MumbaiIndia

***Abstract*-Wireless local area networks (WLANs) are beginning to play a much larger role in corporate network environments. The 802.11g specification is a standard for wireless local area networks (WLANs) that offers transmission over relatively short distances at up to 54 megabits per second (Mbps), compared with the 11 Mbps theoretical maximum with the earlier 802.11b standard. In this paper, focus is on WLAN security technologies designed to improve 802.11 standard by applying security policies. Policy indexing model are implemented using mesh topology (five nodes).Total packet transmission time for authentication and encryption of 802.1X model is evaluated with fixed packet size. The evaluation is done based on policy index in model implementation .The total number of iteration in complete file transmission is fixed. The use of different security configuration will provide both the necessary flexibility to network operators and high level of confidence to end users**

Keywords: 802.1X, RADIUS, security policy index, VPN, WEP

I. Introduction

The 802.11 wireless networks are fast becoming the preferred choice for LAN environments. Given their limited bandwidth (54 Mbps in 802.11g)[1] and the need for security in wireless standards, it is necessary to understand the relative overhead of different security protocols .The MAC sub-layer provides reliable data transmission for the IEEE 802.11 standard similar to a wired network. The IEEE 802.11g WLAN standard can be thought of as an intersection between the 802.11b and 802.11a standards. Like 802.11b, 802.11g operates in the same 2.4-GHz[2] portion of the radio frequency spectrum that allows for license-free operation on a nearly worldwide basis. An important mandatory requirement of 802.11g is full backward compatibility with 802.11b, which both provides investment protection for the installed base of 802.11b clients and extracts a substantial performance penalty when operating in this mode. Firewall hardware and software can be configured to allow safe Internet traffic to enter a network and block unsafe Internet traffic. It uses a series of rules to intercept and analyze data to determine whether the traffic is safe to enter either wired or wireless networks. The purpose of a firewall is to prevent unsafe traffic from entering to the network.

If there is a small group of users, the network can be limited to assigning a certain number of DHCP addresses equal to the number of possible users. RADIUS is a protocol for authentication, authorization and accounting of remote access connections. The user inputs his name and password and submits it to the RADIUS server.[3] The RADIUS server determines if the user is authorized to use the network. Communication between the user and the RADIUS server is encrypted; however, the RADIUS protocol does not provide data encryption. RADIUS is often used with a VPN. The paper is divided into ten sections were section II discusses the overview , section III describes details of IEEE 802.11 standards ,section IV describes the motivation in the field of wireless technology , whereas section V discusses the work carried out in the related fields of wireless security as literature review ,section VI details the problem definition ,section VII describes feasibility study ,section VIII describes the various techniques and methodologies ,section IX depicts the experimental results ,The conclusion based on the results achieved is stated in section IX and X followed by acknowledgement and References is in subsequent sections.

II. OVERVIEW

The project is to provide organizations with guidance for establishing secure wireless networks. Wireless carriers and network access providers must meet consumer demands for: easier access , better Quality of Service, Better security ,More network management tools ,Flexible pricing. The bit rate per user must be increased to support higher-end application, wireless streaming media, Video conferencing. IT should also fulfill the following challenges: Security, Multicast support, Location management support, Scalability. The parameters that are measured here are Encryption time, Packet transmission time, Computational complexity, Space complexity.

III. DETAIL OF IEEE 802.11 VARIOUS STANDARDS

The IEEE802.11 has released multiple set of standards for various operating frequency, and ranges specification. The first release was IEEE802.11 original standard release that was defined 1997 and clarified 1999. Some of these old standards are now obsolete, and some are still active. One would be interested in investigating in details the available standards to determine a suitable standard for the intended application of the WLAN network.

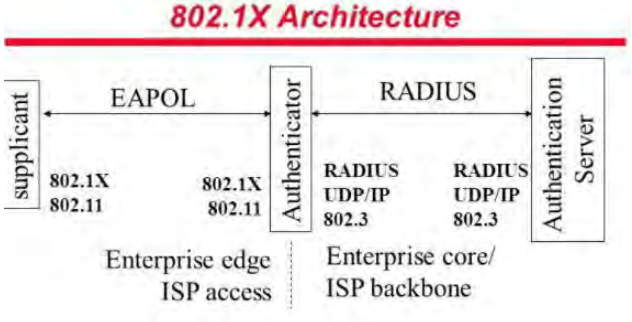


Fig.1 802.1X Architecture

Authenticator acts as a bridge.802.1x is a framework ,not a complete specification in and of itself. The actual authentication mechanism is implemented by the authentication server as shown in fig.1.

[1] IEEE 802.11a- The IEEE802.11a standard was released on September 1999. Networks using 802.11a operate at radio frequency of 5GHz or 3.7GHz and a bandwidth of 20MHz. The specification uses a modulation scheme known as orthogonal frequency-division multiplexing (OFDM) that is especially well suited to use in office settings. In 802.11a, data speeds as high as 54 Mbps are possible. This standard employ the single input, single output (SISO) antenna technologies, and the indoor/outdoor ranges from 35m to 125m for 5GHz operating frequency. The outdoor range goes to 5Km for operating frequency of 3.7G. The IEEE802.11a is less prone to interference compared to with 802.11b due to the high operating frequency of 5GHz.
[2] IEEE 802.11b- IEEE 802.11b standard was released on September 1999 as well. This standard provides 11 Mbps transmission (with a fallback to 5.5, 2 and 1 Mbps) in the 2.4 GHz operating frequency and bandwidth of 22MHz. [3]The 802.11b uses only DSSS (Direct Sequence Spread Spectrum) modulation technique. This standard also employs the SISO antenna technology as in the IEEE802.11a standard. The IEEE802.11b standard was ratified on 1999 from the original IEEE802.11 standard which allowed wireless functionality comparable to Ethernet. The IEEE802.11b standard is prone to higher interference due to the fact that the 2.4GHz frequency range is becoming crowded with carriers, hence increased interference risk. The indoor and outdoor ranges for this standard is 35m to 140m[2].
[4] IEEE 802.11g -The standard 802.11g was ratified in 2003 as an IEEE standard for Wi-Fi wireless networking and it supports maximum network bandwidth of 54 Mbps compared to 11 Mbps for 802.11b. This standard operates at 2.4GHz frequency and bandwidth of 20MHz. This standard uses the OFDM or DSSS modulation schemes. This standard employ the SISO antenna technologies, and its indoor/outdoor range are from 38m to 140m respectively.
[5] IEEE 802.11n -The 802.11n standard was ratified in 2009 and it utilizes multiple wireless antennas in tandem to transmit and receive data[3-4]. The IEEE802.11n standard employs OFDM modulation technique. The antenna technology used with the IEEE802.11n standard is known as Multiple Input, Multiple Output (MIMO). This technology refers to the ability of 802.11n and similar technologies to coordinate multiple simultaneous radio signals. The MIMO increases both the range and throughput of a wireless network. An additional technique employed by 802.11n involves increasing the channel bandwidth from 20MHz to 40MHz. The 802.11n standard support maximum theoretical network bandwidth up to 300 Mbps. The IEEE802.11n indoor/outdoor ranges are 75m, and 250m respectively.
[6] IEEE 802.11ac- IEEE 802.11ac is the fifth generation in Wi-Fi networking standards released December 2013[56]. This standard operating frequency is 5GHz, and bandwidth of 20, 40, 80, 160MHz sectors. The stream rates ranges for these bandwidth sectors are 7.2 - 96.3Mbps for 20MHz, and 15 – 200Mbps for 40MHz, 32.5 - 433.3Mbps for 80MHz, and 65 -
[7] 7Mbps for 160MHz. This standard exhibits better performance, and better coverage compared to IEEE 802. 11a,b,g and n standards. The 802.11ac standard uses a wider channel and an improved modulation scheme that also supports more clients. The IEEE 802.11ac standard utilizes a modulation technique known as multi-user MIMO. This technique allows a set of users or wireless terminals, each with one or more antennas, o communicate with each other. The indoor range is 35m, and there is no recorded max for outdoor range.

IV. MOTIVATION

The convenience of wireless LANs bring benefit to lives. They keep people connected beyond cubicle or office. People can access the Internet as easily from a coffee shop or our couches at home as while sitting at desks. An entry-level wireless Access Point (AP) costs less than $75, a price that makes these devices one of the best-selling computer peripherals since the iPod. Wi-Fi-enabled laptops can connect to your network without any wires. This convenience brings with it a huge risk to your network and data, because when you accept the configuration defaults of many of these inexpensive wireless APs, you run the risk of others connecting to your network or snooping on your data just as easily. Also the 802.11g standard devices are compatible with 802.11a and 802.11b devices. Hence, to provide more security to 802.11g devices and to take advantages of its low hardware cost we have undertaken the project.

V. LITERATURE SURVEY

A literature review surveys scholarly articles, books, dissertations, conference proceedings and other resources which are relevant to a particular issue, area of research, or theory and provides context for a dissertation by identifying past research. Research tells a story and the existing literature helps us identify where we are in the story currently. It is up to those writing a dissertation to continue that story with new research and new perspectives but they must first be
familiar with the story before they can move forward.

1. The solution found is Authentication using VPN is better and Encryption using 802.11 is better.The key findings are File Size=10KB,Packet transmission time 802.1x =17.37 sec VPN=21.99sec,ComputationalComplexity 802.1x=41.16sec,VPN=78.66sec,Space Complexity 802.1x=7.85 KB VPN=8.09KBand the Research gap found is to use 802.1x for encryption and VPN for authentication.[1]
2. The solution found is When more clients were added to the network the author found that overall throughput decreased by roughly 50% with additional client and 67% with two additional clients. The key findings are Transmission rates lowered to 500kbps,Bandwidth=12mbps,Packet sizes= 100,500,1000,1500 bytes and the research gap found is to Implement for 5 nodes.[2]
3. The solution found is the flaws lead to a number of practical attacks that demonstrate WEP fails to achieve its security goals. The key findings are the use of CRC-32 and RC4 can be justified by their speed and ease of implementation, they have discovered several security flaws in the protocol, stemming from misapplication of cryptographic primitives and the research gap is to Use 802.1x for encryption and VPN for authentication.[3]
4. The solution found is WEP has minor impact on FTP Throughput decreases by 7.5% and throughput for HTTP decreases by 50%.The key findings are the combined effects of encryption and authentication for which response time increases

by 268% and throughput decreases by 73%and the research gap is to use 802.1x for encryption and VPN for authentication.[4]

5. The solution found is Open VPN overcomes the weaknesses of WEP and there is an increase in performance of 802.11g with the use of compression. The key findings Throughput Max decrease 15.84%(UDP),Throughput Max decrease 16.91%(TCP), Latency increases for both TCP and UDP traffic as we increase the frame size, Frame loss percentage increases when we increase the frame size (up to 90%-92%).[5]
6. The solution found is EAP-TLS provides stronger authentication than EAP-MD5, They found that using small data amounts resulted in no visible differences between encrypted and unencrypted stream throughput. The key findings are in terms of Authentication Time for various Policies IPSEC=1.432(max) 1.405(min),802.1x EAP(MD5)
   Without IPSEC=1.749(Max) 0.427(min),802.1x EAP(MD5)with IPSEC=1.722(min) 1.749(max),802.1x EAP(TLS)without IPSEC=1.822(min) 3.144(max),802.1x EAP(TLS)with IPSEC=3.117(min) 3.144(max) and the research gap is to use EAP-TLS For authentication[6]

VI. PROBLEM DEFINITION

The experiment is simulated with 5 nodes. Initially a Net-work controller (server) is created with five nodes. Nodes are connected in a mesh network with each other. The further work was done as follows:

Phase I : Selecting a source node and text file to transmit. In this work file size is limited to 12kB.
Phase II : Selecting the Encryption method from the options for encrypting the file.
Phase III : On selecting the desired encryption method, time to encrypt and key length is calculated for that specific file.
Phase IV : After Encryption, authentication method was selected for concerned security level
Phase V : The decryption key was encrypted again using Diffie-Hellman protocol key exchange protocol for secure transfer of file from source to destination.
Phase V I: Random path was generated in the simulation for every file transfer.
Phase VII: On reaching the destination safely computational complexity, space complexity and packet transmission is calculated for security level 2, 3, 4, 5 and 6.

VII. FEASIBILITY STUDY

1. Technical Feasibility – Here one has to test, whether the proposed system can be developed using existing technology or not. It is evident that the necessary hardware and software are available for development and implementation of the proposed system. Hence, the solution is technically feasible.
2. Economic Feasibility - As part of this, the costs and benefits associated with the proposed system compared and the project is economically feasible only if tangible or intangible benefits outweigh costs. The system development costs will be significant. So the proposed system is economically feasible.
3. Legal Feasibility- Legal issues can affect a system's acceptance by users, its performance, or the decisions on whether to use it in the first place—so it is best to consider these explicitly in system design. Clearly, the behavior of those being enrolled and recognized can influence the accuracy and effectiveness of virtually any security system.
4. Operational Feasibility - It is a standard that ensures interoperability without stifling competition and innovation among users, to the benefit of the public both in terms of cost and service quality. The proposed system is acceptable to users. So the proposed system is operationally feasible.
5. Social Feasibility-The acceptability of a security system depends on the social and cultural values of the participant populations. A careful analysis and articulation of these issues and their identification can improve both acceptability and effectiveness.

VIII. Methodology

The Methodology is used to compare security index policies of 802.1X model with VPN model for response time, computational complexity, space complexity, key length and time taken to encrypt in simulator. Although there are a number of combinations that could be chosen for security configurations utilized for the testing of response time, computational complexity, space complexity, key length and time taken to encrypt. This research focuses on those most likely to be present in a corporate network environment. Proposed system architecture is shown in the fig.2 below. The authentication protocols that are examined are MAC address, PPTP, CHAP, HMAC-MD5. However, all levels of encryption, RSA, DES, WEP with 40 bit key are included at some point in the trials. Below is an overview of the security combinations selected for study:

i. Security Level 1 – This entails open association with no encryption on the data flow. This is the base line security scheme used as the starting point for all data comparisons with encryption and authentication.
ii. Security Level 2 – For 802.1X model encryption used is RSA and authentication is MAC address authentication. For VPN model RSA encryption and PPTP authentication were implemented.
ii. Security Level 3 – Open association with a 40-bit WEP key for encryption and HMAC-MD5 authentication for 802.1X model. DES encryption and CHAP (handshake procedure each time the client re-associated with the access point) authentication for VPN model.

In addition to the widespread use of these security settings, it is easy to use the measurement campaign. The combination of various security settings provided a broad look into overhead associated with encryption and authentication, and allows one to draw accurate conclusions on the effects that encryption and authentication have on network performance

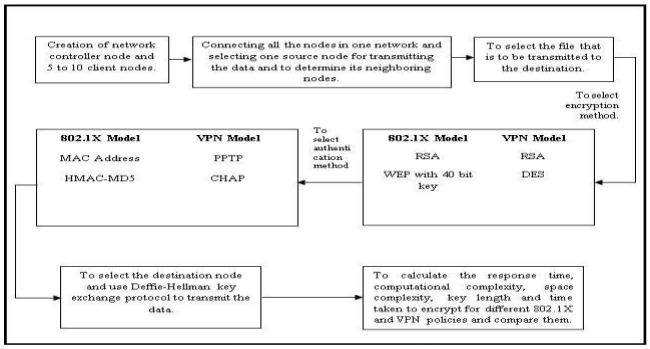
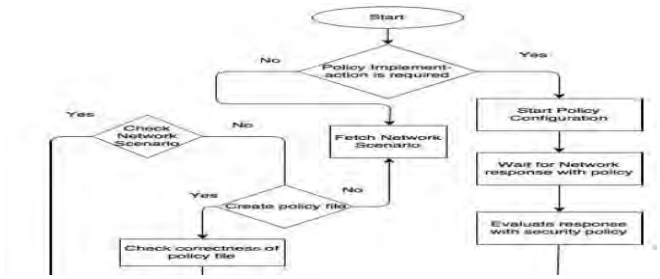


Fig.2 Performance System Architecture



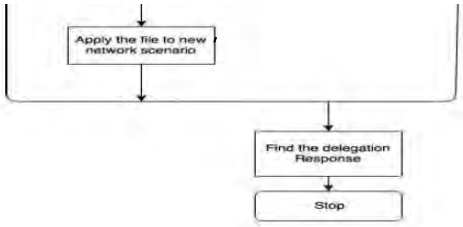


Fig.3 Flowchart for policy selection in optimized manner

IX. RESULTS

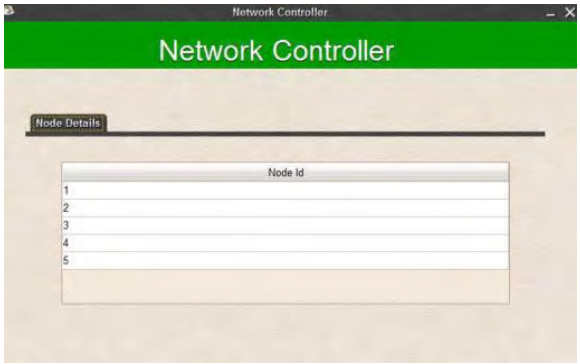


Fig 4 GUI of network

A network is formed by connecting nodes (personal computers). As shown in the figure 4.3 network consist of five nodes. The above network is connected through mesh topology. In Mesh topology every node is connected to every other node present in the network.

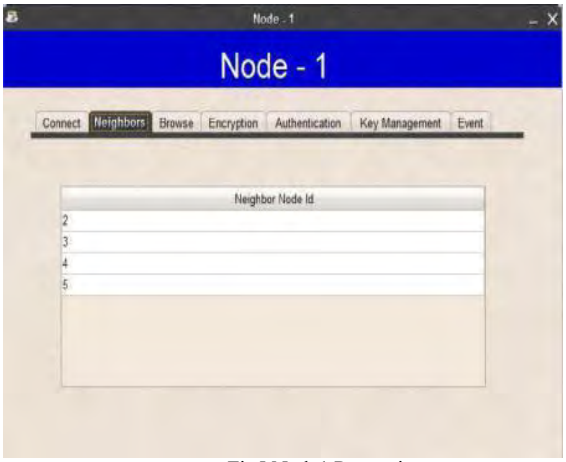


Fig5 Node1 Properties

The above fig 5.1 gives the description of the first personal computer acting as Node1 in a network. In mesh topology every node is connected to every other node present in the network.Since every node is connected in mesh topology and there are total five nodes neighbors for Node1 will be Node2, Node3, Node4, Node5

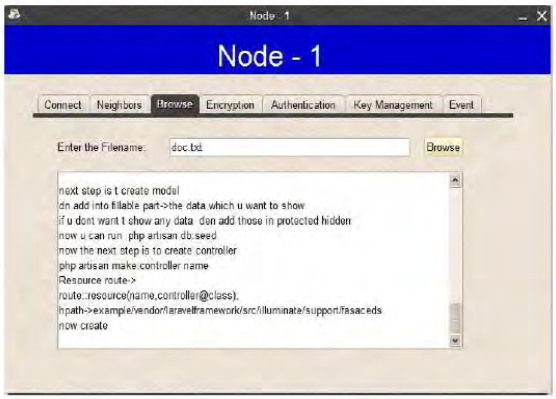


Figure.6 Node1 Properties

The above fig 5.1 gives the description of the first personal computer acting as Node1 in a network. In mesh topology every node is connected to every other node present in the network. Since every node is connected in mesh topology and there are total five nodes neighbors for Node1 will be Node2, Node3, Node4, Node5

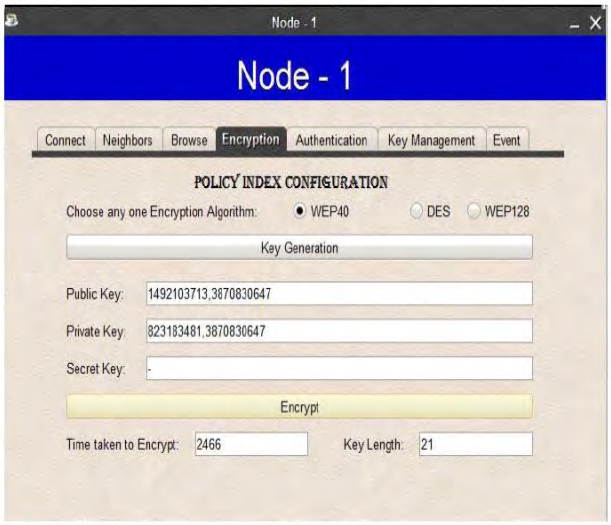


Fig 7 Encryption details (WEP40)

The above fig 5.7 shows Encryption(WEP40) used for file encryption and keys are generated hence time taken
to encrypt which is 2466 and key length is 21 calculated.

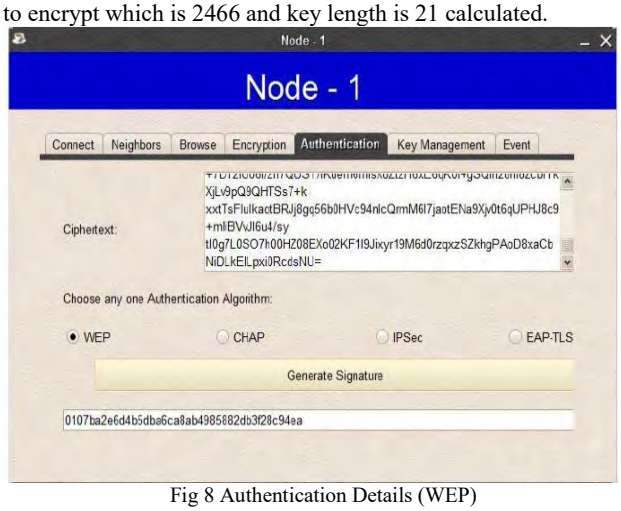


Fig 8 Authentication Details (WEP)

The above fig 5.10 shows Authentication (WEP) used for a file transmission and hence signature is generated.

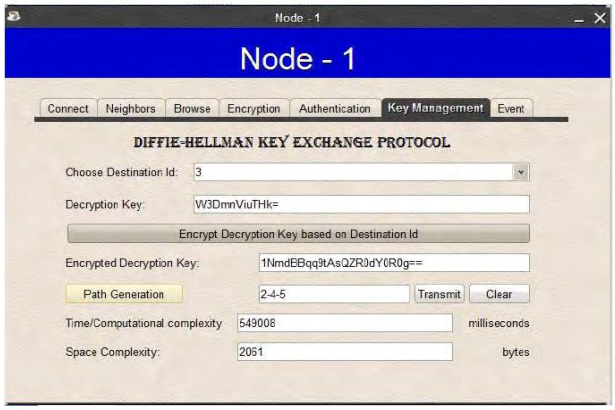


Fig 9 Key Management

Diffie-Hellman Key Exchange Protocol is implemented for key management. Destination node is selected,generated decryption key and encrypted based on destination node. Random path is generated to reach destination node and file is transmitted calculation of time/computational complexity and space complexity is done.

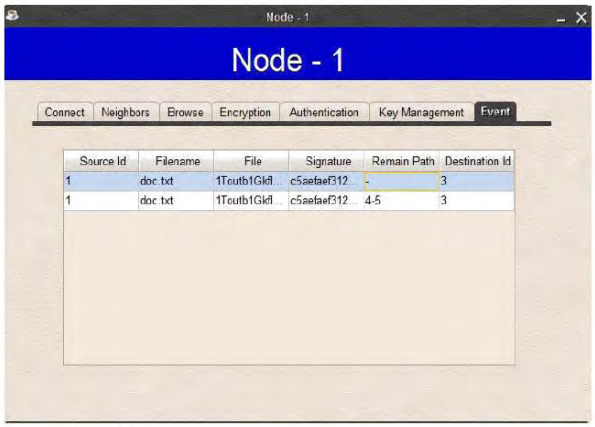


Fig 10

Event Description Events of file transmission is described in event description where following
Parameters are described for each transmission Source Id, Filename, Signature, Remain Path, Destination Id.

X. CONCLUSION

The project provides an in-depth look into the effects that encryption and authentication may have on network performance. From the results it is concluded that encryption on networks can be implemented efficiently, greatly reducing the amount of bandwidth allocated to encryption processes. As the security level goes on increasing the response time also increases proportionally. This approach provides flexibility to the user as well as the network engineers to design security level to use for a specific transfer of data. If the data security is not so important but good response time is required lower policies can be selected. If Security is of utmost importance without caring about the delay than one can chose higher security levels. Thus it can be observed from the results that higher the security level, higher is the computational time and packet transmission time. It can also be observed that authentication using VPN model is better than 802.1X model, whereas encryption for 802.1X model is better than VPN model. The combined effect of encryption and authentication yields good results in VPN than in 802.1X model.

ACKNOWLEDGEMENT

The people we would to thanks for their help in presenting this project seminar. we specially want to thanks our Principal Dr.B.K.Mishra, our HOD-IT and our project guide Dr. Rajesh Bansode whose advise have inspired and shaped much of the content of this project also guiding us on this project work and spending hours working to improve the content of this project. Also we want to thanks to all the staff members who help us to improve. So majority of our credit goes to all staff members.

REFERENCE

[1]. S. Kakade, R. .Bansode , ―Performance Evaluation of 802.11g Architecture Using Security Protocols Based on Index Policy Method,‖ IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), vol.11,Issue 4, Jul.2016

[2]. N. Baghaei ,‖ IEEE 802.11 Wireless LAN Security Performance Using Multiple Clients,‖ Department of Computer Science and Software Engineering University of Canterbury,2003

[3]. N. Borisov, I. Goldberg & D. Wagner ,― Intercepting Mobile Communications: The Insecurity of 802.11," Dept. of Computer Sci. & Software Eng., Canterbury Univ., Christchurch, New Zealand, Jul.2004

[4]. D.Nayak , D.B.Phatak & A.Saxena,― Evaluation of Security Architecture for Wireless Local Area Networks by Indexed Based Policy Method: A Novel Approach,‖ International Journal of Network Security, Vol.7, No.1,July 2008

[5]. A.TahaZamani, J.Ahmad ,‖Wireless LAN Security : IEEE 802.11g & VPN,‖International Journal of Advanced Research in Computer Science and Software Engineering‖ Vol.4, No. 2, Feb 2014]A. K. Agarwal, J. S. Gill & W. Wang ,― An Experimental Study on Wireless Security Protocols over Mobile IP Networks,‖ Technology conference,2004

Appendix C

Graduate Attributes:

Problem Analysis: Identify, formulate, research literature and analyze complex engineering problem searching substantiated conclusions using first principles of mathematics, natural sciences and engineering sciences.

Modern Tool Usage: Create, select and apply appropriate techniques, resources and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

Individual and Team work: Function effectively as an individual, and as a member or leader in diverse teams and in multi-disciplinary settings.

Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations and give and receive clear instructions.

Project Management and Finance: Demonstrate knowledge and understanding of engineering and management principles and apply these to one‘s own work, as a member and leader in a team, to manage projects in multidisciplinary environments.

Design/Development of solutions: Design solutions for complex engineering problems and design system components or processes that meet specified needs with appropriate consideration for public health and safety, cultural, societal, environmental considerations.

Predicting Outbreak of a Disease through Epidemiology using Data from Twitter

Sneha Milind Yadav, Dr. Brijit Marakarkandya
Department of Information Technology, Thakur College of Engineering and Technology Kandivali (E) Mumbai-400101 Maharashtra, India
sneha.m.yadav@gmail.com, bijith.marakarkandy@thakureducation.org

***Abstract*--The project is to make people aware of the Epidemic diseases in a particular location. The people can also search for other non-communicable diseases to check the tweets helpful to them in providing information regarding the disease. This social media (Twitter) data can be analyzed by the system to provide data regarding the diseases by checking the location from the tweets retrieved from Twitter. This will be done by searching for the disease in the system. Analyzing Twitter data can help health organizations such as State Health Departments & large HealthCare Systems to provide health advice & track health opinions of their populations and provide effective health advice when needed. Using tweets most used hash tag is retrieved to find out which disease is discussed to spread the awareness and it is helpful to prevent spreading of Communicable Disease.This can be utilized by anyone who needs information regarding the disease and can be utilized to support clinical providers, public health experts, social scientists, healthcare providers and Government.**

Keywords -- Epidemic, Data Mining, Hashtag, Semantic Analysis.

I. INTRODUCTION

The amount of data exchanged over social media is witnessing a major growth in the last few years. Various studies in different domains investigated extracting useful information from this exchanged data.

Public posts on a social website such as Twitter include personal status, opinion sharing, discussions, marketing, campaigning and so on. Among the material users share on Twitter are tweets related to health and healthcare. Some users share information and updates on their health or the health of loved ones. This is a common behavior by many people who seek support during difficult times. These social posts can be of tremendous value if detected and monitored by healthcare providers: they provide indicators about the general public health, they can help in early detecting of an epidemic, or can alert about ongoing concerns with health care.In addition, people often share their experience with healthcare facilities like hospitals, clinics, and health centers.[5]Twitter is one of the social media that is gaining popularity.Twitter is a gold mine of data. Unlike other social platforms, almost every user's tweets are completely public and pullable. This is a huge plus if you're trying to get a large amount of data to run analytics on. Twitter data is also pretty specific. Twitter's API allows you to do complex queries like pulling every tweet about a certain topic within the last twenty minutes, or pull a certain user's non-retweeted tweets [1].

Nowadays, detecting the diseases that are widely spread in the society and predicting the future stages of the diseases has become very important in modern life. Twitter - a social media platform - a platform used by billions of users around worldwide to express their ideas about various topics, including health conditions. Twitter is used as source for public health information on global scale. To detect the disease and predict the future spread SNS is helpful.

Social media platforms are also useful epidemiological data sources. Twitter is quickly becoming a favorite social media data source for epidemiologists because it can provide more contextual information than search queries. Frequently, people who are experiencing symptoms or have been confirmed to have a specific disease post about it on social media [3].Epidemiology is the study of the distribution and determinants of health-related states or events (including disease), and the application of this study to the control of diseases and other health problems.

The Internet, social media, and mobile phones become bigger parts of everyday life for more people, these data sources will contribute more to epidemiological studies. We are in the midst of a digital revolution, and epidemiologists are figuring out how to mine these data for useful public health information[4].Several Twitter studies have demonstrated that aggregating millions of messages can provide valuable insights into a populationThe objective of the project is to make people aware of the epidemic diseases and non-communicable disease in a particular location.

It will help them to understand the affected areas along with the useful information that people have shared through tweets.

It will predict the outbreak of disease by analyzing the tweets

II. BACKGROUND AND MOTIVATION

As many people are using social media to share different information we are utilizing this fact to make our system more

efficient and reliable by taking the data from Twitter and analyzing various factors like sentiments, location, etc.

Social media (Twitter) data can be analyzed by the system to provide data regarding the epidemic diseases by checking the location from the tweets retrieved from Twitter[4].

This will be done by searching for the disease in the system. This can be utilized by anyone who needs information regarding the disease.

As we conduct our lives online—through Internet searches, social media posts, or the use of our mobile phones. There is a rich treasure trove of data that, though it may provide only a partial glimpse of a disease outbreak, is proving to be quite nimble at capturing the spread of disease. Traditional tracking methods—surveillance of clinics, telephone triage calls, purchases of over-the-counter medications, and school absenteeism, for example—are far more unwieldy. Although researchers caution that these new sources of information cannot completely replace traditional epidemiological methods, they are serving to fill in critical gaps.

Scope of the project is presently specific. Both the user must have this software installed on systems to know regarding the status of the disease

Each step is clearly stated and the user will not face any ambiguity in using the software.The software provides clarity in its functionality even to naïve users.Several Twitter studies have demonstrated that aggregating millions of messages can provide valuable insights into population The objective of the project is to make people aware of the epidemic diseases and non-communicable disease in a particular location.

It will help them to understand the affected areas along with the useful information that people have shared through tweets.It will predict the outbreak of disease by analyzing the tweets.

The amount of digital health related data is becoming more and more huge, being generated both by healthcare industries (e.g. medical records and exams) as well as by social media and virtual networks, where individuals share their experiences and opinions about different topics, including personal health (illnesses, symptoms, treatments, side effects).

While data owned by healthcare industries are often accessible only with restrictions, social media data are generally publicly available, therefore they represent an enormous resource for mining interesting healthcare insights. Among various social networks, the one on-the-edge is Twitter , the micro-blogging service whose restriction of 140 characters for post encouraged the development of a kind of shorthand and speed in composing messages. Twitter has been recently used as an information source to predict and/or monitor real world outcomes.

III. METHODOLOGY

Extracting Tweets:

The Twitter API is simply a set of URLs that take parameters. They URLs let you access many features f Twitter, such as posting a tweet or finding tweets that contain a word, etc.

The Twitter API allows you to access the features of Twitter without having to go through the website interface. This can be useful for doing things like posting tweets or sending directed messages in an automated way with scripts[10].

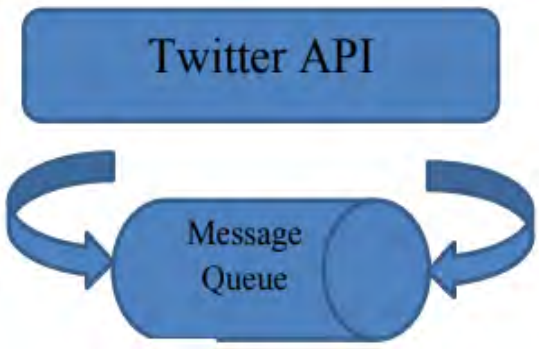


Figure 1: Twitter API

1.Get Tweets by hashtag and store them in the message queue

2.Process(mark as Favorite) Messages in the Queue

- **Data Mining:**

Data mining is defined as a process of discovering hidden valuable knowledge by analyzing large amounts of data, which is stored in databases or data warehouse, using various datamining techniques such as machine learning, artificial intelligence(AI) and statistical.Twitter is a global social media platform and it is nothing less than a goldmine when it comes to data and information. Nearly all tweets are public and easily extractable, which makes it easy to gather large amount of data from twitter for analysis. The fact that twitter data is very specific

makes it extremely good for prediction purposes.



Figure 2: Data Mining Process

Data Preparation:



Figure 3: Data Preparation Process

Data Preparation is the process of collecting, cleaning, and consolidating data into one file or datatable, primarily for use in analysis

Data preparation is necessary to manipulate and transform raw data so that the information content enfolded in the data set can be exposed, or made more easily accessible. This is the first step in data analytics projects for data wrangling and can include many discrete tasks such as loading data or data ingestion, data fusion, data cleansing.Data cleansing is one of the most common tasks in Data preparation.

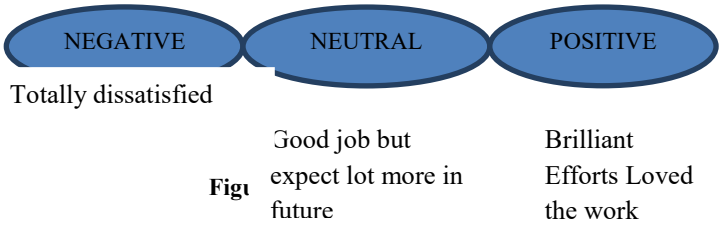
This process is carried out to remove blank spaces, stop words.

- **Stemming Algorithm:**

Stemming is the process of reducing a word to its word stem that affixes to suffixes and prefixes or to the roots of words known as a lemma. Stemming is important in NLUand NLP. Stemming is a part of linguistic studies in morphology and artificial intelligence (AI) information retrieval and extraction. Stemming and AI knowledge extract meaningful information from vast sources like big data or the Internet since additional forms of a word related to a subject may need to be searched to get the best results. Stemming is also a part of queries and Internet search engines[9]

- **Sentiment Analysis**

Opinion mining (sometimes known as sentiment analysis or emotion AI) refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information[7]. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine



Figu

Sentiment Analysis is the process of 'computationally determining whether a piece of writing is positive, negative or neutral. It's also known as opinion mining, deriving the opinion or attitude of a speaker.

Business: In marketing field companies use it to develop their strategies, to understand customers' feelings towards products or brand, how people respond to their campaigns or product launches and why consumers don't buy some products.

Politics: In political field, it is used to keep track of political view, to detect consistency and inconsistency between statements and actions at the government level. It can be used to predict election results as well!

Public Actions: Sentiment analysis also is used to monitor and analyse social phenomena, for the spotting of potentially dangerous situations and determining the general mood of the blogosphere.

- ANOVA(Statistical Analysis)

Analysis of variance (**ANOVA**) is a collection of statistical models and their associated estimation procedures (such as the "variation" among and between groups) used to analyze the differences among group means in a sample.

- Network Density

In a network, the things that are connected are usually called “nodes.”A node might be a person, a computer, or even some hyperlinked text. The connections between nodes are sometimes called “connections” and sometimes called “edges” – but it’s really just the same thing

“Network density” describes the portion of the potential connections in a network that are actual connections. A “potential connection” is a connection that could potentially exist between two “nodes” – regardless of whether or not it actually does. This person could know that person; this computer could connect to that one. Whether or not they do connect is irrelevant when you’re talking about a potential connection[1].

➢ **PROPOSED FLOW**

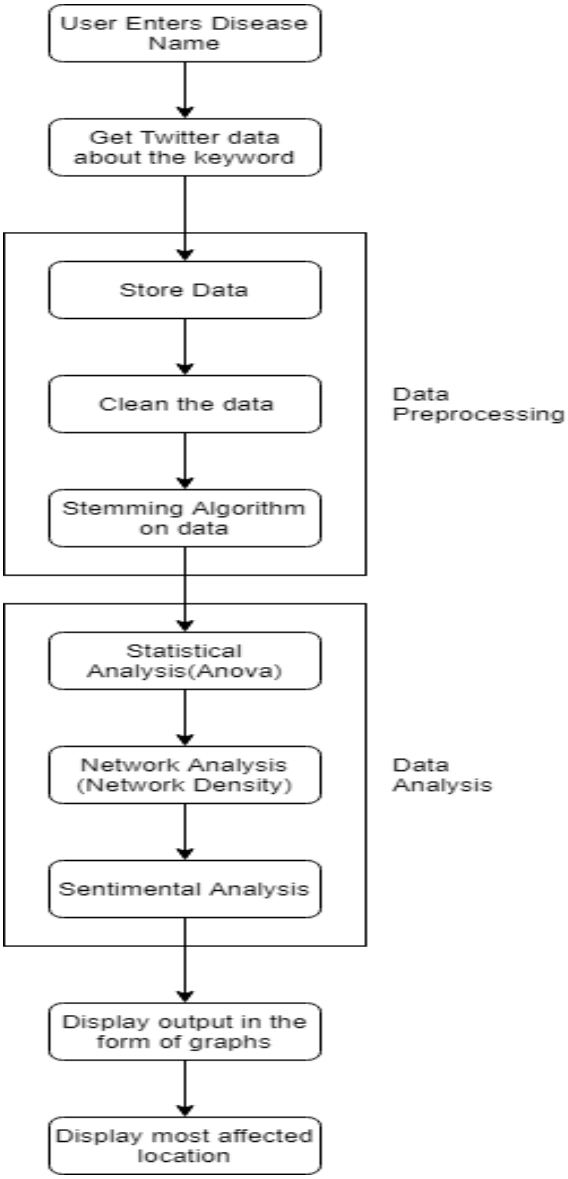


Figure 5: Proposed work flow of the system

IV. CONCLUSION

The study shows social media platform (Twitter) is being used for health related discussions and technologies. Some of the tweets are written by the bloggers and influencers for giving the moral and emotional support through social media platforms. The top frequent tweets show the public awareness and concern with healthcare services.

Social Media is currently playing a fundamental role in the life of many internet Users. Twitter is a social networking and micro blogging service, enabling registered users to read and post short messages, so-called tweets. Twitter has 326 billion Users

Worldwide providing a unique opportunity to understand user's opinion with respect to the most common health issues.

The System results in graph including hash tag and disease name. It will display the information about the disease entered by user using retrieved tweets which help many social healthcare system and medical experts, researchers , advisor to know the location along with the affected areas.

Location of outbreak in case of Epidemic Disease where tweets are more discussed about disease through most used hash tag.

REFERENCES

[1]Purva Grover, Arpan Kumar Kar,Gareth Davies " Technology enabled Health" – Insights from twitter analytics with a socio-technical perspective ELSEVIER/International Journal of Information Management Volume 43, December 2018, Pages 85-97

[2]Kenny Byrd ; AlisherMansurov ; Olga Baysal "Mining Twitter Data for Influenza Detection and Surveillance" 2016 IEEE/ACM International Workshop on Software Engineering in Healthcare Systems (SEHS)

[3]Amir Karami, Alicia A. Dahl, Gabrielle Turner-McGrievy, Hadi H K Kharrazi, George Shaw "Characterizing diabetes, diet, exercise, and obesity comments on Twitter" ELSEVIER/International Journal of Information Management 38(1):1-6 · February 2018

[4]Ahmed Ali, WalidMagdy, and Stephan Vogel "A Tool for Monitoring and Analyzing HealthCare Tweets" HSD 2013, August 1, 2013, Dublin, Ireland.

[5]Xiang Ji, Soon Ae Chun, James Geller "Monitoring Public Health Concerns Using Twitter Sentiment Classifications" IEEE International Conference on Healthcare Informatics 2013

[6]HarshavardhanAchrekar ; AvinashGandhe ; Ross Lazarus ; Ssu-Hsin Yu ; Benyuan Liu"Predicting Flu Trends using Twitter data" IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)2011

[7]Hideo Hirose, Liangliang Wang "Prediction of Infectious Disease Spread using Twitter: A Case of Influenza" Fifth International Symposium on Parallel Architectures, Algorithms and Programming 2012

[8]KrutiNargund ; S. Natarajan " Public health allergy surveillance using micro-blogs" International Conference on Advances in Computing, Communications and Informatics (ICACCI) 2016

[9]KarolosTalvis ; KostantinosChorianopoulos ; Katia LidaKermanidis "Real-Time Monitoring of Flu Epidemics through Linguistic and Statistical Analysis of Twitter Messages" 9th International Workshop on Semantic and Social Media Adaptation and Personalization 2014

[10]P AmruthaValli ; M Uma ; T Sasikala "Tracing out various diseases by analyzing Twitter data applying data mining techniques" International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) 2017

[11]RanjithaKashyap ; AniNahapetian "Tweet analysis for user health monitoring"

4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH) 2014

[15]Michael J. Paul and Mark Dredze "You Are What You Tweet: Analyzing Twitter for Public Health" Association for the Advancement of Artificial Intelligence (www.aaai.org) 2011

[16]https://www.sciencealert.com/how-researchers-can-use-your-tweets-to-study-your-health

Emotion Detection and Feedback Generation

Akshdeep Sharma, Rishabh Gupta, Shivangi Singh, Parth Khurana

Department of Computer Engineering MIT College of Engineering, Pune, India

***Abstract*- Face and emotion detection is one of the current topics which provides solution to various challenges. Human emotions & facial expressions play an important role in non-verbal communication. Humans demonstrate and convey a lot of evident information visually rather than verbally. Although humans recognize facial expressions without efforts or delay. Where as, reliable expression recognition by machines remains a challenge as of today. To automate recognition of emotional state, machines must be taught to understand facial gestures which can be used widely in different environments. For any face and emotion detection system, dataset is the most important part for the comparison of the facial features. The dataset is used for the evaluation of the face and emotion by using different algorithms. In this paper, we propose a system we will analyze the emotions of the students in a classroom and generate the required feedback based on those emotions. For detecting face from the input image we are using haar-classifier and to evaluate the face and emotion detection Support Vector Machine(SVM) is used.**

Keywords: Face Detection, Emotion Detection, Facial Landmarks,Feature Extraction, Haar- classifier, SVM.

I. INTRODUCTION

Face plays significant role in social communication. This is a 'window' to human personality, emotions and thoughts.According to the psychological research,non-verbal part is the most informative channel in social communication.

Verbal part contributes about 7% of the message, vocal – 34% and facial expression about 55% to the effect of the speaker's message.

Emotions are feeling or response to particular situation or environment. Emotions are an integral part of our existence, as one smiles to show greetings, frowns when confused, or raises one's voice when enraged. This is because we understand other emotions and react based on that expression,that enriches the interactions. Computers are "emotionally challenged" i.e. they cannot directly understand human emotions or expressions.

Face is a subject of study in many areas of science such as psychology, behavioural science, medicine and finally computer science.Facial emotion recognition is one of the specific issues of computer vision. Emotions can be classified like fear, happiness, joy, sadness and aggressiveness. Emotional expressions are related to the movements or positions of the muscles under the skin and are a form of nonverbal agreement. Figure 1 shows different facial expressions.



Fig. 1. Facial Expressions(url: https://www.researchgate.net/figure/Universal-facial-expressions-of-seven-emotions-Source_fig1_312160510)

But there are still avariety of limitations that make face & emotion detection difficult. For example:

- Occlusion: faces may be partially occluded by other objects.
- Presence or absence of structural components: beards, moustaches and glasses.
- System cannot detect if a person is faking emotion.
- Pose (Out-of Plane Rotation): frontal, 45 degree, profile, upside down.
- Orientation (In Plane Rotation): Face appearance directly varies for different rotations about the camera's optical axis.
- Imaging conditions: lighting and camera (sensor response, gain control, lenses), resolution.

The remainder of this paper is organized as follows: Section 2, consists of the literature survey of the paper. In

Section 3, the working procedure of each stage of the proposed system is portrayed. Finally, conclusions are presented in section 4.

II. LITERATURE REVIEW

Saumil Srivastava[1],discusses a novel method for Facial Expression Recognition System which performs facial expression analysis in a near real time from a live web cam feed. Primary objectives were to get results in a near real time with light invariant, person independent and pose invariant way. The system is composed of two different entities trainer and evaluator. Each frame of video feed is passed through a series of steps including haar classifiers, skin detection, feature extraction, feature points tracking, creating a learned Support Vector Machine model to classify emotions to achieve a trade-off between accuracy and result rate. A processing time of 100-120 ms per 10 frames was achieved with accuracy of around 60%. We measure our accuracy in terms of variety of interaction and classification scenarios.

Marryam Murtaza[2], introduces automatic face recognition is one of the most emphasizing dilemmas in diverse of potential relevance like in different surveillance systems, security systems, authentication or verification of individual like criminals etc. Adjoining of dynamic expression in face causes a broad range of discrepancies in recognition systems. Facial expression not only exposes the sensation or passion of any person but can also be used to judge his/her mental views and psychosomatic aspects.

Shruti Bansal[4],Facial expressions play a major role in Face Recognition Systems and image processing techniques of Human Machine Interface. There are several techniques for facial features selection like Principal Component Analysis, Distance calculation among face components, Template Matching. This algorithm describes a simple template matching based facial feature selection technique and detects facial expressions based on distances between facial features using a set of image databases. The algorithm involves three stages: Pre Processing, Facial Feature Extraction and Distance Calculations. Then, identify whether a human is smiling or not using the measurement of Euclidean distances between pairs of eyes and mouth region of that face quality requires new tools for size and colour measurement and capturing the fruit side view image, some fruit characters is extracted by using detecting algorithms.

S. Sharmila[12],describes that Student engagement has been a key topic inside the educational training. The three specific styles of engagement of the students in a class are: behavioural, emotional, and cognitive. The time period behavioural engagement is commonly used to describe the scholar's willingness to participate within the getting to know system. Emotional engagement describes a scholar's emotional attitude toward learning. Cognitive engagement is a chief part of overall learning engagement. From the facial expressions the involvement of the students in the magnificence can be decided. Commonly in a lecture room it's far difficult to recognize whether the students are able to understand the lecture or no longer. So that you can know that comments form will be collected manually from the students. However those feedbacks given by using the students will now not be correct. Hence they will no longer get proper comments. This hassle can be solved by means of the use of a facial emotion evaluation. From the facial expression the emotion of the students may be analysed and required feedback can be generated.

III. PROPOSED SYSTEM

The proposed system will be able to analyse the emotion or expressions of the students in a classroom and generate the required feedback. This feedback can be used by the teacher to evaluate student engagement and improve interaction.

SYSTEM ARCHITECTURE:

The proposed system architecture is given below:

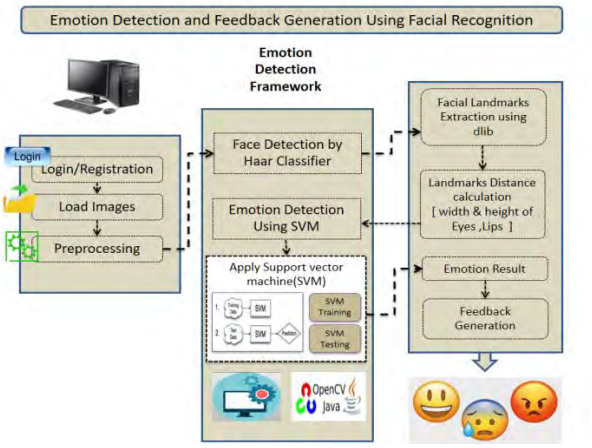


Fig. 2. System architecture of the proposed system.

Figure 3 demonstrates overall Activity diagram of the proposed scheme:

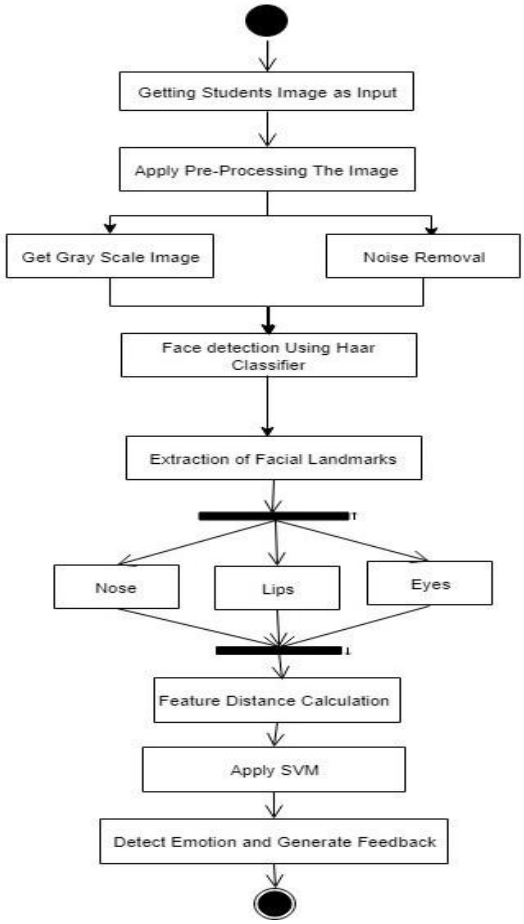


Fig.3. Activity diagram of the proposed system.

Procedural steps:

The proceduralstepscomprises of three key stages: face detection, Feature Extraction and Expression classification. Figure 3 demonstrates the overall block diagram of the proposed scheme. Each stage of the scheme is described below in detail.



Fig. 4. Block Diagram of the proposed Scheme.

Facial expression recognition is composed of three major steps:

(1) Face detection:

Open-CV a popular image processing library is used in the proposed system. Pre-processing of image (like noise removal, RGB to Grey-scale conversion etc.) is neededand can be easily done using Open-CV'spre-defined library functions. After the pre-processing is done Haar-Classifier is used for Face detection to detect all the student faces in the image.

HAAR-CLASSIFIER:Haar- classifiers is an effective facedetection method proposed by Paul Viola and Michael Jones. It is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images. Then we need to extract features from it. For this, Haar features shown in the below image are used. Each feature is a single value obtained by subtracting sum of pixels under the white rectangle from sum of pixels under the black rectangle.

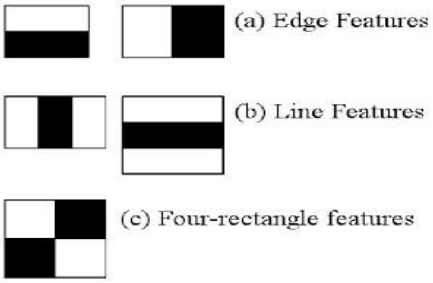


Fig. 5. Different featuresextracted from an image: (a) Edge Features, (b) Line features (c) Four-rectangle features. (url:https://www.researchgate.net/figure/Haar-features_fig1_284253914)

(2) Feature extraction:

The next step is feature extraction. On all the faces that are detected in the first step we will be extracting facial landmarks using the dliblibrary. The pre-trained facial landmark detector inside the dlib library is used to estimate the location of 68 (x, y)-coordinates that map to facial structures on the face.Figure 5 represents facial landmarks:

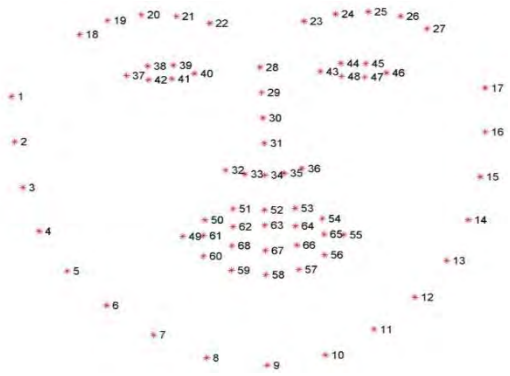


Fig. 6. Facial landmarks in an image.(url: http://gigantesdescalzos.com/182665/draw-filter-on-face-dlib/facial-draw-filter-on-face-dlib-landmarks-with-opencv-and-pyth-pyimagesearchrhpyimagesearchcom-a-functial-regressi-jpg/)

Then the features can be evaluated by calculating the Euclidian distance between two points (like in above image when a person is smiling the distance between the points 49-55 will increase). Examining the image, we can see that facial regions can be accessed via simple indexing:

- The lips can be accessed through points [49, 55].
- The right eyebrow through points [18, 22].
- The left eyebrow through points [23, 27].
- The right eye using [37, 40].
- The left eye with [43, 46].
- The nose using [28, 31].
- And the jaw via [1, 17].

These seven features can then be stored for every face in our database (MySQL).

(3) Expression classification or recognition:

The third step is to classify or recognize the emotion in the image based on the features extracted. A supervised machine learning algorithm (SVM) can be used for this purpose. SVM will be pre-trained to recognize the emotions in the image and then our database will be passed to SVM as testing data and it will map all the faces in the image to an emotion. Then the percentage of the every emotion in the image will be evaluated and based on that a rating will be generated as a feedback.

Support Vector Machine: The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N—the number of features) that distinctly classifies the data points. Figure 7 demonstrates data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane becomes a two-dimensional plane.

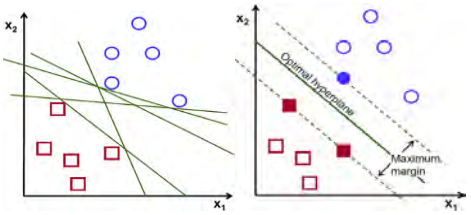


Fig.7. (url: https://towardsdatascience.com/support-vector-machine-vs-logistic-regression-94cc2975433f)

Applications:

Health care, Games, E-learning, Education System.

IV. CONCLUSION

This system proposes an efficient scheme for Emotion detection and Feedback generation using Image Processing

techniques. The system will perform various operations such asface detection, feature extraction and expression classification and based on the classified emotions, feedback canbe generated and engagement of the students in the classroom can be evaluated.

V. ACKNOWLEDGMENT

This work was supported by our mentorprof. AnitaThengade. We would like to also thank the students of MIT College Of Engineering(MITCOE), Pune.

VI. REFERENCES

[1] Saumil Srivastava, Krishna Asawa, "REAL TIME FACIAL EXPRESSION RECOGNITIONUSING A NOVEL METHOD", the International Journal of Multimedia & Its Applications (IJMA) Vol.4, No.2, April 2012.

[2] Marryam Murtaza, Muhammad Sharif, Mudassar Raza, and Jamal Hussain Shah "Analysis of Face Recognition under Varying Facial Expression: A Survey", The International Arab Journal of Information Technology, Vol. 10, No. 4, July 2013.

[3] Dolly Reney and Dr.NeetaTripaathi, "An Efficient Method to Face and Emotion Detection", 2015 Fifth International Conference on Communication Systems and Network Technologies IEEE, 2015.

[4] Shruti Bansal and Pravin Nagar "EMOTION RECOGNITION FROM FACIAL EXPRESSION BASED ON BEZIER CURVE," International Journal of Advanced Information Technology (IJAIT) Vol. 5, No. 3/4/5/6, December 2015.

[5] Surbhi , Mr. Vishal Arora , "A Face Identification Technique for Human Facial Image", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (6), 2012,5390-5393.

[6] Deepak Ghimire, Joonwhoan Lee, Ze-Nian Li, SunghwanJeong, "Recognition of Facial Expressions Based on Salient Geometric Features and Support Vector Machines", Springer.

[7] ByoungChul Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information", www.mdpi.com/journal/sensors 30 January 2018

[8] Dr. Sanjay Kumar and Ayushi Gupta, "Facial Expression Recognition: A Review" Special Conference Issue: National Conference on Cloud Computing & Big Data.

[9] Saeed Turabzadeh, Hongying Meng , Rafiq M. Swash, MatusPleva and JozefJuhar "Facial Expression Emotion Detection for Real-Time Embedded Systems" Technologies 2018, 6, 17; doi:10.3390/technologies6010017 www.mdpi.com/journal/technologies

[10] Miss Nidhi G. Gupta, Dr. S. A. Ladhake "FACE DETECTION AND FACIAL EXPRESSION RECOGNITION SYSTEM USING 2DPCA" International Journal For Technological Research In Engineering Volume 2, Issue 7, March-2015.

[11] Devi Arumugam and Dr. S. Purushothaman "Emotion Classification Using Facial Expression". (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 7, 2011.

[12] S. Sharmila, Dr. A. Kalaivani," AUTOMATIC FACIAL EMOTION ANALYSIS SYSTEM FOR STUDENTS IN CLASSROOM ENVIRONMENT" International Journal of Pure and Applied Mathematics.

[13] KRISHNA MOHAN KUDIRI, ABAS MD SAID AND M YUNUS NAYAN," HUMAN EMOTION DETECTION THROUGH FACIAL EXPRESSIONS" Journal of Theoretical and Applied Information Technology 30th June 2016. Vol.88. No.3.

Face Recognition using Raspberry PI

Shruti AnantAmbre Mamata VijayMasurekar Shreya Shivaji Gaikwad

Department of Information Technology St. Francis Institute of Technology, Mumbai-400103, India
shrutiambre225@gmail.com, masurekarm@gmail.com, shreyagaikwad99@gmail.com

***Abstract*—In an age where public security is a priority, there is a growing need for autonomous systems capable of monitoring hotspots to ensure public safety. Face recognition technology could create an increased level of security, enable businesses and governments to save money on CCTV monitoring staff, and increase business productivity by automating processes such as attendance monitoring. In recent decades, such a system would have been unfeasible to implement due to cost and technological restraints. This study aims to explore a real-time face recognition system using easily-attainable components and libraries, such as Raspberry PI and Dlib, Face Recognition library and Open Source Computer Vision Library (OpenCV). It also covers various face recognition machine learning algorithms.The results show that in real-time applications, the system runs at 2 frames per second and recognizes faces despite Raspberry PI's limitationssuch as low CPU and GPU processing power.**

Keywords—Face recognition, Haar cascade, OpenCV, Python, Raspberry PI.

I. INTRODUCTION

The temporal lobe of the brain is responsible for face recognition in humans and is an essential part of the human perception system. Similarly, in machine learning systems, facial recognition is a technology capable of autonomously identifying or verifying a person from a digital image or video in real-time. This is accomplished by comparing and analyzing patterns, such as contours and facial features [1]. Compared to other methods of identification such as physiological biometrics like fingerscan, irisscan and behavioral biometrics like voicescan, signaturescan, face recognition have the advantage of its non-invasive and user-friendly nature. A real-time face recognition system can be scrutinized from almost any location that has a computer with an internet connection. Facial images can be recorded and recognized from a distance without interacting with a person, which is particularly beneficial for security and surveillance purposes. Practical applications such as real-time crowd surveillance, criminal identification, human-computer interface, and the prevention of unauthorized personnel from accessing restricted areas, etc. are made possible with face recognition. Furthermore, for government security surveillance, personal information such as name, address, criminal record, etc. can be obtained by further analyzing recognition results. Face recognition can also be used for generic purposes in institutions such as schools, shopping malls and other public and private enterprises for access control and as a content-based database management system.

Face recognition is a two-part procedure, face detection must precede face recognition. Real-time face detection was made attainable by the work of Viola and Jones [2]. In this paper, Haar cascade algorithm proposed by Viola and Jones, histogram of oriented gradients (HOG) and Linear Support Vector Machines (SVM) are used for face detection and Geitgey's Face Recognition library for Python and command line, OpenCV and Dlib were used for face recognition.The algorithms can be used with relatively low-cost hardware, i.e. Raspberry PI and Raspberry PI (RPI) camera. The proposed system was cost-effective and reliable.

The paper is structured as follows: Section 2 covers literature review on substantial findings within face recognition. Section 3 focuses on the system design, components and its operation. In Section 4, results and analysis have been noted, and finally, section 5 concludes the study.

II. LITERATURE REVIEW

Ishita Gupta et al. [3]proposed a system for face detection and recognition using Raspberry PI, Principal Component Analysis (PCA) and the Eigenface approach. For detecting faces, Haar feature-based cascade classifier is used which uses positive and negative images (images with and without faces, respectively) to train the classifier. The PCA is an algorithm that reduces number of (possibly) correlated variables into smaller number of uncorrelated variables. While capturing and training, various images of positive and negative types are created. To display results, LCD as well as command terminal are used. LCD shows the name of a person whose face is detected with 'present' status, and in the terminal window, their name is printed with a 'presentOk' status.

Ali Akbar Shah et al. [4] proposed a system where an LED glows if a face is detected within a specific range. Haar classifier, developed by the Viola-Jones algorithm is used for face detection. The Eigen features of the face for tracking its position are detected using MATLAB and Raspberry PI. This system is using a camera and LED on the array of LEDs. The LEDs will glow where the face is located and it will track the face within its commanded limit. The frame is given as an input to Haar classifier and further it is given to Eigenface where Eigen features detects eyes, nose and mouth. This information is passed to the geometric transformation where without modifying the color of the image, the adjustment of pixels takes place. By specifying the LED (ON) condition, the LED will glow if the face enters a specific range. But this system is limited for single person's face only.

Ayman A. Wazwaz et al. [5] proposed a system for optimizing face detection and recognition in real-time using Raspberry PI, OpenCV and Python. The faces in frames sent by the camera to Raspberry PI are detected using the Boosted Cascade of Simple Features (BCOSF) algorithm, cropped by Python-based program and the cropped faces are routed to one of the computers (Linux based servers) connected by TCP/IP based over an Ethernet connection. The cropped faces are recognized using the Local Binary Pattern Histograms (LBPH). If any face is unfamiliar and the recognition system is unable to recognize it, the face is passed on to another computer. But as the image size reduces, the quality of image reduces which results in increase in the average error rate. Factors like

lightning, distance and camera resolution affect the results of system. The more servers, the more accurate the result will be. The cost and security between servers should be considered while designing the system.

Priya Pasumarti et al. [6] used Raspberry PI, OpenCV, Haar cascade, LBPHRecognizer, and Viola-Jones framework for face detection and recognition. The total system is divided into 3 modules - database creation, training the dataset, testing, sending alert messages as an extension. In the first part the database has been created by taking the input image, converting the images to grayscale and then storing it with an ID. In the training phase LBPH face recognizer is initialized, then the faces and IDs are trained using LBPH and are later saved as an xml or yml file. In the testing phase, the system has used Viola Jones algorithm for face detection. In this algorithm, each feature is represented as a single value obtained by subtracting the summation of white rectangles from black ones. As the number of classifiers increase, it becomes more complex. This has been overcome by using integral images. As the classification takes place in stages, the regions that pass through all the classifiers are detected as faces and local binary pattern algorithm is used for face recognition. Adaboost machine learning algorithm has been used to cascade the classifier and thus increase the efficiency.

Umm-e-Laila et al. [7]proposed a system uses Raspberry PI, Python,C++, Java, MATLABetc. for implementation. The program is designed to use all the algorithms of OpenCV for greater efficiency and speed. The system graphical user interface (GUI) has been designed to detect and recognize faces in real-time using a webcam. After the camera is activated the system uses algorithms such as LBPH to detect the faces in real-time. The system GUI includes the option for choosing a specific algorithm before starting face detection. It has been concluded that Raspberry PI eliminates the machine dependencies, and LBPH algorithm gives better results in terms of accuracy, while Fishface algorithm provides better results in terms of time consumption. It has also been observed that most of the algorithms perform similarly because of Raspberry PI's processing power.

III. PROPOSED SYSTEM

A. *System Architecture*

The system design in Figure 1 consists of PI camera module that will be used for real-time video streaming, capturing input frames and for forwarding the frames to the Raspberry PI. The Raspberry PI will detect and recognize the faces in the frame with the resultant output being shown in a real-time video stream on connected devices. The Raspberry PI is connected to an ethernet and the devices to Wi-Fi or cellular data.

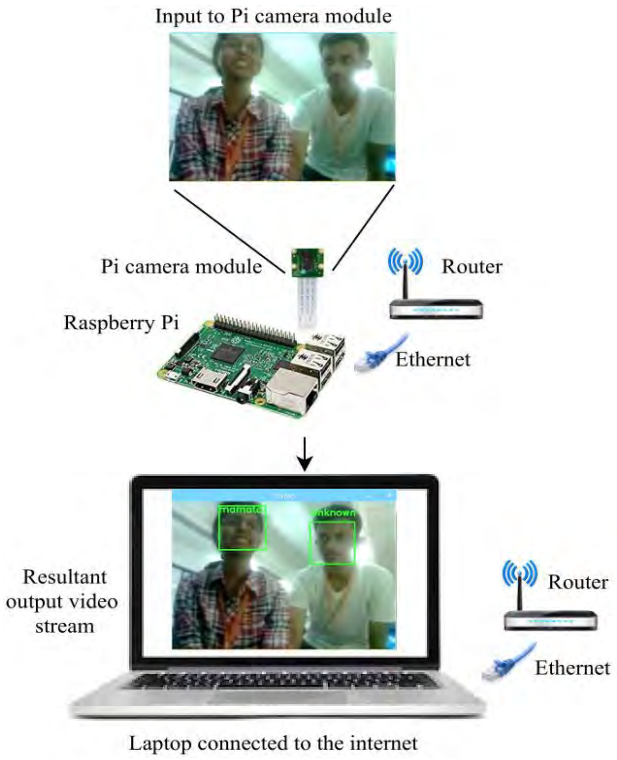


Fig. 1. System design

The system workflow is as follows:

- RPI camera connected to the Raspberry PI live streams video.
- The Raspberry PI detects faces from captured frames, computes face embeddings, and compares the vector output to the known database and labels the most matched face based on k-NN classification
- The Raspberry PI and devices are connected over a virtual network using Virtual Network Computing (VNC), where the Raspberry PI acts as a server and the devices as clients. The Raspberry PI should be connected to an ethernet connection and the devices must be networked with TCP/IP.
- Devices connected to the Raspberry PI can see the labels for recognized faces in a real-time video stream.
- If the face is not recognized, an unknown label is shown for the detected face.

B. *Components*

1) *Hardware*

Includes Raspberry PI 3 b+ model, Raspberry PI camera module v1, ethernet cable and power supply.

a) *Raspberry PI 3 b+ Model*

Raspberry PI is a small and affordable computer that can be plugged to a monitor or an LCD and can be used with a keyboard and a mouse. It can be used to build devices that are Internet of Things (IOT) or sensor based. Raspberry PI is mainly used for projects that don't require much processing power or storage space.

PI 3 Model B makes use of 1.2 GHz 64-bit Quad Core Processor and consists of 1GB RAM. It has an onboard Wi-Fi and Ethernet Port and also consists of one microSD slot.

b) Raspberry PI camera module v1

The Raspberry PI camera module can record videos and capture still images. The camera module is compatible with all Raspberry PI models.It has a native resolution of 5MP. The camera module supports a sensor resolution of 2592 x 1995 pixels and 1080p30, 720p60 and 640 x 480p60/90 video modes.

2) Software

Python language was used for programming face recognition.

a) Haar cascade classifier

Viola and Jones proposed a machine learning object detection algorithm known as Haar cascade, to identify objects in images and videos based on the concept of features [2]. The cascade function is trained on a huge dataset of negative and positive images. For face detection, the algorithm requires positive images that contain faces and negative images that don't contain faces to train the classifier, which then is used to extract the features from.

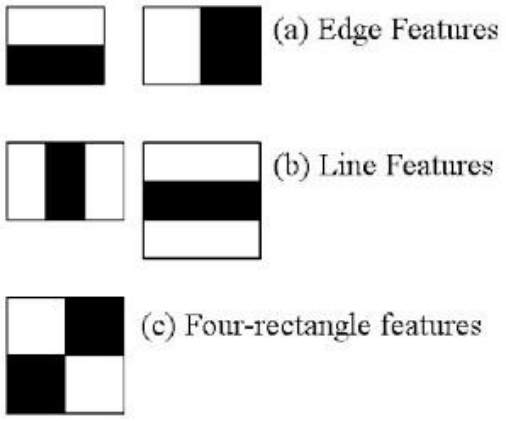


Fig. 2. Haar classifier patterns [8]

Every feature is a single value obtained by calculating the difference between the sums of pixel intensities in each region. Due to the quantity of features used, integral image concept is applied to prevent an increase in computation time. It also helps to simplify pixel calculations. Adaboost is used for feature selection, reducing the complexity of classifiers and training them. After which each feature is applied on the training images and for every feature, the best threshold is found to classify the faces as positive or negative. However, since it is inefficient and time consuming, cascading classifiers are used to obtain the best features from a face by grouping the features into a stage of classifiers and processing or discarding the face region according to the stages of features passed.

b) OpenCV

Open Source Computer Vision Library (OpenCV) [9] is a machine learning library with programming functions for real-time computer vision and image processing. OpenCV library consists of many built-in packages for face recognition. The library includes linear and non-linear image filtering, geometric image transformations, changing color spaces, smoothing images, image thresholding, histograms, and so on. OpenCV library includes functions for algorithms such as Haar classifier, Histogram of oriented gradients (HOG), Eigenfaces, Fisherfaces and Local Binary Patterns Histograms (LBPH).

c) Histogram of oriented gradients

The Histogram of Oriented Gradients is a feature descriptor used in image processing for object detection. The HOG descriptor technique counts occurrences of gradient orientation in localized portions of an image - detection window, or region of interest (ROI).

The HOG descriptor algorithm is as follows:

- Divide the image into small connected regions called cells, and for each cell compute a histogram of gradient directions or edge orientations for the pixels within the cell.
- Discretize each cell into angular bins according to the gradient orientation.
- Each cell's pixel contributes weighted gradient to its corresponding angular bin.
- Groups of adjacent cells are considered as spatial regions called blocks. The grouping of cells into block is the basis for grouping and normalization of histograms.
- Normalized group of histograms represents the block histogram. The set of these block histograms represents the descriptor [10].

d) Support Vector Machine

Support Vector Machine (SVM) is a machine learning algorithm used for classification and regression. It uses supervised learning to group the data into two categories. It is trained with a collection of categorized data. The aim of SVM algorithm is to determine which category a new data point belongs to. It should not only classify the data but also draw a margin between the two categories as wide as possible.

e) k-Nearest Neighbor

k-Nearest Neighbor (k-NN) is known for being a lazy learning algorithm and is used for classification and regression predictive problems. k-NN is easy to interpret and has low calculation time. k-NN classifier, classifies test points by finding the most similar class among the k-closest examples. "k" is the number of training set items that are considered for classification. Larger k gives smoother boundaries and better generalization but it is important that the locality is preserved.

$$d(p,q) = d(q,p)$$
$$= \sqrt{(q_1 - p_1)^2 + (q_3 - p_3)^2 + \cdots + (q_n - p_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$$

Fig. 3. Euclidean distance [11]

Distance metric functions such as Euclidean distance or Manhattan distance are commonly used to apply the k-NN algorithm. These functions are used to compute the distance between input sample and the training samples. The distance calculated is sorted and the nearest neighbors are determined on the k-th minimum distance. The categories of the nearest neighbors are gathered, and the category that is the majority is selected as the predicted value of the input sample. As k-NN has no training stage, test stage is computationally expensive. It is generally used applications such as pattern recognition, gene expression, measurement of document similarity.

f) DlibLibrary

Dlib[12] is an open source toolkit built using C++ containing machine learning algorithms like classification, clustering, regression, data transformation, structure prediction. The designs are modular, easily executable, and can be implemented by using C++ API. It is used solve real world problems, both in industries and academia. It has its application in the domain of robotics, embedded devices, face recognition. Deep matrix learning tool of Dlib can be used for face recognition. It is a pretrained ResNet model that recognizes faces with accuracy of 99.38%.

g) Face RecognitionLibrary

Face Recognition [13] is a Python programmed library that can be used for recognizing faces. It is built using Dlib's state of art face recognition. This library also provides command line tools. It can be used for finding faces, manipulating facial features and for identifying faces in pictures.

C. Working

Many algorithms can be used for face detection and face recognition. The algorithms proposed in this paper for face detection are Haar cascade, HOG and Linear SVM. For face recognition, the system will make use of Geitgey's Face Recognition Library for Python, OpenCV and King's Dlib library. The open source Dlib library, OpenCV library and Face Recognition library for Python contain built-in face recognition algorithms that have been used in this system.

The system consists of three subsystems - database creation, dataset training, and testing.

a) Database creation

Creation of a folder containing images of a person. The folder name should be the name of the individual whose photos are contained within.

b) Dataset training

As per the flowchart in Figure 4, Dlib's HOG and Linear SVM algorithms are used for face detection. For every image in the dataset, the person's name is extracted from the path, the image is converted to RGB, faces are localized in the image using HOG and Linear SVM and face embeddings are computed with Dlib's deep metric network. A pickle with 128-d embeddings for each face in the dataset along with their names is created.

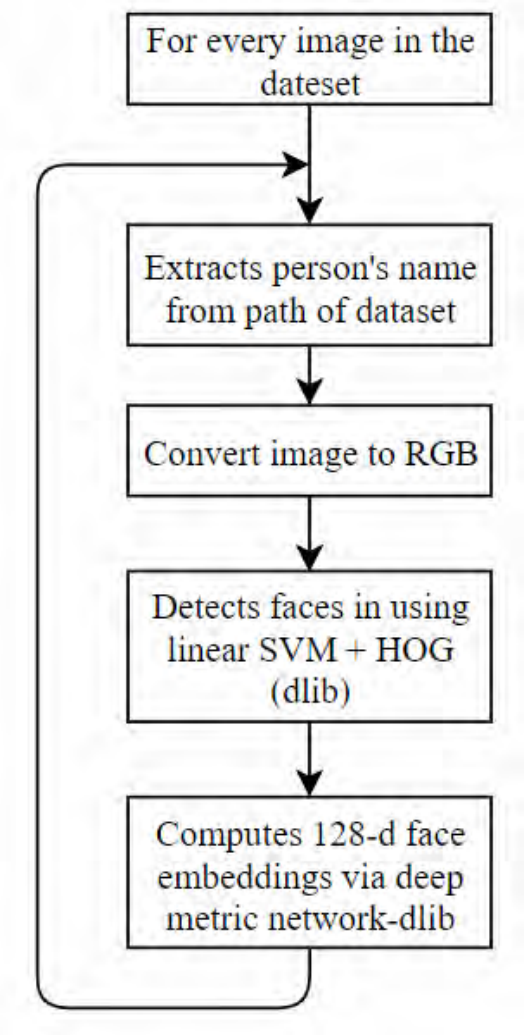


Fig. 4. Flowchart for dataset training

c) Testing

RPI camera module video streams to connected devices in real-time. A frame is captured and the frame is then preprocessed. OpenCV's Haar cascade method is used to localize and detect faces for which the frame is preprocessed by resizing and converting it to grayscale. OpenCV is then used for bounding box. The frame is then converted to rgb for face recognition. Face Recognition library is then used to compute facial embeddings for each face in the bounding box. The computed facial embeddings are then compared to the facial embeddings of the dataset and based on k-NN classification, if matches are found, number of votes are given to each name and the name with the most corresponding votes is selected. The predicted name is then used to label the faces in the frame. The resultant output is video streamed in real-time for other connected devices to view. See Figure 5.

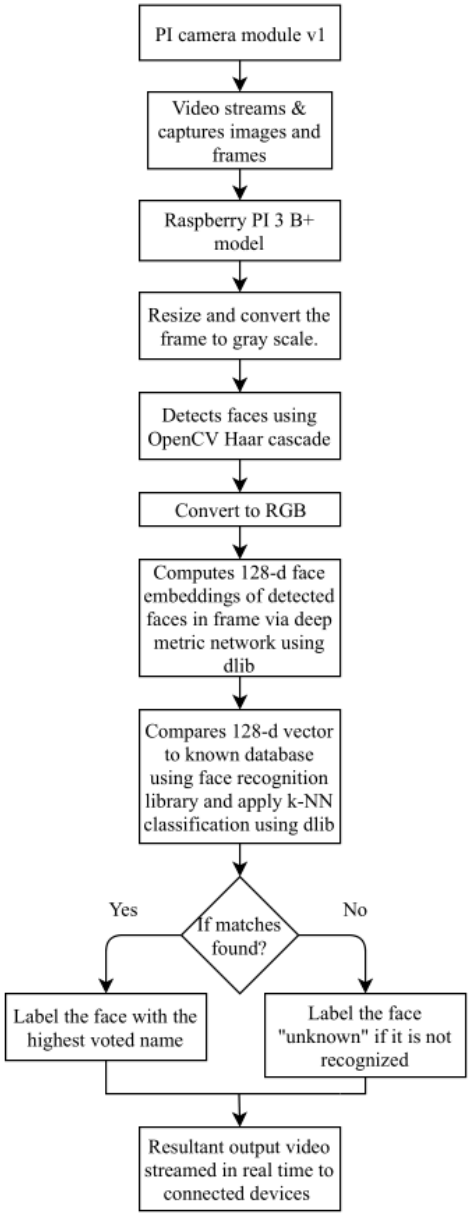


Fig. 5. Flowchart for testing

IV. RESULTS AND DISCUSSION

The resulting figures show the output. In Figure 6, Mamata, Shreya and Shruti's facesare detected accurately.

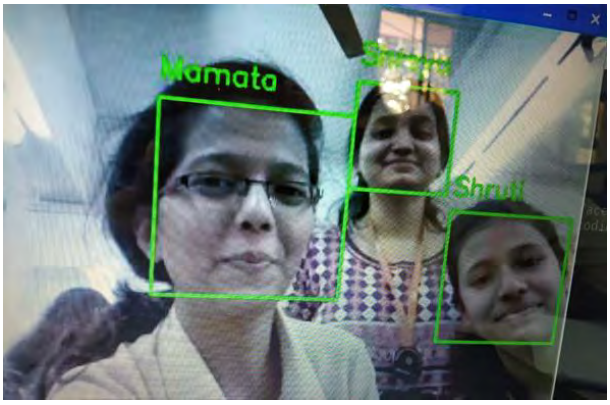


Fig. 6. Output showing recognized faces

In Figure 7, Mamata's face is detected and the person whose photos are not in the dataset is labelled as unknown.

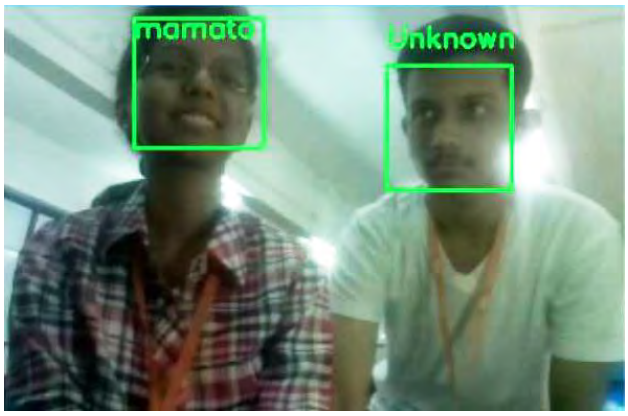


Fig. 7. Output showing recognized and unrecognized faces

Because of Raspberry PI's processing and GPU power limitations, the frame rate was noted to be 1-2 FPS. Due to the low framerate, an error causing inconsistencies in facial recognition can be seen in Figure 8.

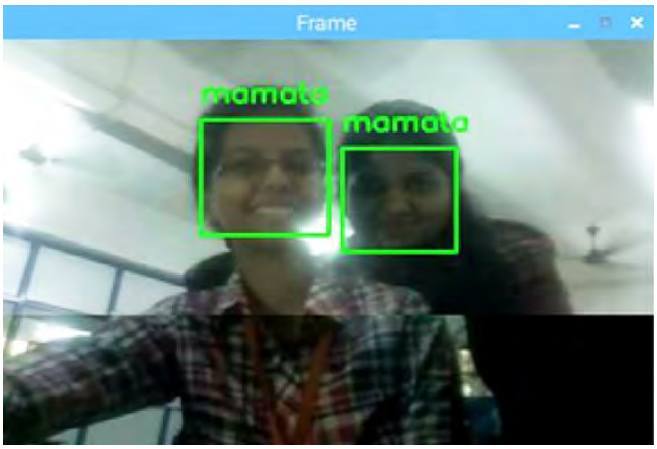


Fig. 8. Output showing error

V. CONCLUSION

The system was able to recognize the faces in the video stream in real-time and thereby, labeled them correctly. The desktop interface of Raspberry PI was remotely accessed from multiple devices through Virtual Network Computing (VNC). The system was cost effective and reliable. However, while it can be highly accurate, one can only obtain aframes per second (FPS) rate of 1-2. The Raspberry PI while powerful for such a cheap and small device, is limited in terms of processing power, memory and especially without GPU power. Factors such as lighting, camera resolution and distance affected the face recognition process as well. Face recognition is the first step in many of its applications such as expression analysis, attendance monitoring, security systems, surveillance and human computer interface.

ACKNOWLEDGMENTS

We would like to thank our mentor, Prof. Vivian Lobo for his support and encouragement. We would also like to thank Mr. Thomas Calvert forhis diligent proofreading,which greatly improved the research work.

REFERENCES

[1] “What is facial recognition? - Definition from Techopedia”, [online] Available: https://www.techopedia.com/definition/32071/facial-recognition. [Accessed on February 08,2019]

[2] P. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features", Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, Dec. 2001.

[3] Ishita Gupta, Varsha Patil, Chaitali Kadam and Shreya Dumbre, “Face detection and recognition using Raspberry PI”, 2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), Pune, India, pp. 83-86, Dec. 2016.

[4] Ali Akbar Shah, Zulfiqar Ali Zaidi, Dr. Bhawani Shankar Chowdhry and Dr. Jawaid Daudpoto, “Real-time face detection/monitor using Raspberry PI and MATLAB”, 2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), Baku, Azerbaijan, Oct. 2016.

[5] Ayman A. Wazwaz, Amir O. Herbawi, Mohammad J. Teeti and Sajed Y. Hmeed, “Raspberry PI and computers-based face detection and recognition system”, 2018 4th International Conference on Computer and Technology Applications, Istanbul, Turkey, May 2018.

[6] Priya Pasumarti , P. Purna Sekhar, “Classroom attendance using face detection and raspberry-pi”, International Research Journal of Engineering and Technology (IRJET), vol. 05, pp. 167-171, March 2018.

[7] Umm-e-Laila, Muzammil Ahmad Khan, Muhammad Kashif Shaikh, Syed Annas bin Mazhar and Khalid Mehboob, “Comparative analysis for a real-time face recognition system using Raspberry PI”, 2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA 2017), Putrajaya, Malaysia, November, 2017.

[8] “OpenCV: Face Detection using Haar Cascades”[online], Available: https://docs.opencv.org/3.4.3/d7/d8b/tutorial_py_face_detection.html [Accessed February 08,2019]

[9] “OpenCV library”[online], Available: https://opencv.org/, [Accessed on February 08,2019]

[10] "Histogram of Oriented Gradients (HOG) descriptor”[online], Available: https://software.intel.com/en-us/ipp-dev-reference-histogram-of-oriented-gradients-hog-descriptor [Accessed on February 08,2019]

[11] “Euclidean distance” [online], Available: https://en.m.wikipedia.org/wiki/Euclidean_distance [Accessed on February, 08 2019]

[12] ”dlib C++ Library” [online], Available: http://dlib.net/ [Accessed on Februrary 08,2019]

[13] “GitHub - ageitgey/face_recognition: the world’s simplest facial recognition api for python and the command line”, [online] Available: https://github.com/ageitgey/face_recognition, [Accessed on February 08,2019]

[14] “Raspberry Pi Face Recognition - PyImageSearch”[online], Available: https://www.pyimagesearch.com/2018/06/25/raspberry-pi-face-recognition/ [Accessed on February 08,2019]

Transport Vehicle Selection Predictor

Shubham Matale[1] Apurva Taru[2] Pratik Mahamuni[3] Prof. Sukhada Bhingarkar[4]
[1,2,3,4]*Department of Computer Engineering*
MIT College Of Engineering
Pune, India
[1]mataleshubham@gmail.com [2]apurvataru8855@gmail.com [3]pratikraj.mahamuni@gmail.com
[4]sukhada.bhingarkar@mitcoe.edu.in

***Abstract*—Vehicle Selection Predictor is a new and important concept in transportation studies. In recent years, prediction model was designed for the prediction of prices of vehicles. In this project, we are trying to build a prediction model for vehicle selection based on its usage. In this project, we consider the problem of vehicle selection for transportation using a multi-criteria decision-making approach. The problem includes several conflicting factors which are economic and technological factors. The vehicle has its own specifications/factors like load that it can handle, average of the vehicle, design of vehicle etc. Based on these factors the transport agency should choose a vehicle for a particular delivery. Hence, to predict the selection of vehicle depending on its specifications, the system will be designed by using machine learning algorithms. While designing the model feature selection is a very important aspect. This paper mainly focuses on feature selection and building the predictive model.**

Keywords— Classification Algorithm, Feature Selection and Eval-uation, Machine learning, Predictive model

I. INTRODUCTION

Vehicle selection predictor is an act of predicting the suitable vehicle for the required route for the transport agency. In the transport system, choosing a vehicle for a particular route is a difficult task. As there are many options available while choosing a vehicle, it makes confusing to choose a suitable vehicle.

All the transport management responsibilities, vehicle selection, and allocation are sure to present one of the greatest conundrums. There can be some complicated factors to be consider like the load capability, engine power, fuel capacity of the vehicle, etc. Achieving the goal of getting the best value up front and throughout a vehicles service life while providing the right equipment for the job. By considering this, the system will try to reduce the efforts required for the selection of vehicles with the help of Machine Learning Algorithm.

Transport Management Systems (TMS) are gradually extended with new features to improve reliability, such as planning vehicle routes. This project will be an improvement over the existing TMS, which will provide the vehicle selection based on machine learning. It will truly automate the transport planning, as the present systems are not effectively able to plan trips of vehicles according to their full efficiency.

Prediction of vehicle depends on various vehicle specifications such as the load capacity of the vehicle, the power of the engine etc.[1]. The system will be designed by using machine learning algorithms. The purpose is to select the vehicle from all other available vehicles for a respective route, by comparative study of popular classification and prediction techniques. This work gives a brief introduction towards techniques, application, and challenges of the prediction system. This project is as important to the individual as much as to the public too.

In this paper, section II overviews previous work done related to transportation engineering and machine learning. Section III discusses problem solution and it gives an idea about overall proposed system's architecture. Section IV discusses about methodologies used. Finally, section V is about conclusion and future work.

II. RELEVANT WORK

In [1], the author explained various factors affecting transportation. The performance, design, and operation of the vehicle are affected by various factors such as a human, vehicle, acceleration characteristic, breaking performance etc. These factors influence the geometric design and design of control facilities. Variant nature of the vehicle and road surface have importance for the smooth and efficient performance as well as the smart planning of the transportation vehicles.

In 2014, R. Prytz, S Nowaczyk, T Rgnvaldsson, S Byttner investigated unsupervised and supervised methods for predicting vehicle maintenance [2]. These methods are data-driven and use big amounts of data which may be streamed, onboard data or historical data from the off-board database. These methods depend on telematics gateway that enables vehicles to communicate with the back-office system.

Author Serhat Aydn and Cengiz Kahraman considered the problem of bus selection for public transportation using a hybrid multi-criteria decision-making approach [3]. The problem includes various contradictory factors such as social, economic and technical factors

The author brings integrated approach of fuzzy analytic hierarchy process (AHP) and simplicity of fuzzy VIKOR methodology together. To test the methodology, a case study of Ankara, the capital of turkey is given. In this system, the four-level hierarchy was established along with three experts. These experts are utilized for assessing the pairwise comparison matrices. The weights of criteria were deter-mined using fuzzy AHP and then alternative ranks by fuzzy VIKOR. Sensitivity analysis is also made to see how sensitive decision predicted. It helps to change in parameters of methodology. Buckley's fuzzy approach is finally implemented to solve the problem.

In 2014 author Visalakshi, S., Radha, V. studied basic concept of feature selection, the process of feature selection, evaluation criteria and various approaches used in feature selection which are applicable to all databases [4]. From the state of art, feature selection methods are flexible and capable of providing a solution to any kind of problem faced when performing feature selection. From the literature study, instead of using filter and wrapper methods separately, better embed both methods for selecting relevant features, improves in classification accuracy, speed up the process and reduce the error rate. Still, there is a scope that genetic algorithm and ensemble (filter and wrapper) are capable of handling multi-dimensional dataset.

In the article [5] author explained how different methods shows different variables are important, or at least the degree of importance changed. The result need not be conflict, because each method has a different approach and perspective for how a variable can be useful.

III. PROPOSED SOLUTION

The problem of vehicle selection for transportation is solved by us-ing a multi-criteria decision-making approach. The problem includes several conflicting factors which are economic and technological factors.

Fig. 1 shows the basic architecture of the proposed system.It describes modules such as feature selection,separation of training and testing dataset etc. It gives an insight into what are the different modules of the system and how they interact with each other.

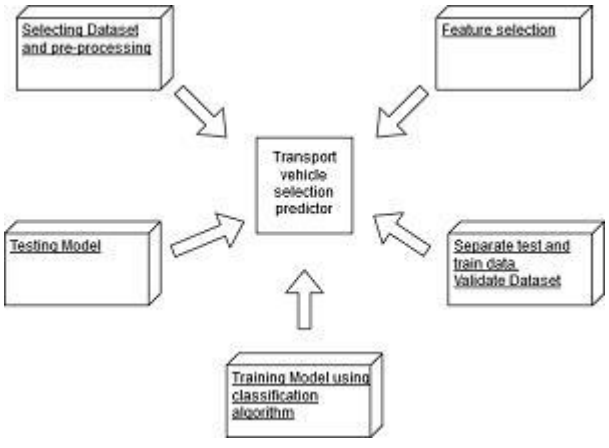


Fig. 1. Block Diagram

In accordance with the problem statement, the proposed system will train the model to classify the various vehicles

Company Name	Fuel Tank Capacity
Model Number	Turning Radius
Engine Type	Max Speed
Engine Cylinder Count	Wheelbase
Displacement	Overall Vehicle length,
Max Power	Width, Height
Max Torque	Ground Clearance
Transmission Type	Payload
Gearbox	Type of Location

based on various parameters and predict the suitable vehicle for a particular condition. The proposed system will act as a decision support system to the user who needs to select the vehicle. This system will automate the task of selecting the vehicle taking into consideration parameters like location, load type, load capacity etc. As shown in fig. 1, all the modules are explained as follows:

A. Dataset Selection and Preprocessing

In this phase, data of various transportation vehicles will be collected with the following attributes:

Various data pre-processing methods such as Data Cleaning, Inte-grating, Transformation, Reduction etc will be applied to generated data[6].

B. Feature Selection

Feature selection becomes more important when there is a large set of features available. It is not necessary to use every feature for the system. It is necessary to assist algorithm by providing only those features that are really important. Selected feature subsets give better results than a complete set of features for the same algorithm. The methods used for feature selection is wrapper methods which contain two sub-methods.

Wrapper methods: In wrapper methods, a subset of features is used for training the model. Based on the inferences that are drawn from the previous model, it is decided to add or remove features from the subset. The problem is essentially reduced to a search problem. These methods are usually computationally very expensive .

Types of Wrapper Methods are:

Forward Selection: Forward Selection starts with no variable in the model. For all predictors which are not in the model, check their p-value if they are added to the model. Choose the one which affects the result. This method is repeated until no new variables added to the model [4].

Backward Elimination: Backward elimination starts with all the features and removes the least significant feature at each iteration which improves the performance of the model. This process is repeated until no improvement is observed on the removal of features [4].

Recursive Feature Elimination: It is a greedy technique which aims to find the optimal solution i.e. best performing feature subset. It continuously creates models and separates the best performing or worst performing features at each iteration. It creates the model with the remaining features until all the features are exhausted. After that, it ranks the features based on the order of elimination of features [4]. Fig. 2 shows, the working of forward selection and backward elimination.

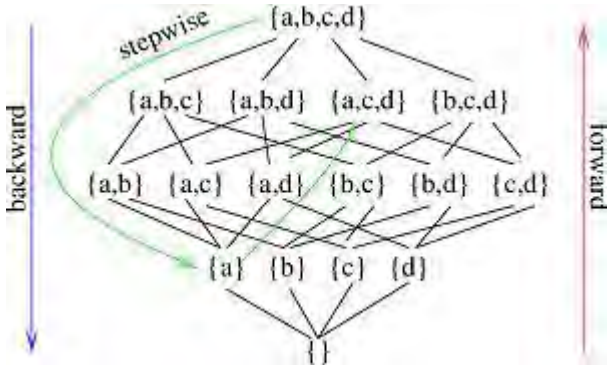


Fig. 2. Feature Selection Methods

For example, if there are 4 features such as [a,b,c,d]. In forward selection initially it start with no feature set. In first step, it adds one feature and checks for accuracy [a], [b] etc. In next step it adds another feature and check for accuracy with various combination [a,b], [b,c] etc. The combination with most accuracy is selected and another feature is added to check the accuracy [a,b,d], [b,c,d] etc. In similar manner the process is repeated untill the set of best features with good accuracy is achieved [a,b,c,d]. In contrast backward elimination starts with all the features initially and eliminates features one by one according to the accuracy contribution of that particular feature.A step one contains[a,b,c,d]. In further step it contains [a,b,c], [b,c,d] etc. In similar manner steps are repeated to select best features set.

C. Dataset Validation

Validation techniques are used to lower the error rate of the machine learning (ML) model. It is more effective when dataset available is too less. Two proposed methodologies for dataset validation are as follows.

K-Fold Cross-Validation :

In this method data is split into k subsets of equal size by random sampling. In the next step, each subset in turn is used for testing and the remainder for training. The advantage is that all examples are used for both training and testing. The error estimates are averaged to yield an overall error estimate [7]. Error rate is derived from equation 1

$$E = 1/K \sum_{i=1}^{K} E(i) \quad (1)$$

Where,
E is error rate.
k in number of subsets.

In fig. 3, the colored part is for testing and remaining is for training purpose. For example, consider initially data is divided in 4 subset. In each experiment one subset is used for testing and remaining for training.

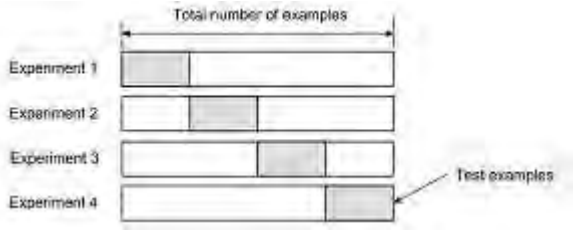


Fig. 3. K-Fold Cross-Validation Method

Leave-One-Out Cross-Validation (LOOCV)

If the dataset has N examples, then N experiments to be performed for Leave-one-out cross-validation. For every experiment, training uses N-1 for training and one example is use for testing . The average error rate on test examples gives the error[7]. Error rate is derived from equation 2

$$E = 1/N \sum_{i=1}^{N} E(i) \quad (2)$$

This method is generally preferred over the K-Fold Cross-Validation because it does not suffer from the intensive compu-tation, as number of possible combinations is equal to number of data points in original sample or n.

As shown in fig. 4, a single data point for testing and remaining is used for training purpose. This is repeated for all experiments. Then the error is averaged for all the experiments to give overall efficiency.

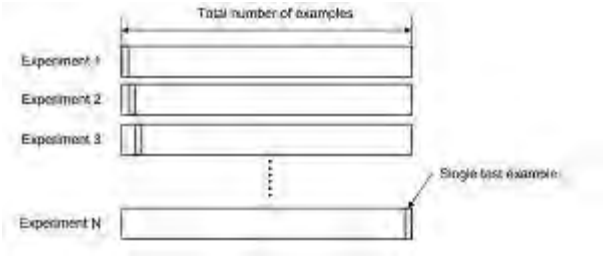


Fig. 4. LOOCV Method

D. Training Model

In machine learning, classification is a supervised learning method. In this method the computer program learns from the data input given to the algorithm and then uses this program to classify new observation.

The first algorithm deals with clustering.

1. K - Nearest Neighbour

K - Nearest Neighbors is used in the field of pattern recog-nition. It learns by analogy, i.e by comparing a given test tuple with training tuples that are similar to it. The training tuples have n

attributes, every tuple represents a point in n-dimensional space. All training tuples are stored in n-dimensional patterns.

K-Nearest Neighbors searches the pattern space for the k training tuples that are closest to the unknown test tuple. Closeness is defined using distance metrics such as Euclidean distance [8].

Step 1: A positive integer k is specified, along with a new sample. (k is the number of nearest neighbors).

Step 2: Select the k entries in the database which are closest to the new sample (calculate the distance between those points).

Step 3: Find the most common classification of these entries (one having the shortest distance).

Step 4: Assign that classification to the new sample.

In Fig. 5, the test sample (which is inside the inner dotted circle) should be classified either to the first class of circle or to the second class of triangles.

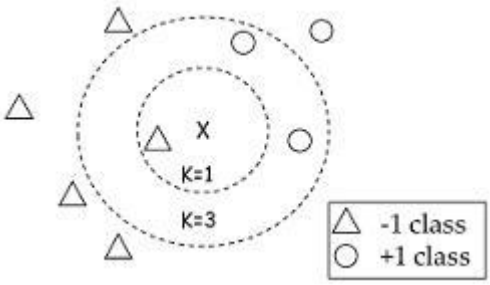


Fig. 5. K Nearest Neighbour

Where,

E is error rate,

N is number of examples.

the inner dotted circle. For example, if k = 5 then it is assigned to the first class (1 circles versus 4 triangles outside the outer dotted circle).

2. Neural Network

Neural network consists of many simple processing units, which are wired together in a complex communication network as a human brain. Each of the processing unit is called a neuron. It consists of input signals, weights assigned to each of the input, processing function f which computes the summation of weighted input and output signals [9] .The basic structure of a neuron is as shown in fig 6.

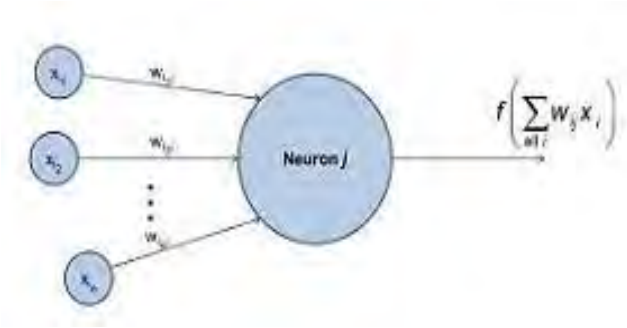


Fig. 6. Simple Neuron

The steps involved in the algorithm are :

The n inputs are denoted by x(1),x(2)...x(n).Weights are denoted as w(1),w(2)...w(n).

The input values are multiplied by their weights and those are summed as equation 3

$$v = \sum_{i=1}^{n} w(i)x(i) \quad (3)$$

The output is some function y = f(v) of weighted sum.

IV. DISCUSSION

There are overall 17 features of dataset but some of them are not necessary so wrapper methods are applied to select the features. During the experiment, among the 3 wrapper methods, backward elimination has given the better feature set where label is Location Type. Intracity, Intercity, Hilly, Highway are four classes of location type. Intracity represent the particular vehicle is best for within a city or minimum distance transportation, similarly Intercity for two cities. Hilly shows the mountain roads transportation, and Highway is for long distance smooth roads.

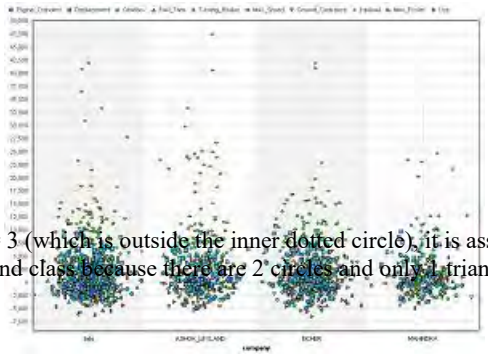


Fig. 7. Scatter Diagram of Selected Feature Set

Fig. 7 shows the scatter diagram of selected feature set. Scatter diagram represent the distribution of feature dataset along the x-axis which represent the company name and y-axis represent the values of features.The selected attributes are as follows:

Engine Cylinder Count	Max Speed
Displacement	Ground Clearance
Gearbox	Payload
Fuel Tank Capacity	Max Power
Turning Radius	Use

Validation of dataset is important aspect for good accuracy of model. Among the LOOCV and k-fold validation, K-fold cross validation is exhaustive in the sense that it needs to train and validate the model for all potential combination and for moderately large number of subsets, which becomes computationally infeasible. In contrast, number of possible combinations in LOOCV is equal to number of data points in original sample hence it does not suffer from the intensive computation so LOOCV has better performance. It also leads to higher variance and get estimates of test error with lower bias. Therefore LOOCV is to be preferred over K-Fold Cross Validation for validation.

If k = 3 (which is outside the inner dotted circle), it is assigned to the second class because there are 2 circles and only 1 triangle inside

In this paper, two classification algorithms are used; the K-Nearest Neighbor(KNN) and Neural Network(NN). Cross validation is used to classify and validate the dataset. The accuracy output of the both classifier is 67.49% and 72.71% respectively. KNN is computationally expensive as it need to store all the training data, whereas in neural network once model is trained, the training data is no longer needed to produce new predictions.

Table 1 represents the performance matrix for the Neural Network classifier.

TABLE I
PERFORMANCE MATRIX

	Intercity	Intracity	Highway	Hilly	C. Precision
P. Intercity	32	9	5	0	69.57%
P. Intracity	5	23	1	0	79.31%
P. Highway	5	1	80	21	74.77%
P. Hilly	0	0	11	19	63.33%
C.Recall	76.19%	69.70%	82.47%	47.50%	63.33%

C = Class, P = Predicted

It gives overall accuracy of 72.71%. Label is location type for that particular vehicle. Hence algorithm will predict the location type for the vehicles and as per the requirement, suitable vehicle from available dataset will be selected.

V. CONCLUSION AND FUTURE WORK

The results are discussed on the comparison of classification of the vehicle using the KNN and NN. The results were observed through the percentage of accuracy of the KNN and NN. In conclusion NN gives better result with backward elimination for feature selection and LOOCV for dataset validation compared to KNN for the system. The proposed system briefly examine the potential use of classification based machine learning techniques to solve industry problem of vehicle selection.

Future work includes the improvement of the system's performance as well as other modules for the transport management system will be added like minimum distance route plan, ERP etc. Also in future maintenance of vehicle will also be considered as an attribute for classifier building.

REFERENCES

[1] Factors affecting transportation, by NPTEL.

[2] Machine learning methods for vehicle predictive maintenance using off-board and on-board data, Halmstad University. Serhat Aydn , Cengiz Kahraman,

[3] Visalakshi, S., Radha, V. (2014). A literature review of feature selection techniques and applications: Review of feature selection in data mining.

[4] 2014 IEEE International Conference on Computational Intelligence and Computing Research. DOI:10.1109/iccic.2014.7238499Vehicle selection for public transportation using an integrated multi criteria decision making approach: A case of Ankara, Journal of Intelligent Fuzzy Systems 26 (2014) 24672481, DOI:10.3233/IFS-13091, IOS Press

[5] Visalakshi, S., Radha, V. (2014). A literature review of feature selection techniques and applications: Review of feature selection in data mining.

[6] 2014 IEEE International Conference on Computational Intelligence and Computing Research. DOI:10.1109/iccic.2014.7238499

[7] Feature selection guidelines, https://www.machinelearningplus.com/ machine-learning/feature-selection/, [Accessed on :2018-11-10]

[8] Vehicles selection guidelines, http://www.nzdl.org/gsdlmod?e=d-00000-00—off-0fnl2.2–00-0—-0-10-0—0—0direct-10—4——-

[9] 0-1l–11-en-50—20-about—00-0-1-00-0–4—-0-0-11-10-0utfZz-8-

[10] 00cl=CL1.4d=HASH31111503a8027f96d6931c.9.3gt=1, [online], [Accessed on :2018-08-12]

[11] Machine Learning: Validation Techniques"https://dzone.com/articles/machine-learning-validation-techniques",[online],[Accessed on :2018-08-12]

[12] A Quick Introduction to K-Nearest Neighbors Algorithm "https://medium.com/@adi.bronshtein/a-quick-introduction-to-k-

[13] nearest-neighbors-algorithm-62214cea29c7",[online], [Accessed on :2019-02-01]

[14] NEURAL NETWORKS "https://www.doc.ic.ac.uk/ nd/surprise˙96/journal vol4/cs11/report.html", [online], [Accessed on :2018-08-12]

Investigating Enterprise Software System Integration Prototypes Involved in Integrating Disparate Software Systems

Dr. Nazneen Ansari, Sahil Raut
Department of Information Technology St. Francis Institute of Technology Mumbai, India
naz2ensfit@gmail.com, sahilraut172@gmail.com

***Abstract*—Organizations possess legacy systems and applications that are valuable assets and significant to current business operations. One of the main challenges in today's organizations is how to efficiently and effectively integrate legacy systems and applications with other systems and applications. Disparate systems have different platforms. Therefore, in order to integrate disparate systems with different platforms and infrastructures, integration technologies and solutions are required. Various approaches of integrating disparate enterprise systems are comprehensively studied and different solutions are examined. Therefore, this research work is conducted to cover the technical details of integrating disparate enterprise systems. In this work, a significant amount of literature is reviewed, and different ERP–BI enterprise software system integration prototypes are investigated.**

Keywords—disparate software systems, enterprise software system, integration prototypes

I. INTRODUCTION

Software systems play a crucial role in an organization as they integrate key information present within the organization [1]. Many organizations spend loads of money on such systems. To obtain returns on such investments, these systems must be used for many years. Depending on the organization size, the software system lifespan varies. Nowadays, many software systems are available in the market aiming to increase an organization's productivity and efficiency. The desire to integrate such systems turns out to be a priority for business managers [2]. Reducing the effort required for software integration is on a rise in organizations [3]. ERP consists of organizing and handling an organization's assets efficiently and productively. Business intelligence (BI) is a quintessential way to encourage an organization's core competency. It is an array of know-hows that expedite data collection, integration, and analysis in an organization thereby enabling the process of decision making. If ERP and legacy systems have previously been implemented in an organization and the organization attempts to integrate a BI system, then ERP–BI integration would be extremely challenging due to a variety of factors. Limited literature is available for integrating ERP and BI systems. Most of the studies emphasize on factors that influence the adoption of BI by ERP systems. This research work investigates different ERP–BI enterprise software system integration prototypes.

The rest of this paper is organized as follows. Section II provides an overview on enterprise software integration. Section III discusses enterprise software system integration prototypes. Section IV compares different enterprise software system integration prototypes. Section V concludes the study.

II. LITERATURE SURVEY

According to the Institute of Electrical and Electronics Engineers (IEEE) standard computer dictionary, integration is defined as "the process of merging either software and hardware components or both in a system" [4, 5]. A number of generations of systems are used by several organizations that rely on countless technology developed over years. These technologies are responsible for huge support to organizations. Regrettably, there exist many systems that are unable to communicate with each other. But, there is for all time a choice of substituting old systems with new ones. The only problems that arise are cost and time. But because of time and cost, we simply cannot discard these systems because they contain humongous of data and information that are of extraordinary importance to organizations.

The abovementioned problems can be solved via system integration. System integration was primarily restrained to technical aspects such as for linking hardware components. However, as facts on IT evolved, system integration came into existence in software too.

The waves of technology are striking the landscape of an organization with incredible force. For organizations, in order to meet customers changing demands, it is now or never to implement breakthrough technologies. By thoroughly learning and understanding system integration, organizations progress to ride the wave of technologies without losing existing systems functionality.

System integration offers organizations the capacity to get disparate systems together and make such systems work as a single system. System integration has grown into a function that provides power to an enterprise's business revolution journey. The market of system integration is sustained by the request for enterprise application software (EAS) together with SCM, ERP, data management, human resource (HR) applications, product lifecycle management, and BI.

With organizations enthusiastic about increasing returns from savings made on software applications and perfecting business operation efficiency such as superior worker output, better-quality workflow procedure, reduced upkeep costs, and value-added data integrity, system integration market is imagined to achieve more statures forthcoming years.

Since ERP systems were not initially designed to provide real-time reports to massive users, the entire system was unable to facilitate the decision support function. Progressively, organizations are focusing on BI tools that create data collected by ERP, CRM, and other important data-intensive applications. BI systems can pull the data in ERP systems and then perform analyses and deliver superior reporting, which help users for making timely and accurate decisions [6].

There are numerous organizational benefits, for example, increased profitability, decrease in costs, and increased efficiency, from system integration. Integration solutions enable using data and functionality exemplified in the existing applications of organizations or legacy systems instead of replacing these systems with new ones. In addition, these systems provide long-run benefits, for example, organizations can gain an instant and real-time view of all their data and operations, which help in making better decisions. Furthermore, they provide flexibility to rapidly adapt business processes to accommodate growth and meet new business challenges.

When compared to individual purchasing of ERP and BI systems, a combined application is a more profitable option, making such systems affordable to even low-end small and medium businesses and small and medium enterprises.

Chou et al. [7] state that the integration of BI and ERP systems can strengthen corporate decision-making capability by utilizing the analytical capability of BI system and data managerial capability of ERP system.

III. Enterprise Software System Integration Prototypes

There are several enterprise software system integration prototypes. Some of them are discussed below:

A. Point-to-Point

It is one of the most conventional approaches for integrating ERP–BI systems [8]. It is useful only when very few systems need to be integrated. In this approach, individual applications are loosely coupled, which in turn offers independency among applications. The main disadvantage of this approach is that as the number of applications increase, the number of connections also increase exponentially. Moreover, maintenance of this integration approach is a horrendous. Thus, such an integration approach is unfeasible for large organizations [9].

B. Middleware

With the help of a middleware, all-purpose or general interfaces can be arranged for. Such interfaces permit applications to distribute messages. Middleware performs operations like routing, altering, combining, splitting, and transferring. Note that establishing a middleware is a bit complicated.

General interfaces in a middleware allow communication between disparate systems. Middleware helps in forming a loosely coupled architecture. Moreover, it supports several applications with least maintenance cost as compared to p2p integration [10].

In order to attain a desired integration infrastructure, it is essential to use an open source middleware application. Microsoft BizTalk, Microsoft message queuing (MSMQ), and active message queuing (ActiveMQ) are some of the examples of commercial middleware integration product. The advantage of using such products is low cost. Moreover, these products provide 1:1, 1:many, and many:many integrations. In addition, safety, reliability, and publish/subscribe mechanism are some of the features of commercial middleware integration products. These products are thoroughly tested and are stable in nature. Another major advantage of these products is that there is always a support site that will provide detailed solutions on the Internet to users if a user faces any problems [11].

Integration can be achieved via middleware toolsets using SOA and Web service design. This approach is helpful when there are numerous integrations to be built and maintained since the tools can be expensive [12, 13]. Middleware helps in providing an arbitration point between applications. It has a common interface for passing messages between integrated applications [14]. Middleware-based integration approach provides a number of integrated applications and requires little maintenance. In addition, complex operations such as transformation, aggregation, routing, separation, and message conversion can be performed using this approach [15].

The only disadvantage of this approach is establishing the middleware and changing existing applications to use middleware APIs. Many companies such as IBM, Oracle, Microsoft, Sun, and Computer Associates [16] offer integration technology as part of an e-business platform and provide middleware-based integration. For example, IBM offers WebSphere—an application and integration software platform—that includes a middleware layer to interface with prominent business applications. Companies that offer integration technology are Tibco, Vitria, DataMirror, Propelis, etc. Even though they offer integration solutions, they not only differ in terms of technology but also application focus.

One of the common integration approaches observed today is middleware integration. Some of the most common middleware solutions are Oracle Fusion Middleware, IBM WebSphere, Informatica, TIBCO, and WebLogic. Middleware offers ready-made connectivity to applications. Furthermore, there are some middleware that offer packaged integration processes.

Middleware includes functions like creating, sending, receiving, and routing structured and unstructured data. Nowadays, middleware provides embedded logic, business rules, and an application process flow across web services. Lately, cloud-based solutions are made available by middleware providers, which can be implemented quickly [17].

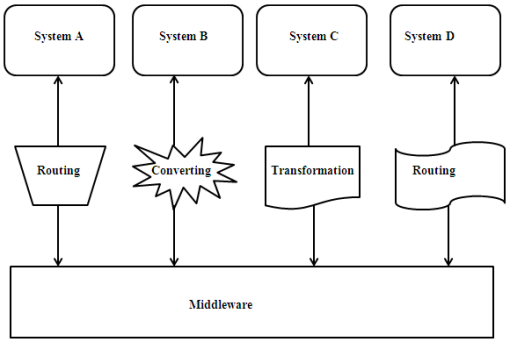


Fig. 1. Middleware approach

C. Custom Programming

Best suited for one-to-one integration projects. It is highly expensive, difficult to maintain, and takes a lot of time to build and test. It is less widespread than other approaches because the links are all handmade. It is also difficult to maintain.

Depending upon an organization's requirements, customized software applications are developed to serve as an integration infrastructure. By doing so, the organization receives not only the complete control of middleware but also the complete control of features that are needed for implementation. Non-functional supplies (i.e., operation and security) can be used to develop a custom programming solution. In addition, it has been observed that the developed solution supports 1:1, 1:m, and m:m integration topologies. Moreover, the development cost associated with custom programming is less.

Use of internal development tools, APIs, web services, etc. are some of the means of creating a customized integration

solution by organizations. Only those organizations can do this those who have skilled developers and adequate resources.

Customized integrations can be performed through mashups. Concurrent integration, quick execution, and display of sensitive data to users without storing the data outside a firewall are some of the advantages of using mashups.

D. Connectors

They are open-source applications that offer explicit connectivity among systems. Communication between ERP and BI systems takes place via connectors. Data exchange is handled through API, which is installed either in an ERP or a BI system.

- ***Advantages:***
  1) Simple to deploy only if the requirements of an organization are straightforward.
  2) Because connectors are designed for a specific BI system, they offer supplementary services other than connections.
- ***Disadvantages:***
  1) Because of specificity, connectors cannot be used universally (i.e., if an organization wishes to install several BI systems, then a new connector needs to be deployed).
  2) As per a connector's availability, clients need to select BI systems.
  3) Connectors are generally suitable for small organizations where massive data exchange does not take place [18].

Connectors include an organizational console for observing data flow. They provide a supple configuration procedure for plotting and altering data on the basis of extensible stylesheet language transformations (XSLT). Connectors synchronize ERP with BI, which means that when any change occurs, information flows to and fro. Connectors contain semantic renovations so that mappings are available out of the box.

Native connectors are suitable for organizations that prefer single-vendor solution. They provide quick customer–master synchronization. Moreover, they can be implemented from a few days to 2–3 weeks.

E. Single Source

Buy BI from the same supplier that provides your ERP solution. There is also a possibility that complete integration is already built-in. For organizations with large investments in ERP, BI options are better than ever. Most ERP vendors now provide various built-in BI solutions. There have been a lot of developments in this domain during the last few years. ERP vendors saw BI as a group to sustain traditional growth levels. Many of the large BI vendors have been acquired.

Oracle started the trend by acquiring Hyperion, SAP acquired business objects, and IBM followed suit by acquiring Cognos. Ease of integration and value for money are the key reasons behind choosing BI from ERP vendors. "If ERP vendors fulfill key performance indicators with their BI products and score well on other vendor evaluation criteria, then it's a plus point to get your ERP vendor on board for BI". The advantages of clubbing BI and ERP vendors, to list a few, would be easy integration, packaged licensing costs, lower system ownership costs, familiarity with vendor interfaces, etc.

F. Third party

Find a third-party BI supplier who has already developed an integration approach with your ERP brand and version. Organizations find deficiencies in BI products offered by ERP vendors. Easy integration may not possibly be the sole criteria that organizations can opt for. Specialized needs will play a major role in evaluating an ERP vendor for BI. If you are going to need advanced analytical capabilities that BI provided by ERP vendors do not offer, then it is logical to deploy a third-party BI. Choosing between the two often requires a thorough assessment of major issues, including cost and integration.

G. Integration Framework

To enable enterprises to integrate legacy and firsthand systems using different technologies on premise and cloud applications as well as address scalability challenges, enterprise integration frameworks are developed. To reduce cost and time of integration projects, enterprises are increasingly inclining toward integration frameworks. The parameters that should be considered while choosing an integration framework are as follows [19]:

- *Current integration needs:* Before choosing an integration framework, enterprises should not only consider the need for integration with external systems but also jot down different types of transport models.
- *Out-of-the-box connectors:* While comparing different integration frameworks, enterprises should analyze their need for out-of-the-box connectors. Many integration frameworks offer an array of out-of-the-box connectors, which help in accelerating the integration.
- *Future integration needs:* A good integration framework should not only support current but also future needs. A good software development kit will assist in developing new connectors.
- *Integrated development environment (IDE):* For developing integration flows, evaluating integrated development environment (IDE) of a framework is essential. Some frameworks have intuitive IDE, while others require users and different parameters, respectively, to configure complex uniform resource identifiers and make the integration process longer and more complex.
- *Open source or commercial products:* Enterprises need to choose between open source and commercial products. Both have their own merits and demerits. Open source products are cheaper and easily modifiable, whereas commercial products provide many connectors and are based on service level agreement levels.

H. Productized Integration

They are the ones that assist in carrying out integration between well-liked and widespread applications. Here, integration is presented as a ready-made solution that can be installed, altered, and progressed with marginal jeopardy. They cover common integration design patterns and business procedures. Moreover, they contain ready-made data mappings, restorations, and process transpositions. They are available in the form of integration and packaged applications and connectors. Application vendors are the ones that tender productized integration.

IV. Comparison of Different Enterprise Software System Integration Prototypes

Table I shows comparison of different integration prototypes.

TABLE I. COMPARISON OF DIFFERENT INTEGRATION PROTOTYPES

Sr no.	Prototype names	Advantages	Disadvantages
1.	***p2p***	Easily integrates 2 to 3 applications, excellent speed, and simple to understand and implement	Continuance and expandability proliferate an integration solution's complexity level. Communication program needs to be written.
2.	***Middleware***	Affords standard interface, offers functionalities to carry out tasks such as transformation, routing, totaling, transfiguring, easy data passing, and least upkeep cost.	Introducing a middleware and modifying prevailing applications to use the introduced middleware is complicated.
3.	***Custom programming***	Suitable for 1:1 integration projects.	It is pricey, problematic to maintain, and requires huge development and testing time.
4.	***Connector***	It keeps a reference of corresponding objects and contains a record of semantic conversions and flow of activities.	They are application specific. General connectors are difficult to create.
5.	***Single source***	Stress-free integration, packed accrediting costs, less system costs, and friendly vendor interfaces	BI products that have some limitations or shortcomings are offered by ERP vendors.
6.	***Third party***	It possesses sophisticated logical competences.	Assessment issues, i.e., cost and integration, need to be performed and managed accordingly.

V. CONCLUSION

If ERP vendors fulfill key performance indicators with their BI products and score well on other vendor evaluation criteria, then it is a plus point to get an ERP vendor on board for BI. However, organizations find deficiencies in BI products offered by ERP vendors. If organizations require advanced analytical capabilities that are needed by BI systems and cannot be fulfilled by ERP vendors, then it is logical to deploy a third-party BI system. Choosing between the two options often requires a thorough assessment of major issues, including cost and integration.

There are different approaches for coupling disparate enterprise software systems. Depending upon the number of applications to be integrated, expertise, budget, and time, different approaches can be applied for coupling disparate software products.

There is no single approach for integration. Depending upon integration requirements, different approaches can be selected. Integration solutions and framework are also available for synchronizing disparate enterprise software systems. There is no unique framework or solution. Solutions can be built or purchased from vendors. Moreover, various prototypes of integrating ERP and BI systems were studied and compared.

REFERENCES

[1] Ansari, N., & Thampi, G. T. (2015). "Investigating contemporary software engineering challenges involved in integrating disparate software systems," *International Journal of Computer Science and Information Technologies*, vol. 6 (4), pp. 3849–3854, 2015.

[2] Larrucea, X. (2008). Method engineering approach for interoperable systems development. *Software Process: Improvement and Practice*, 13(2), 127-133.

[3] Mazhelis, O., Tyrvainen, P., & Viitala, E. (2007, December). Modeling software integration scenarios for telecommunications operations software vendors. *In Industrial Engineering and Engineering Management, 2007 IEEE International Conference on* (pp. 49-54). IEEE.

[4] Rognerud, H. J., & Hannay, J. E. (2009, October). Challenges in enterprise software integration: An industrial study using repertory grids. *In Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement*, pp. 11-22. IEEE Computer Society.

[5] Radatz, J., Geraci, A., & Katki, F. (1990). IEEE standard glossary of software engineering terminology. *IEEE Std*, 610121990(121990), 3.

[6] Patil, N. V., & Kshirsagar, M. C. (2015). Enterprise Application Integration using Service Oriented Architecture with Web Service Aggregation Pattern.

[7] Chou, D. C., Bindu Tripuramallu, H., & Chou, A. Y. (2005). BI and ERP integration. *Information Management & Computer Security*, 13(5), 340-349.

[8] Fenner, J. (2011). Enterprise application integration techniques.

[9] Yang, L., Jones, B. F., & Yang, S. H. (2006). Genetic algorithm-based software integration with minimum software risk. *Information and Software Technology*, 48(3), 133-141.

[10] Patil, N. V., Kshirsagar, M. C., & Jaypal, P. C. (2014). Enterprise Application Integration using Service Oriented Architecture with Generic Pattern. *INPRESSCO International Journal of Current Engineering and Technology*, 4(5).

[11] Smaavik, T. F., & Øvstetun, N. T. (2007). An Evaluation of messages-based Systems Integration with respect to Performance.

[12] Ansari, N., & Thampi, G. T. (2016, April). Investigation of different approaches, feasible solutions, and estimates for disparate enterprise solution synchronization. *In Computing, Communication and Automation (ICCCA), 2016 International Conference on* (pp. 214-218). IEEE.

[13] Ansari, N., & Thampi, G. T. (2015, October). Template creation to merge disparate software solutions by adapting software engineering principles. *In Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on* (pp. 1339-1343). IEEE.

[14] e-Zest, Enterprise Application Integration, Re-imagining enterprises through Digital Transformation, [Online] Retrieved from http://www.e-zest.com/eai/

[15] Java World from IDG, Enterprise application integration using J2EE, [Online], Retrieved from http://www.javaworld.com/article/2074488/enterprise-java/enterprise-application-integration-using-j2ee.html

[16] Destination CRM.com: From the editors of CRM Magazine, Middleware: Directing Enterprise Traffic, [Online] Retrieved from http://www.destinationcrm.com/Articles/Editorial/Magazine-Features/Middleware--Directing-Enterprise-Traffic-46640.aspx

[17] Knorr, E. (2006). Software as a service: The next big thing. InfoWorld, 3, 2-4.

[18] SEGÉŇ, J. (2014). Connecting ERP and e-commerce systems (Doctoral dissertation, Masarykova univerzita, Fakulta informatiky).

[19] Persistent, Choosing the Right Enterprise Integration Framework, [Online] Retrieved from https://blog.persistent.com/index.php/2014/01/28/3077/#.VrcUQvl97IV

A Comparative Study on Recent Trends and Applications in Cloud Computing

Jaishree Tawade, Vikas Kaul
Department of Information Technology, Thakur College of Engineering, Mumbai

***Abstract*—Cloud computing is a storage technique which stores most of our data onto cloud i.e internet. Cloud provides on demand service to the users. The cloud service provider gives access to users to store data on to their platform and use it whenever required. Cloud computing is changing our lives by contributing to society in all sectors like Agriculture, Banking, Military, Business, Web Applications, Health Care, Home security etc. In this paper we discussed application of cloud computing in daily life and how it is changing our lives and the concerned security issues.**

Keyword: Cloud Computing, On Demand Service, Security.

I. INTRODUCTION

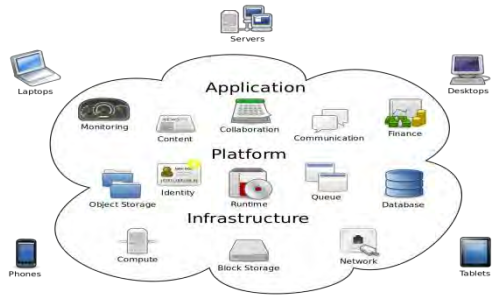


Fig 1. Cloud Computing

As per the definition provided by the National Institute for Standards and Technology (NIST) "cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".[3]

A. Cloud service models:

Software-as-a-Service (SaaS): It is a software distribution model where a third-party provider hosts applications and makes them available to customers over the high-speed internet connection.

Platform-as-a-Service (PaaS): It is a middle layer which gives the organizations, institutions or companies a freedom and framework for developers to develop their own applications and deploy them and make customers within their company to access the resources.

Infrastructure-as-a-Service (IaaS): Infrastructure is most vital among the three service models because it is the basic need to launch the organization's services over the internet in a cloud platform, to make their services available to clients and applications to run them smoothly.

B. Cloud computing deployment models:

Public Cloud: The cloud services are easier to install and less expensive or even charge free, the applications, hardware and bandwidth are provided by the service provider, and are scalable, the user avail can only those services that they are interested.

Private Cloud: As the name suggests, its services, infrastructure is solely operated and maintained by an organization. The services are made available on proper authentication, priority is being given towards the client's data security.

Community Cloud: Here the cloud resources are shared by an organization which is of common interest for every participant which is being part of a community, whose needs are similar.

Hybrid Cloud: It is a combination of two or more cloud deployment models like (public, private, community) it enables cloud application portability, multi-tenant, resource sharing.

C. Cloud computing characteristics

On-demand self-service: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically.
Measured service: Public cloud providers like Amazon allow companies to avoid large upfront infrastructure investment, so the small companies can afford the workloads as per their requirement.

Broad network access: Capabilities are available over the network that promotes use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).

Rapid elasticity: Capabilities can be elastically provisioned and released, scale rapidly outward and inward commensurate with demand dynamically.

Resource pooling: The resources like storage, servers, memory, Processing Unit, Network and virtual machines can be pooled and utilized by multi-tenant fashion with dynamically provisioning and provisioning of resources

II. APPLICATIONS

- Smart farming:

Smart farming is new revolution in the traditional agricultural system which uses modern technologies like GPS, Cloud, Internet of Things, Sensors, Soil Scanning, Data management and Precision etc which increase quality and quantity of food.
According to Wolski[1] The agriculture industry faces tremendous pressure to increase crop production and yields to meet future consumer demand for food. This problem is compounded by the population growth of the planet, severely

limited natural resources (arable land and water), unpredictable weather patterns, the requirement of sustainability, and climate change, among other variables.

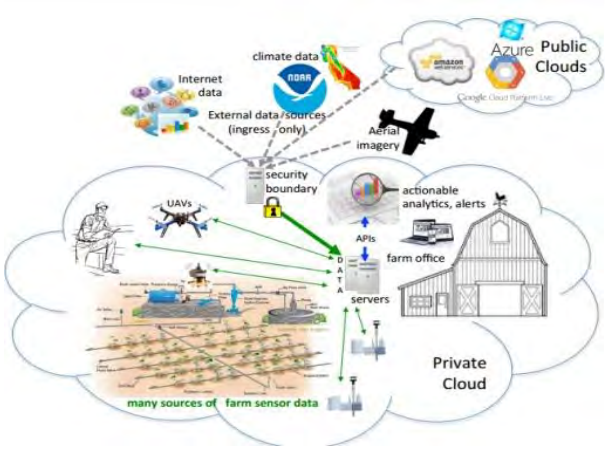


Fig. 2. Smart Farming

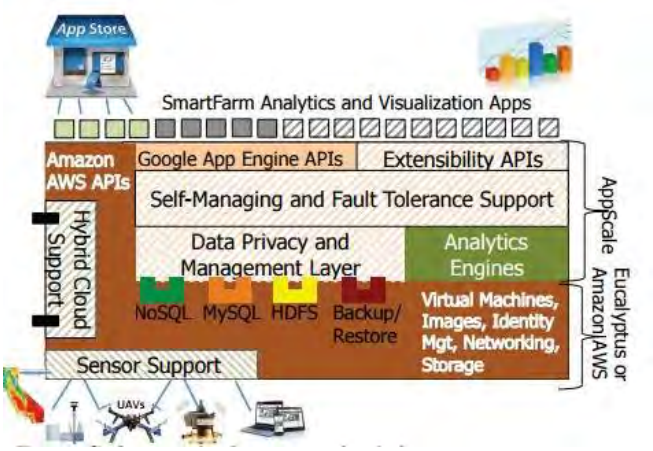


Fig. 3. The UCSB Technology

In an effort to make farms more productive, farmers are increasingly turning use computing technologies for environmental measurement and data analysis. Computing systems can automate this process to facilitate faster problem diagnosis, more accurate outcome prediction, and proactive decision making.

Smart farming is intelligent farming system help the farmers to keep an eye on their farms remotely. The UCSB[1] is an on-going project which is introducing a new software application an App store as shown in fig3 which includes

- Open source, secure, on-premise (private) cloud
- Easy to use, self managing, lights-out automation
- Linking disparate sensor technologies and analysis engines
- To give farmers actionable insights, alerts, accurate prediction

UCSB SmartFarm Technologies

• Private cloud is API compatible with Amazon and Google Clouds
– Any application or service that executes in public cloud can also execute in on-farm without modification
– Facilitated via AppScale and Eucalyptus private cloud systems
– Hybrid configura8ons
• Move computation to the data on-farm; control all data that leaves farm
– Precision farming “app store”
• Or perform computation in the public cloud; shared datasets and services
• Leverages emerging Internet-of-things (IOT) & big data technologies
– Unified data ingress from disparate sensor systems
• Images and data
– Machine learning and data mining Streaming (think Storm) and batch processing (think Hadoop/Spark)

AppScale is an open source distributed runtime system that makes it very easy to write and deploy network-accessible applications (including mobile device back-ends) in high level languages. AppScale is commercially supported, freely available, and is API-compatible with Google’s cloud platform App Engine. This means that applications that run on App Engine, also run on AppScale and thus SmartFarm, without modification. AppScale runs on public cloud infrastructures (Amazon Elastic Compute Cloud (EC2), Google Compute Engine, Microsoft Azure) or on local cluster resources within virtual machine instances over Eucalyptus. Eucalyptus is an open source cloud infrastructure that runs on local cluster resources and provides highly available, fault tolerant virtual machine management. Eucalyptus is API compatible with Amazon EC2 and Simple Storage Service (S3) so that any EC2 instance/service can run on Eucalyptus without modification. AppScale and Eucalyptus together in SmartFarm provide us with a production-ready, highly available private cloud system and our users with a vast ecosystem of freely available cloud software from Google and Amazon.

As a result they have developed a Smart Farm app for wine grape growers called **RootRApp**. RootRApp employs anthocyanin sampling via Near Infrared (NIR) sensing in combination with Ordinary Kriging to implement automated differential harvesting. This use of sampling and Kriging constitutes an example of statistical analytics used as a control mechanism since the harvester uses the Kriging interpolations to separate grapes as it picks them into two different quality batches automatically.

- Smart Transportation:

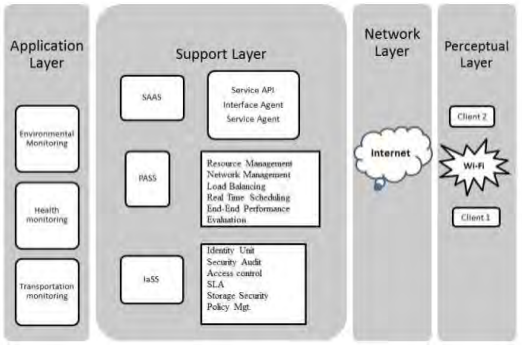


Fig 4. Smart Transportation

As the number of accident is increasing, according to a survey USA is having highest frequency of accidents; the STS i.e “System for a smart transportation” gives futuristic transportation systems. Smart transportation uses advanced technologies such as wireless

sensor networks, distributed system architectures, cloud computing, sensing and actuating, control and detection in multiple types of automobiles on roads to improve safety of the passengers, optimized service of the transportation system and real time traffic situation. STS is coming fast towards the integration of virtual technology and cloud computing in the transportation field.

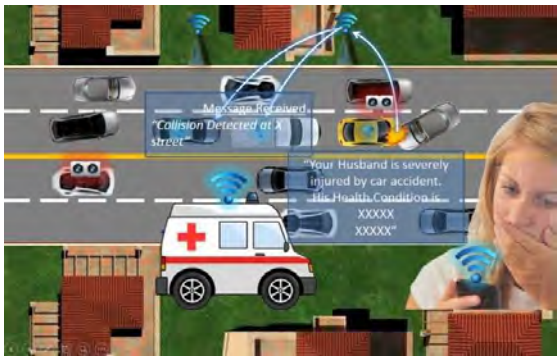


Fig 5. An example for smart transportation

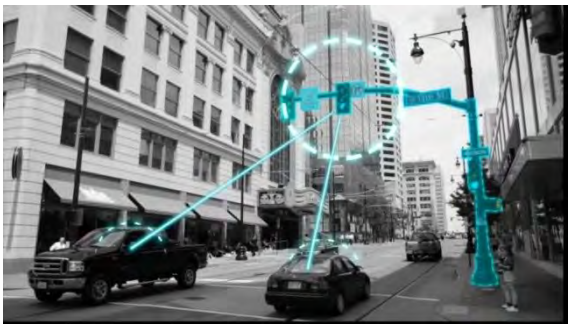


Fig 6. Smart Driving Car

The Paper[2] explore new possibilities to enhance the framework of the STS to a valuable cloud computing system, which includes both machine side and client side. The author has proposed techniques which consider an internet enabled STS system which can play a very crucial role through using Cloud computing to enhance accident prevention/monitoring and control. The STS system find the destination of Accident and send information to concerned people using cloud which support permanent services like computing, amateur radio, aviation etc

The Fig.4 shows the architecture of STS where four layers is presented from cloud computing to end-user (Vehicle). The communication layer contains a transportation stratum for sharing of information between end-user via an ad hoc network or to a cloud server in the event of an accident. In the transportation stratum, if we have a Wi-Fi module and a Bluetooth module connected using their respective interfaces to the processor. Here, there are two processors for processing the data. First is the main processor and the second one is the subsidiary processor. The main processor will just share the data (video, audio, information about the location using GPS/WiFi etc.) to the clients in the range of ad hoc network or to a cloud-based server using Internet services. The last part is the conventional cloud which is responsible of data collector stratum. It consists of various modules for collecting and transferring data and information like camera modules for rear and front, LCD module, GPS Module, pressure sensor for accident detection, temperature sensor for protecting the data which could be destroyed from excessive temperature, proximity sensor for vehicular lane change detection, and speed sensor.
Fig5. Shows a STS system where we can see the information is collected using the collectors namely front and rear camera module for recording the pre and post-accident data. The information gathered is processed using a processor. The information is simultaneously sent to the end users who are nearby the ad-hoc network without any intervention from third party electronics device so that they can receive the news about the nearby vehicles and the drivers then decide for themselves whether to overtake the preceding cars or not. In the event of accident, the video recorded using the collectors and the information about the location specified by the GPS module are transferred to a Wi-Fi-based server so that the concerned authorities can have a look at the evidence stored. Therefore the reliability is redundant resource utilization and availability can be understood as the possibility of attaining the resources whenever they are needed with the consideration to the time it takes for these resources to be provisioned. Irrespective of employing architectures having attributes for high reliability and availability, the services in cloud computing can experience in Denial of service attacks, Performance slowdowns, Equipment outages and natural disasters. In order to remove FUDD (fear, uncertainty, doubt, and disinformation) probably the reliability, availability and security are important. The level of reliability and availability of cloud resources must be considered as a serious problem in the organization's planning to set up the cloud infrastructure in order to provide effective services to the consumers.

The paper[5] mainly focus on STS based on reliable cloud networks especially in vehicular safety of drivers while driving on highways roads. This system is going to support various services such as economy, traffic congestion, pollution and comfortable driving. Further, it would be easier for the investigation end to collect the information statistics of the accident (if any) from the cloud server. This system will avoid the road rage as people would be priory aware about the traffic statistics. Thus this system would act as an alert eye for the travellers. Hence, cloud computing provides a highly scalable and on demand computing platform to End to End user with high elasticity and availability to the Smart Transportation Services.

- Smart Mobile Banking:

Cloud computing has changed the way of communication between banking and its customers. A large number of banks are now adopting cloud technology to fulfill their needs including less budget, effectiveness business, green Information Technology (IT) business and time saving using Mobile banking[7].

Cloud computing provides some services in Banking sector as BPaaS, SaaS, PaaS, IaaS. The BPaaS stands for Business Process-as-a-Service use cloud computing in business as in terms, including billing, payroll. The SaaS stands for Software-as-a-Service including accounting, Customer Relationship Management (CRM), and Enterprise Resource Planning(ERP), invoicing now PaaS stands for Platform-as-a-Service is a Fulfil platform for applications, interface and it decrease IT cost & decrease the number of hardware & software and hosting environment. Now the last one is IaaS stands for Infrastructure as-a-Service allows a business to buy those resources as a fully outsourced service that purchasing servers, software's, data center space and network equipment.

Mobile Banking using cloud computing can be explained as a way whereby the bank customers communicate with a bank by a mobile device and use a cloud computing as growing technology

for instance, mobile phone with internet or PDA (Personal Digital Assistant). Mobile Banking is used by smart phone and other devices such as tablets to perform online banking tasks including knowing account balance, transferring payments, and funds between two accounts and locating an ATM (Automated Teller Machine)[6]

The general architecture of CMB (cloud mobile banking) is shown in figure7. Mobile devices are connected to the mobile operators via base station including access point or satellite that control the connections and functional interfaces between the mobile devices and mobile network operator. Mobile users requests and send information such as ID and location are transmitted to the central processors that are connected to servers providing mobile network services. The financial sectors like banks, both are connected to Internet Service Providers (ISPs). ISP provides internet connections to the world community. Subscribers and banks request are delivered to a cloud through the internet. In cloud, the cloud controllers process the request to provide mobile user and banks with the corresponding cloud services. Mobile user can take the banking facility from cloud computing as corresponding bank details are residing on data centres of cloud computing[7].

Benefits:

Utilization of Time: We can use Mobile Banking through cloud computing 24*7 hours so it is very convenient and most easy and a great choice for operates financial services for most mobile phone owners in rural areas cause everywhere is cloud.

Increase Adaptability: Using cloud computing, banks will enjoy upgrade adaptability ratios and operating leverage. Using cloud computing technology and business operations can be much more effectively aligned, the cloud gives to mobile banking a golden opportunity to decline complexity.



Fig 7. Smart Mobile Banking System

- Smart Military Security**:**

Leaking of personal information, secret government document, confidential secrets about our country defense, military secret is now a day big serious problem facing today. Uploading an information on cloud can now easily hacked by hackers.

However, the problem comes to military users if scientist develop a new weapon for military and he want to send a launching code to military admirals /chiefs through cloud, how he can trust cloud that he's codes will be safely delivered to admirals.



Fig 8. Smart Military Weapon Detection System

The paper[5] proposed a system which use steganography, watermarking, image encryption and visual cryptography schemes to protect military weapons data in clouds. Steganography allows users to hide the weapons launch code in image captcha. Visual cryptography shares the image captcha in shares which is depend on number peoples in group in military. Image encryption will apply on each share of captcha. After this watermarking is apply on each share for authentications between users and cloud. For receiving the launch code receivers have to from de-watermarking, image decryption then visual cryptography to get captcha and launch code. Our studies show that the proposed approach achieves good security performance and securing the future of country.

The Application will be able to generate weapons launching code which is in text format will hide in image captcha. After this image captcha, will breaks into shares. After making shares using visual cryptography. after this watermarking is applied on each pixel of image shares. Image encryption is applied to encrypt image shares. after all this process shares are send through email. When it's come to receive mail, decryption is applied on each share then De watermarking is applied after this visual cryptography is done to collect share and generate original image. Then stenography is used to get hidden weapons launching codes from image captcha.

- Smart Health care system

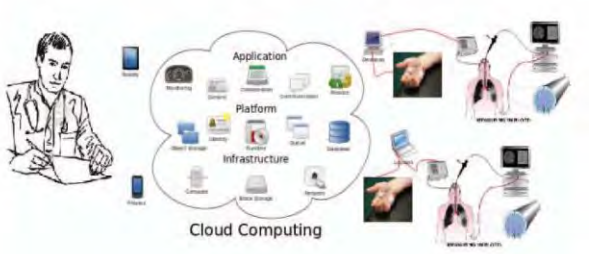


Fig 9. Smart Health Care System

Ex: ClearDATA**,** Dell's Secure Healthcare Cloud, IBM Cloud

According to paper [8] the problem under study consists of improvement in health care system using cloud. They passionately dedicated to help medical fraternity to find health status of vital organs of the patient's body at early stage that support effective treatment by introducing innovative and high quality hand carried non-invasive health care systems and devices. Cloud based Health Care is the integration of cloud computing and health monitoring. The computing device enables

the delivery of accurate medical information anytime anywhere by means of internet. Cloud based healthcare system consists of a computing device and number of sensors mounted on patient's body. In this paper we present a Cloud based Intelligent Healthcare Monitoring System (CIHMS), which can provide medical feedback/assistance to the patient through cloud (if data already available) or hospital. The suitable sensor is/are to be used to obtain adequate data related to patient's disease.

In this paper[8], the author presents a cloud based health monitoring system that uses
- The Wearable sensors for collecting data from patients, data may be tempura, BP, ECG, from patients those data send to the server through sensor gateways.
- cloud for storing/mining the data and intelligently predicts patient's health status. It also provides feedback to patient through the computing device.

The physician then can attend on patient directly or send necessary precautions to patient via communicating device. Moreover, so far there is no automated medical server used in the field of healthcare as it requires large number of specialists to monitor the patient's health status data. That motivated to work for a cloud based intelligent medical server for health care applications that will aid the specialists in the health care. In addition simulated the novel cloud computation model where each doctor's cell phone or PDA is considered as node. The data from patient and physician will be collected and stored on cloud. The cloud is equipped with computation module. The computation module accepts the new data from patient and processes it, then compares the obtained result with the existing result. If found suitable, the patient will be communicated with feedback. In case if the suitable record is not found, the appropriate physician will be intimated via phone call or sms.

The physician can get the information about patient from cloud. The physician may perform various procedures to diagnose the disease and send the report to the respective patient. Figure.9 Illustrates overall functioning of mobile healthcare system. The steps being carried out are as follows:

- The patient is provided with wearable sensors.
- The signal from the sensor is captured in a Node called sink through which data will be forwarded to cloud via Wi-Fi or Bluetooth where the data is stored for further processing or future reference.
- The data is then forwarded to CIHMS's physician via phone/PDA (CIHMS system).
- The CIHMS system allows ubiquitous access to patient's data and medical information in real-time.
- The medical data is stored & processed in CIHMS's server where database about different patient's health conditions will be maintained, and can be used for establishing diagnostics and treatments.
- CIHMS server centralises the received data from body sensors and makes it available globally.

The paper discusses the simulation of primary health care system on intranet. It also discussed the multidisciplinary endeavour such as cloud for healthcare systems that provides greater benefits for patients and hospitals. In this paper introduces a novel cloud computing model that can efficiently compute the data from healthcare devices and dynamically diagnose the disease also, an intelligent way of saving only crucial data. The cloud based health care system using bio-sensor network places forward some future works such as finding the most effective mechanism for ensuring security for the data[8]

- Cloud storage:

The number of cloud storage providers online seems to grow every day. Each competing over the amount of storage they can provide to clients.

Right now, Dropbox is the clear leader in streamlined cloud storage allowing users to access files on any device through its application or website with up to 1 terabyte of free storage.

Google's email service provider Gmail, on the other hand, provides unlimited storage on the cloud. Gmail has revolutionized the way we send emails and largely responsible for the increased usage of email worldwide.

Facebook is a mix of the two, in that it can store an infinite amount of information, images, and videos on your profile. They can then be easily accessed on multiple devices. Facebook goes a step further with their Messenger app, which allows for profiles to exchange data.

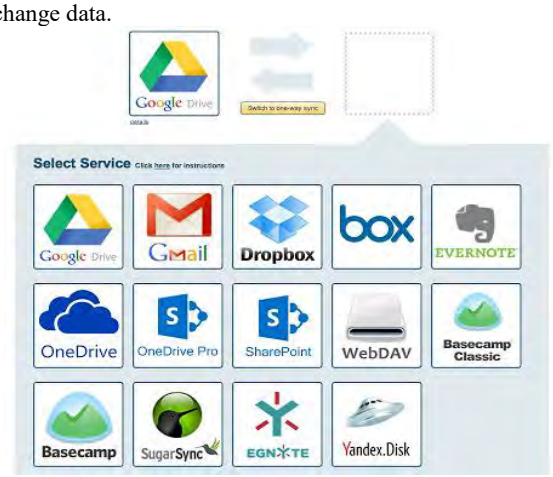


Fig.10 Cloud Storage System

- Marketing Cloud Platforms

Ex: Maropost Marketing Cloud, Hubspot, Adobe Marketing Cloud

A marketing cloud is an end-to-end digital marketing platform for clients to manage contacts and target leads. Maropost Marketing Cloud combines easy-to-use marketing automation and hyper-targeting of leads. At the same time, ensuring emails actually arrive in the inbox, thanks to its advanced email deliverability capabilities.

In general, marketing clouds fulfill a need for personalization. This is important in a market that demands messaging be "more human." That's why communicating that your brand is here to help, will make all the difference in closing.

- Cloud Computing in Education:

Ex: SlideRocket, Ratatype, Amazon Web Services

Education is increasingly adopting advanced technology because students already are. So, in an effort to modernize classrooms, educators have introduced e-learning software like SlideRocket.

SlideRocket is a platform that students can use to build presentations and submit them. Students can even present them

through web conferencing all on the cloud. Another tool teachers use is Ratatype, which helps students learn to type faster and offers online typing tests to track their progress.

For school administration, Amazon's AWS Cloud for K12 and Primary Education features a virtual desktop infrastructure (VDI) solution. Through the cloud, allows instructors and students to access teaching and learning software on multiple devices.

- Cloud Computing for Government

Uses: IT consolidation, shared services, citizen services

The U.S. government and military were early adopters of cloud computing. The U.S. Federal Cloud Computing Strategy, introduced under the Obama administration, was instituted to accelerate cloud adoption in all departments.

According to the strategy: "focus will shift from the technology itself to the core competencies and mission of the agency." The U.S. government's cloud incorporates social, mobile and analytics technologies. However, they must adhere to strict compliance and security measures (FIPS, FISMA, and FedRAMP). This is to protect against cyber threats both domestic and abroad.

Cloud computing is the answer for any business struggling to stay organized, increase ROI, or grow their email lists. Maropost has the digital marketing cloud solutions you need to transform your business.

III.CONCLUSION

Cloud computing is the fastest growing field of IT providing enormous benefits to Business of all sizes also playing a vital role in our daily life improving our lifestyle in more systematic and sophisticated way. Cloud Computing is contributing in all fields like Farming, Military, Transportation, Health care, Education etc. Amazon, Azure, Google and IBM Cloud platforms are the fastest growing network environments. Security in clouds is a key challenge, much vulnerability in cloud still exits and hackers continue to exploit these security holes. So more concentration is needed on security challenges.

REFERENCE

[1]. "Improving Agriculture Sustainability Using Modern Information Technology UCSB Tech" Report 2016-04 May 2016 by Chandra Krintz, Rich Wolski, Nevena Golubovic, Benji Lampel, Varun Kulkarni Dept. of Computer Science UC Santa Barbara Balaji Set huramasamyraja*, Bruce Roberts Dept. of Industrial Technology Dept. of Plant Sciences Fresno State University Bo Liu Dept. of Bio Resource and Agricultural Engineering Cal Poly University, San Luis Obispo)

[2]. Secure and reliable cloud networks for smart transportation services Dhananjay Singh, Madhusudan Singh, Irish Singh, Hoon-Jae Lee.

[3]. "A COMPREHENSIVE STUDY ON CLOUD COMPUTING PARADIGM"Ab Rashid Dar1 , Dr. D. Ravindran2

[4]. "A Research Homomorphic Encryption Scheme to Secure Data Mining in Cloud Computing for Banking System" by Sneha Sakharkar, Shubhangi Karnuke, Snehal Doifode, Vaishnavi Deshmukh.

[5]. "ADVANCE SECURITY IN CLOUD COMPUTING FOR MILITARY WEAPONS" by Akash Bagul, Puja Sonawane, Laxmi Sawant ,Rohit Doshi

[6]. Stuart J. Barnes and Brian J. Corbitt, "Mobile Banking: Concept and Potential", International Journal of Mobile Communications, pp 1-17. 2003

[7]. "Cloud Mobile Banking:A Growing Technology " Vibhor Agarwal1 , Akshay Kumar Siddhu ,Shikha Garg.

[8]. "Cloud based Intelligent Healthcare Monitoring System "Mr.Khyamling A. Parane, Mr.Naveenkumar C.Patil, Mr.Shivananda R. Poojara, Mr.Tejaskumar S. Kamble

Improvement in Productivity and Income using Farming Assistance

Mr. Jugal Parulekar,Ms. Shraddha Pathare ,Ms. Kajal Dhakate,Ms. Chaitali Waghulde, Mrs. Surekha Janrao

Department of computer, Terna Engineering College, Nerul, Navi Mumbai

***Abstract* - No wadays mobile phones have become a promising tool in agriculture because the cost of the device is accessible and their computing power allows a variety of practical application to assist diverse farming tasks. According to the observations, Information and Communication Technologies (ICT) state that the cell phones play crucial role in farmers' lives. The traditional methods used by the farmers are very slow and undependable, especially in India. This results in considerable portion of crop getting affected by the bacterial attacks and lack of information, harming the land and degrading its productivity as well. On an average, damage to crops and agricultural production due to these reasons exceeds 40% of the total yearly production. Our key idea is to provide subsidy to the farmers who are Below Poverty Line (BPL) for buying pesticides and fertilizers by scanning their ration card. The presented paper also suggests various ways in which a farmer can utilize Mobile Computing (MC) on their handsets using application called "AGRO", to assist them for relatively better cultivation, yield and merchandise.**

Keywords: Farming Assistance, Image Processing, Segmentation, Firebase Database.

I. INTRODUCTION

Agriculture is the basic reason of production of food and raw material, which eventually is the reason of survival of the population. In India, agriculture is the primary occupation and most of the population is dependent on it. However, here is also need to review and revitalize the mechanism for updating the technology. In the upcoming years agriculture will see major changes. Unlike the earlier 'green revolution' which had a foundation of advanced pesticides and fertilizers, now the agriculture will be revolutionized with the help of technology. Every developing economy has agriculture sector as irreplaceable pillar and so does India. In India the agriculture sector contributes close to 20% of Gross Domestic Product (GDP). Either directly or indirectly, 60% of total population of India depends on agriculture. The majority of Indian farmers, which includes small-scale producers, are often unable to access the information and technological resources that could increase the yield and lead to better prices for their crops and products. The widely spread network of mobile phones could be the game changer in this problem. It will put agriculture field at its peak. The main purpose of this project is to develop a mobile phone-based solution that helps in farm management, which will lead to improvement in agricultural yield and helps in maintenance of the farms.

II. RELATED WORK

The main aim of modern farming technique using android app is to enhance transparency in the agriculture commodity marketplace by providing market price information, facilitating collective buying of inputs and collective selling of product. Farmers most depend on weather forecast to decide as what work to do today and tomorrow.

Earlier approaches which focused on how illiterate n below poverty line farmers would get help are referred as follows: -

We referred paper "Kisan Vikas" in which its main focus was on information and communication technology (ICT) and e-governance application that will help- farmers. The objective of the application was to provide government helplines, weather forecasts, news, mandi (market) prices of crops to the farmer. There is also an in-built database in which the farmer can keep a track of his inventories, harvests, seeds and fertilizer purchases, vehicles and equipment, etc.

The second paper referred as" Maha Farm" also focuses on same agriculture activities such as crop data, market prices, weather and new update. The only difference is that it updates news by RSS feed.[5]

The third paper referred as "Android based solution for Indian agriculture management design paper". In this paper the application developed helps farmer to keep the track record of what's happening in farm and what needs to be done next. Its major focus is on language that is it provides different language for interaction between farmers and application. It also allows users to plan activities, organize staff, monitor agri - inputs and agricultural machineries use, undertake economic and financial analysis of activities i.e. allows farmers to record cropping, machinery procedures and access this information.[2]

The fourth paper referred as" Krishi Ville". It takes care of the updates of the different agricultural commodities, weather forecast updates, agricultural news updates. The application has been designed taking Indian farming in consideration. [1]

III. RESEARCH CHALLENGES IN AGRICULTURAL DOMAIN

- Indian farmers are troubled by several problems. They can be natural problems or can be a problem occurred by human being. The major problem of Indian agriculture is that they are not highly literate which causes big issue in farmers for understanding different and new forms of farming technique. Pollution and environmental issues are also the different forms faced by Indian agriculture.
- Indian farmers never use proper management for crop cultivation. A further problem is a lack of co-ordination along the agricultural value chain from farm inputs to farm processing, which increases the cost of production and lower down revenue for farmers. [2]

- Farmers also rely on weather forecast as what to do now and what not to which causes them to wait for cultivation and farming. [4]
- The farmers do not get full benefits of government services so advances in Information Communication Technologies (ICT) to foster e-governance is important as the farmers will get the full benefits of services provided by the government. [5]

IV. PROPOSED METHODOLOGY

- In the proposed system, a farming assistance application based on android operating system is developed which is platform independent.
- Here, the farmers will first register and use this app.
- They can access different functions such as weather and news feed update and can get details of different crops.
- Farming tools on rent will also be available through the app.
- Farmers will upload the image of ration card and the admin will verify whether the respective farmer is BPL or not.
- The government will provide 50 % subsidy for BPL farmers while buying pesticides.

The flow of the proposed framework is as follows:

1. The application will help farmers who are below poverty line, by scanning ration card color through which government will provide subsidies to them. For this purpose Registration will be done with help of android registration GUI page
2. After registration, the farmer can access their account with the help of Login page.
3. They will also get an expert advice or a message by just clicking or putting up their query on application.
4. They will be updated with the weather information on daily basis.
5. The farmers can ask for the farming tools by updating a query in the application.
6. After updating the query the customers(farmers) will revert back based on his/her requirement.

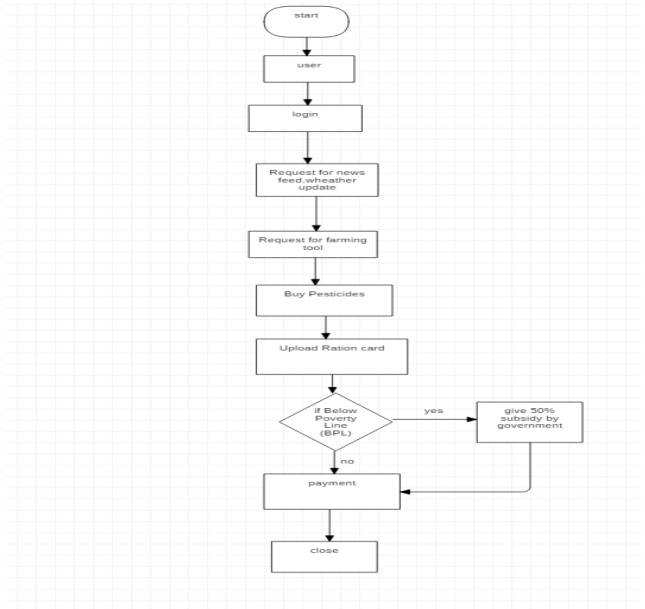


Fig 1.1 Flow Chart Diagram

V. CONCLUSION

Different apps are developed and used by farmers for their specific tasks. All these apps have different functionalities and need to be used separately. Our app is a one stop solution to all that agricultural information and needs. Through this app, location specific information is also provided. It provides a highly authentic and reliable database on agriculture. This data is accurate and is acquired from reliable sources. Thus, unlike other farming apps which are static in nature, our app is dynamic and will prove to be better for use. Also, as all the required information and functionalities are bundled in single app along with support of native language for the farmers, it becomes easy for utilisation.

REFERENCES

[1] Manav Singhal, Kshitij Verma, Anupam Shukla, "Krishi Ville – Android based Solution for Indian Agriculture", IEEE Transactions, 2014.

[2] Singhal M., Verma K, Shukla A., "Android Based Solution Indian Agriculture", IEEE 5th International Conference Advance Network and Telecommunication System (ANTS), DEC-2011

[3] Ghosh, S Garg , A.B; Sarcar, S.; Sridhar, P.S.V.S. "An interface for Indian farmer", IEEE Students Technology Symposium(TechSym), March-2014, pp-259-63.

[4] Santosh G.Karkhile ,Sudarshan G.Ghuge, "A Modern Farming Techniques using Android Applications" ,IJIRSET ,OCT-2015.

[5] Arpit Narechania , " Maha Farm– Android Based ICT Solution in Indian Agriculture to Assist Farmers ".

[6] Aniket bhave,Rahul joshi, Ryan Fernandez, "Maha Farm-an android based solution for remunerative agriculture",APRIL-2014

[7] Suporn Pongnumkul, Pimwadee Chaovalit and Navaporn Surasvadi – "Applications of smartphone-based sensors in agriculture : A Systematic Review Of Research ", JULY 2015

[8] Sumitha Thankachan,Dr. S.Kirubakaran , "E-Agriculture Information Management System" ,IJCSMC ,MAY 201Jesus Martin T

Smart Water Management System and IOT

Shraddha Shrikrishna Adgaonkar, Rutuja Pradip Ambekar, Bhavana Dyaneshwar Deshmukh, Arvind Pandhari Meshram

Electronics & Telecommunication, NDMVPS'S KBTCOE, NASHIK
shraddhaadgi1998@gmail.com, rutujaambekar16@gmail.com, bdeshmukh579@gmail.com

***Abstract*—This paper presents an IOT device which help to manage and plan the usage of water. This system can be easily installed in residential societies. Sensors placed in the tank which continuously informs the water level at the current time. This information will be updated on the cloud and using an android application, user can visualize the water level on a Smartphone anywhere that is connected to Internet. According to the level of water in the tank the motor functioning will be automatically controlled, at low level of water motor will automatically turn on and when tank is about to fill up it will cut off.**

Key words —IoT, Level Sensor, Android application, Arduino and LCD.

I. INTRODUCTION

Water is one of the maximum critical fundamental desires for all residing beings. Water is an important part of ordinary life. Due to global environmental scenario, water control and conservation is crucial for human survival. In an Internet of Things (IoT) ecosystem, two things are very important: the Internet and physical devices like sensors and actuators. Internet of Things (IoT) has taken the world by storm since its conception. The idea of connecting everything by wireless technology sums up IoT. We can connect anything using the sensors designed specifically for objects. The main purpose of sensors is to collect data from the surrounding environment. Sensors, or 'things' of the IoT system, form the front end. These are connected directly or indirectly to IoT networks after signal conversion and processing. IoT is the network of objects, devices, automobiles, houses and other items embedded with electronic sensors, and connectivity to enable them to talk to each other and execute functions. IoT is evolving fast alongside with the latest innovation occurring in wireless technology and embedded technologies, With Microcontrollers working on low powers introduced that are perfect for remotely deployed IoT systems to connect us and work for years without any maintenance has made the IoT not only for luxury functions but also for needful data aggregation as for defense systems. The devices participating in IoT are designed to be interoperable with different vendors of embedded controllers as well as with different wireless technologies. IoT is progressing with millions of things connecting each day to generate large amount of information resulting in useful future actions. This work focuses on a solution for 'Water management' in urban areas with the help of IoT. Water is precious and the supply needs to be regulated. Water demand is exponentially growing high with the increase in population of the urban areas. To maintain the supply demand ratio proper, It is important to have systems to prevent any water loss & hence we have designed an IOT system with we can plan usage of water according to the availability. People with the help of smartphones can now connect to the internet without the need for a computer, while still offering the same functionality but through different means. With the introduction of advanced software and hardware devices, smartphones are now powerful devices and have become an important part of people's daily lives. A major aspect is how the Smartphone is able to connect and communicate with other devices. A field that is recently gaining popularity is Water Level Monitoring which can also use smartphones as information or functionality hubs.[1], [2].

II. MOTIVATION

Water is precious and the supply needs to be regulated. Water demand is exponentially growing high as there is a rapid increase in population in urban areas. To control & minimize existing problems thise system proposed. To maintain the supply demand ratio, It is important to have systems to prevent any water loss & hence an IOT system designed which will plan usage of water according to the availability.

III. OBJECTIVES

- To be aware of the water level and save water accordingly.
- To make our work easy and save our time as this system works automatically.
- To relate our physical devices through internet of things and practically use it to alert system using smart phone.
- To monitor the water level so that they help in detecting the wastage of water and measures can be taken to avoid unnecessary overflowing of water in the areas where monitoring is a difficult task.

IV. ANALYSIS AND DESIGN APPROACH

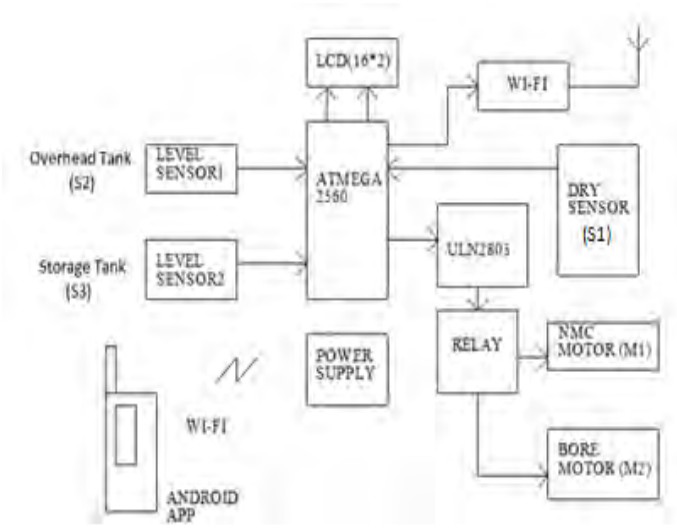


Fig 1 Block Diagram of System

A. ARDUINO ATMEGA2560

ATMEGA2560 based on widely explored esp8266 System on Chip from Express if. It combined features of WIFI access point and station + microcontroller and uses simple LUA based programming language. It is Ardiuno-like hardware I/O. There is no much difference between Arduino Uno and Arduino Mega except later comes with more memory space, bigger size and more I/O pins.[3]

a. Arduino Atmega2560 Features

1) High Performance, Low Power Atmel AVR 8-Bit Microcontroller
2) Advanced RISC Architecture
3) High Endurance Non-volatile Memory Segments
4) Endurance: Up to 64Kbytes Optional External Memory space
5) Atmel QTouch library support
6) JTAG (IEEE std. 1149.1 compliant) Interface

B. LEVEL SENSORS

Level sensors detect the level of water in the tank. Level sensors are used to monitor and regulate levels of water within a tank. A level sensor is one kind of device used to determine the liquid level that flows in an open system or closed system. We have made the capacitive sensor using signal conditioning circuit to detect the level of water in percentages for monitoring. They are operated by using a probe for checking level changes. These level changes are transformed into analog signals. The probes are generally made of conducting wire by PTFE insulation. But, stainless steel probes are extremely responsive and hence they are appropriate for measuring non-conductive substance granular or materials with low dielectric constant. These types of sensors are very simple to use and clean as they do not have any moving components. Here the level sensors used to determine the level of liquid whether that is high or low.

C. DRY SENSOR

Dry sensor is used to detect the presence of water in the NMC Pipeline. Depending upon this output of sensor the motor will be turned on and off automatically.

D. ULN2803

It is high voltage and current transistor array IC used with microcontroller where we need to drive high power loads. The ULN2803 device is a 50 V, 500 mA Darlington transistor array. The device consists of eight NPN Darlington pairs that feature high-voltage outputs with common-cathode clamp diodes or switching inductive loads. The collector-current rating of each Darlington pair is 500 mA. The Darlington pairs may be connected in parallel for higher current capability. Applications include relay drivers, hammer drivers, lamp drivers, display drivers (LED and gas discharge), line drivers, and logic buffers. The ULN2803 device has a 2.7-k series base resistor for each Darlington pair for operation directly with TTL or 5-V CMOS devices.

E. WIFI-MODULE

The ESP8266 is a low-cost Wi-Fi microchip with full self-developed high performance MCU and microcontroller capability. This is the medium to transfer the information on the android application. This module has a powerful enough on-board processing and storage capability that allows it to be integrated with the sensors and other application specific devices through its GPIOs with minimal development up-front and minimal loading during runtime.[4]

a. Wifi-Module Features

1) Processor: L106 32-bit RISC microprocessor core based on the Ten silica Xtensa Diamond Stan-dard 106Micro running at 80 MHz[5]
2) Memory:
   i. 32 KiB instruction RAM
   ii. 32 KiB instruction cache RAM
   iii. 80 KiB user-data RAM
   iv. 16 KiB ETS system-data RAM
3) 3. IEEE 802.11 b/g/n Wi-Fi
4) 4. Integrated TR switch, balun, LNA, power ampli_er and matching network.
5) 5.WEP or WPA/WPA2 authentication, or open networks
   i. GPIO pins
6) 7.SPI
7) IC (software implementation)
8) 9. IS interfaces with DMA (sharing pins with GPIO)

F. LCD

LCD (Liquid Crystal Display) screen is an electronic display module. The proposed system uses 16x2 LCD display is very basic module. These modules are preferred over seven segments and other multi segment LEDs. The real time monitored information will be displayed on this LCD. This LCD helps us to see the information manually.

G. *RELAY*

Relays are switches that open and close circuits electromechanically or electronically. Relays are used where it is necessary to control a circuit by a separate low-power signal, or where several circuits must be controlled by one signal. Relays control one electrical circuit by opening and closing contacts in another circuit. This relay is used to turn the motor ON and OFF.

H. SOLENOID VALVE

A solenoid valve is an electromechanical device in which the solenoid uses an electric current to generate a magnetic field and thereby operate a mechanism which regulates the opening of fluid flow in a valve. Solenoid valves differ in the characteristics of the electric current they use, the strength of the magnetic field they generate, the mechanism they use to regulate the fluid,and the type and characteristics of fluid they control. The mechanism varies from linear action, plunger-type actuators to pivoted-armature actuators and rocker actuators. The valve can use a two-port design to regulate a flow or use a three or more port design to switch flows between ports. Multiple solenoid valves can be placed together on a manifold. Solenoid valves are the most frequently used control elements in fluidics. Their tasks are to shut off, release, dose, distribute or mix fluids. They are found in many application areas. Solenoids offer fast and safe switching, high reliability, long service life, good medium compatibility of the materials used, low control power and compact design.

V. FLOW CHART

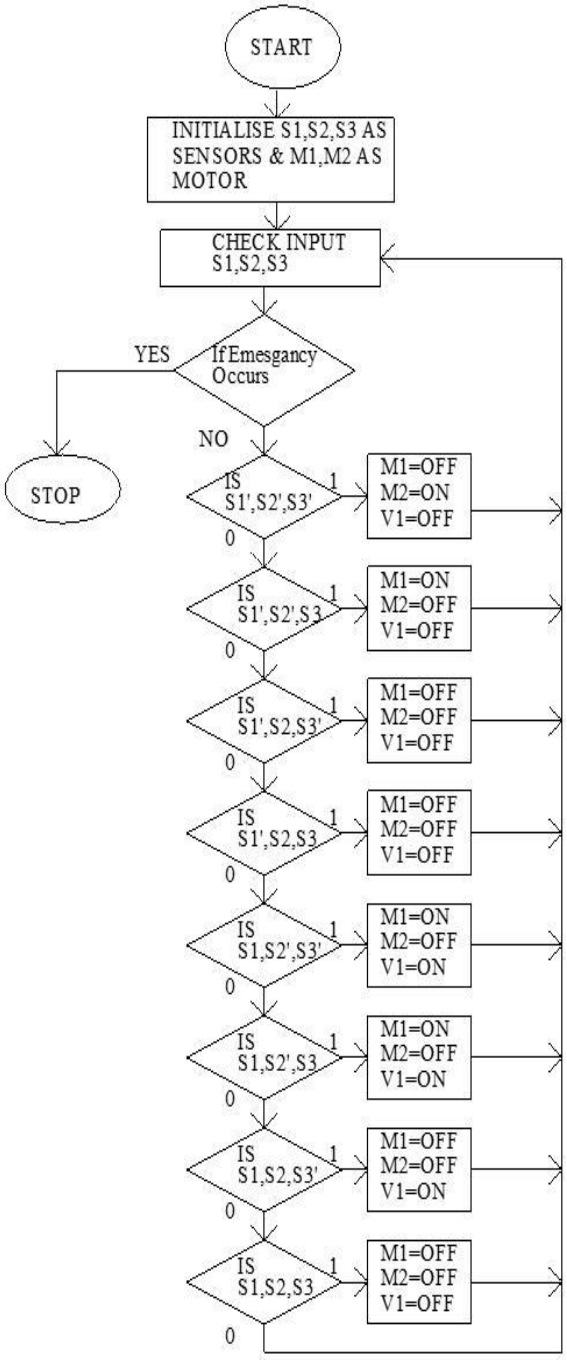


Fig 2 Flow chart

VI. WORKING

Initialize S1, S2 and S3 as sensor1(dry sensor), sensor 2 and sensor3(HIGH/LOW level sensors) which are placed in OHT and ST respectively.V1 as Solenoid valve, M1 and M2 as motors.

Now, check the output of s1 is HIGH / LOW simultaneously, checking whether V1 is ON/OFF. Then checking the conditions of S2 & S3 is HIGH /l LOW.

Conditions to be satisfied:

1. If s1's output is LOW then V1 will be OFF and simultaneously S2 & S3's output will be LOW and then M2 will be ON.
2. If s1's output is LOW then V1 will be OFF and simultaneously S2 is LOW & S3's output HIGH then M1 will be ON.
3. If s1's output is LOW then V1 will be OFF and if S3 is LOW & S2 is HIGH then no motor will be ON
4. If s1's output is LOW then V1 will be OFF and if S3 & S2 is LOW then no motor will be ON
5. If s1's output is HIGH then V1 will be ON and S3 & S2 are LOW then M1&M2 motor will be OFF
6. If s1's output is HIGH then V1 will be ON and if S2 is LOW & S2 is HIGH then M1 will be On
7. If s1's output is HIGH then V1 will be ON and S3 is LOW & S2 are HIGH then M1&M2 motor will be OFF
8. If s1's output is HIGH then simultaneously S3 & S2 is also HIGH then no motor will be On and V1 will be OFF automatically.
9. This process will be continued otherwise in emergency, hence, checking every condition.

- Sensor wires in the overhead tank will detect the level of water. Level sensor is used as sensor. Sensing will be done by operating sensor by detecting level of the water. When a particular water level will be sensed and the corresponding level will be send to the controller it will display the corresponding level of water on the LCD.
- This data will be fetched by the WI-FI application and will be displayed to the end user. In this android application recent and previous water levels will be displayed along with date and time.
- When the output of dry sensor goes high , NMC motor will automatically turn on and when output of dry sensor goes low, bore motor will turn on automatically.
- In this android application recent and previous water levels will be displayed.
- This system presents an IOT device which help to manage and plan the usage of water. This system can be easily installed in residential societies. Sensors placed in the tank which continuously informs the water level at the current time. This information will be updated using an android application, user can visualize the water level on a Smartphone anywhere that is connected to Internet. According to the level of water in the tank the motor functioning will be automatically controlled, at low level of water motor will automatically turn on and when tank is about to ll up it will turn o .This work focuses on a solution for Water management in urban areas with the help of IoT.[6]

VII. APPLICATIONS

1. Irrigation systems.
2. Industrial Automation.
3. Home Automation.
4. Institutions like Colleges,schools,hospitals.

VIII. ADVANTAGES

1) Cost savings: water management costs can be reduced through improved asset utilization, process efficiencies and productivity. Customers and organizations can benefit from improved asset utilization (e.g., smart water irrigation units that eliminate manual operation) and service improvements (e.g., remote monitoring of irrigation conditions)
2) Asset utilization: with improved tracking of assets (machinery, equipment, tools, etc.) using sensors and connectivity, companies can benefit from transparency and visibility into their assets and supply chains. They can easily locate assets and run preventive maintenance on critical pieces of infrastructure and machinery.
3) Productivity increase: Productivity is a critical parameter that affects the profitability of any organization. IoT allows real-time control, new business models, process optimization, resource conservations, service time reduction, and the capability to do all of this globally, reducing the mismatch of required vs. available skills and improving labor efficiency.

IX. LIMITATIONS

1) Privacy/Security: With all of this IoT data being transmitted, the risk of losing privacy increases.
2) Safety: As all the household appliances, industrial machinery, public sector services like water supply and transport, and many other devices all are connected to the Internet, a lot of information is available on it. This information is prone to attack by hackers.
3) It would be very disastrous if private and confidential information is accessed by unauthorized intruders.

X. CONCLUSION

In this project, we have been discussing about the elements of Water tank system and Smart phone. The proposed system actualized the discontinuous notice framework using the current tank system. There are a few preferences. Our connotation of this study is to become a line up to a flexible and competitively priced, simple configurable and most significantly, a transportable system which may solve our water wastage drawback. We have used ESP and level sensor which reduces cost effectively and makes this project economical. Also, this project doesn't require special different tank for it, existing water tanks can be used. We have successfully implemented this project.

REFERENCES

[1] Shifeng Fang, Li Da Xu, Senior Member, IEEE, Yunqiang Zhu, Jiaerheng Ahati, Huan Pei, Jianwu Yan, and Zhihui Liu "An Integrated System for Regional Environmental Monitoring and Manage-ment Based on Internet of Things", IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, VOL. 10, NO. 2, MAY 2014.

[2] Final Technical Report to UK Department for International Development, "Smart Water Sys-tems", April, 2011.

[3] Joy Shah1 P.G. Student, Department of Information Technology, Dharmsinh Desai University,Nadiad, Gujarat, India, "An Internet of Things Based Model for Smart Water Distribution with Quality Monitoring", International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET), V Vol. 6, Issue 3, March 2017.

[4] Tomas Robles,Ramon Alcarria,Diego Martn,Mariano Navarro3,Rodrigo Calero3,Sofa Iglesias3 "An IoT based reference architecture for smart water management processes".

[5] Smart Water Measurements: Literature Review Berhane Gebremedhin Water-M Project Center for Wireless Communications (CWC) Department of Communications Engineering, University of Oulu, June 13, 2015.

[6] Patawala Amatulla .H, Bansode Navnath .P, Bhong Yogesh, Prof. Zadbuke Ashwini .S, "IOT Based Water Management System for Smart City", International Journal of Advance Research, Ideas and Innovations in Technology,(Volume3, Issue2)

Heart Attack Detection using sensors and IOT

Miloni Nisar[1], Purvesh Dave[2], Shridhar Kamble[3].
Department of Information Technology, Thakur College of Engineering and Technology, Mumbai, India

***Abstract*-The proposed system consist of geophones which are used to sense bed vibrations caused by the ballistic force. The geophones helps in monitoring person's heart rate during sleep. This mechanism for heart rate monitoring does not require a special kind of bed. This geophones are placed at the side of the bed, which allows person to sleep in a comfort manner as well as person can move around and change position freely during sleep. Geophones are integrated with Arduino, which connected to the ac amplifier. The vibrations sense by the geophones are converted from digital signal to analog signal which is to be displayed on to the monitoring screen. Geophones technique only resolve the problem for monitoring heart rate on bed. To overcome this problem wristband is used. Wristband helps in detection of the heart rate and it is also used for early detection of heart attack using machine learning algorithms. Pulse detection sensor is used to detect pulse rate. Geophones and Wristband techniques will wirelessly send alert to 5 recommended people.**

Keywords– Heart Rate Detection ,Arduino board, Wireless Transmission of data,sms or calling alert to doctor, predicting the heart attack.

I. INTRODUCTION

A wireless sensor network design for real-time monitoring and detection cardiovascular disease. This system incorporates wireless sensor network technology with other wireless technologies such as cellular network, wireless LAN, and broadband network, for efficient and fast delivery of health alerts. This proposed system consists of a wearable wireless sensor system, control system, heterogeneous wireless network system, two phase real-time data analysis and visualization system, and the warning system. All these together will provide the ECG signal analysis of a cardiac patient, remote monitoring, delivery of warning to a doctor, relative, and the hospital, and an extended service of transferring the ECG signals and previous records of the patient to his doctor in a remote location. This system can be used for providing enhanced health care services to the rural areas of the developing countries that are facing shortage of efficient specialized doctors. Thus the cardiovascular disease causing the death of patient can be reduced immensely by the implementation of this proposed system, in both rural and urban areas.

Using Geo phones to sense bed vibrations caused by ballistic force has shown great potential in monitoring a person's heart rate during sleep. It does not require a special mattress or sheets, and the user is free to move around and change position during sleep. Earlier work has studied how to process the geophone signal to detect heartbeats when a single subject occupies the entire bed. In this study, we develop a system called VitalMon, aiming to monitor a person's respiratory rate as well as heart rate, even when she is sharing a bed with another person. In such situations, the vibrations from both persons are mixed together. VitalMon first separates the two heartbeat signals, and then distinguishes the respiration signal from the heartbeat signal for each person. Our heartbeat separation algorithm relies on the spatial difference between two signal sources with respect to each vibration sensor, and our respiration extraction algorithm deciphers the breathing rate embedded in amplitude fluctuation of the heartbeat signal.

Background

Many heart rate sensing devices are available in the market, and they give high accuracy rate in providing the pulse rate.FitBit Versa is a wrist band watch used for sensing the heart rate and pulse rate of the body and it also tracks other activities, for instance while jogging it counts the kilometers and the footsteps of the user.

Secondly, FitBit band also plan the user dual activity which has to be performed .This activity is based on the past records. Apple Watch Series 1 is a kind of the wrist watch which sense the heart rate, maintain some diagnostics and integration to this it is also used for picking up the call and showing content of the iPhone, iPad, and iMac on the watch.

This devices are used only for sensing the heart rate while the device is on the hand but our system which can sense the ballistic force with the help of the geophones.

Importance of the project

The project aims at creating asystem which is able to sense the heart rate using geophone sensor and the pulse band sensor intelligent but lightweight healthcare assistant. This application will be able to help healthcare and other user suffering from heart diseases. This also aims at having a basic machine learning capability for heart attack prediction.

In an era where medical diagnosis and treatment has made massive leaps, it is unfair to deny anyone's proper healthcare owing to a lack of means.

One often sees a lack of a proper software to help it out in such cases. Software can avoid the necessity of having to revisit doctors, or to spend hours self-researching complex medical terms and appendices, as many diagnosed patients tend to do. The use of a band here can calculate the rate every minute and pass the data to the database.

Objectives and Scope of the project

The proposed system consists of geophones which are used to sense bed vibrations caused by the ballistic force. The geophones help in monitoring person's heart rate during sleep. This mechanism for heart rate monitoring does not require a special kind of bed. These geophones are placed at the side of the bed, which allows person to sleep in a comfort manner as well as person can move around and change position freely during sleep. Geophones are integrated with Arduino, which connected to the ac amplifier. The vibrations sense by the geophones are converted from digital signal to analog signal which is to be displayed on to the monitoring screen.

As Geophones technique only resolve the problem for monitoring heart rate on bed, to overcome this problem pulse- wrist-band is used which is equipped with BLE Nano Circuit, Pulse Sensor, 3V Battery along with nRF Toolbox application for BLE for demonstrating live activity of heart rate.

Pulse-wrist-band captures the heart beats continuously and sends data to nRF Toolbox application using BLE Nano which can be analyzed using machine learning algorithms to predict early detection of heart attack using abnormal patterns of heart beats. Pulse-wrist-band will send alert to 5 recommended people in emergency.

Other scopes of this project are:

This system can be used for providing enhanced healthcare services to the rural areas of the developing countries which are facing shortage of efficient specialized doctors. Thus, the cardiovascular disease causing the death of patient can be reduced immensely by the implementation of this proposed system, in both rural and urban areas.

Heart hate monitoring will help patient understand the nature of the body system.Cost Effective system with multiple features embedded on a single unique system.Wristbands and geophones help in early detection of the heart attack or stroke and help save lives.

It helps in alerting you and your near ones about sudden changes and help can be provided in medical emergency.

Besides the features above, we also aim to make the software robust, accurate and fast.

Summary

Healthcare, though a new venture for the artificial intelligence\ Machine learning domain, remains one of the most crucial domains of public service. With big data growth in biomedical and healthcare communities, accurate analysis of medical databenefits early disease detection, patient care, and community services.



It is pertinent that we utilize mathematical modelling to iron out any imperfections and aberrations, so that we have the rigorous standards of accuracy that the medical diagnosis field requires.

Many virtual assistants are available on the market, and they give high accuracy in handling input, particularly Siri and Google Assistant. Although not anywhere close in accuracy, the application understands most commands with enough accuracy to extract results through them.

The project aims at creating an intelligent but lightweight healthcare assistant. In an era where medical diagnosis and treatment has made massive leaps, it is unfair to deny anyone proper healthcare owing to a lack of means. This project aims to make healthcare less exclusive than it currently is and make it easier to find and avail services. This application will be able to help healthcare and other hospital staff work efficiently.

Stakeholder perspective is crucial to the success of this project. It is also vital to involve the patients, as the patients being comfortable with the equipment, both hardware and software, is necessitous to the smooth working of the technology. Potential users seemed open to the idea, as anticipated.

Stakeholders here include - Doctors and nurses, Health insurance companies, Hospitals, Health department (State as well as National/Federal), Biotechnology manufacturers.

This application will be able to help healthcare and other hospital staff work efficiently. This also aims at having a basic machine learning capability for disease prediction.

The main objective is to offer a way for doctors and patients alike to easily perform mundane healthcare activities as well as advanced prediction services.

II. DESIGN AND IMPLEMENTATION

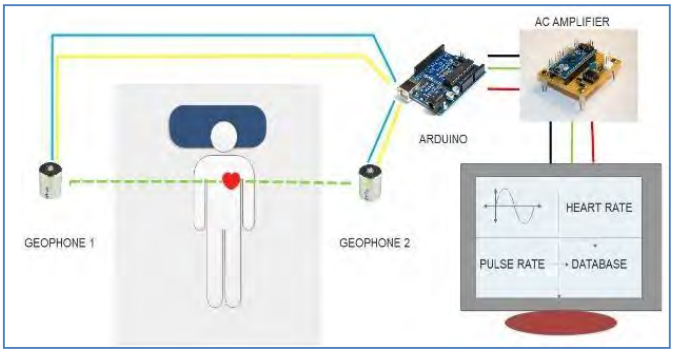
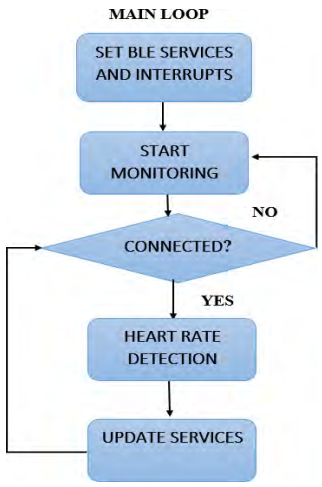


Fig 1.: Geophone Sensor Based Heart Rate Detection and Monitoring.

R

Fig 2.: Heart Rate Detection and Monitoring using Pulse Sensor Band.

Fig 3.: Flow Chart for Heart rate detection.



Outputs:

We were able to build a pulse sensing wristband which is able to sense the pulse rate of the body.

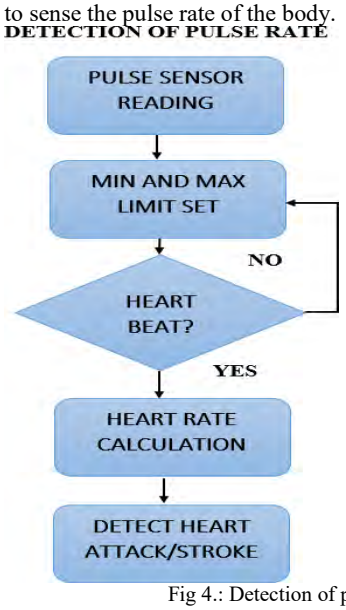


Fig 4.: Detection of pulse rate.

Besides this, we were also able to add some basic features such as transferring that data to the device and showing the pulse rate on the device. The device which is required can be a monitor or mobile device which can be connected to BLE Nano and can easily get the output.

Outcomes:

The project intends to correctly analyze the pulse rate using pulse band sensor and displaying it on the screen. It uses BLE Nano which is for transferring data from the sensor to the device.

Discussion:

AI systems are now a part ofa variety of specializations from stock trading to the setting of real estate prices; from frauddetection to translationof vernacular languages; from designing our daily shopping lists to forecastingwhich movies

t

we might like and diagnosis of medical condition.

This is merely the beginning. Not long from today, AI will be used to advance our understanding of biological phenomenon through analysis of highly dense and huge datasets, help us invent new drugs and make treatments more thorough and personal. Self-driving vehicles canrevolutionize transportation and allow new paradigms in planning of architecture. Machines will run our homes more efficiently, make businesses more productive and help predict risks to society.

Before AI systems can be deployed in healthcare applications, they need to be 'trained' through data that are generated from clinical activities, such as screening, diagnosis, treatment assignment and so on, so that they can learn similar groups of subjects, associations between subject features and outcomes of interest. These clinical data often exist in but not limited to the form of demographics, medical notes, and electronic recordings from medical devices, physical examinations and clinical laboratory and images.There are several techniques that can be used to correctly classify and predict when a new tuple is entered.

Future Scope

Currently the project focuses on gathering a structured authentic data. As medical data and analysis is a sensitive field, the project intends to find an assimilation of the results of various algorithms and compare the accuracy.

It's a challenging task for monitoring and detect heart rate and predict heart attack/ stroke from pulse rate itself. There are other parameters also by which heart attack/ stroke may occur such as patient's diabetic, blood pressure history and all. So, we can consider these parameters also to develop enhanced health monitoring system.

Deployment

Deployment can be a major challenge in software and hardware which involves machine learning and software which will help in detecting the early heart attack from the given pulse rate and heart rate which is collected by geophones and wrist pulse band. Huge amounts of processing is required just to get it up and running. Data generated in the process is humongous as well.The ML model will be deployed using Google firebase or spring. The cloud platform will provide a base for future developments where flexibility, portability and reliability is required.The GUI by itself is not very difficult to deploy. We have multiple ways of offering interfaces, such as web interfaces and command line interfaces. It will expand to a desktop GUI in the future as well.The web interface will be deployed using heroku or AWS, depending on the relative cost of hosting.The command line can be used as a windows executable file. Users can simply download it and run it natively on their windows PCs

III. CONCLUSION

The project aims at creating a system which is able to sense the heart rate using geophone sensor and the pulse band sensor intelligent but lightweight healthcare assistant. This application will be able to help healthcare and other user suffering from heart diseases.Our heartbeat separation algorithm relies on the spatial difference between two signal sources with respect to each vibration sensor, and our respiration extraction algorithm deciphers the breathing rate embedded in amplitude fluctuation of the heartbeat signal.Proposed health monitoring system will monitor the heart beat fluctuation using geophones and monitoring, detection and analysis will be done using pulse-wrist-band which is equipped with BLE Nano Circuit, Pulse Sensor and 3V Battery along with nRF Toolbox application for BLE for demonstrating live activity of heart rate.Geophones are integrated with Arduino, which connected to the ac amplifier. The vibrations sense by the geophones are converted from digital signal to analog signal which is to be displayed on to the monitoring screen.Pulse-wrist-band captures the heart beats continuously and sends data to nRF Toolbox application using BLE Nano which can be analyzed using machine learning algorithms to predict early detection of heart attack using abnormal patterns of heart beats. Pulse-wrist-band will send alert to 5 recommended people in emergency.

REFERENCES

[1] Z Jia, A Bonde, S Li, C Xu, J Wang, Y Zhang, RE Howard, P Zhang, "Monitoring a Person's Heart Rate and Respiratory Rate on a Shared Bed Using Geophones", Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems, 2017.

[2] ZhenhuaJia, Richard E. Howard, Yanyong Zhang, Pei Zhang, "HB-Phone: A Bed-Mounted Geophone-Based Heartbeat Monitoring System", Proceeding of 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Pages 275-276, 2016.

[3] Chacko John Deepu, Zhihao Chen, JuTengTeo, Soon Huat Ng, Xiefeng Yang and Yong Lian, "A Smart Cushion for Real-Time Heart Rate Monitoring", IEEE Biomedical Circuits and Systems, DOI:10.1109/BioCAS.2012.6418512, 2012.

[4] FadelAdib, Hongzi Mao, Zachary Kabelac, Dina Katabi,Robert C. Miller, "Smart Homes that Monitor Breathing and Heart Rate" in 15 Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Pages 837-846, 2015.

[5] Kala John Kappiarukudil, ManeeshaVinodini Ramesh, "Real-Time Monitoring and Detection of Heart Attack Using Wireless Sensor Networks", SENSORCOMM '10 Proceedings of the 2010 Fourth International Conference on Sensor Technologies and ApplicationsPages 632-636, July 18 - 25, 2010.

[6] Chao Li, Xiangpei Hu, Lili Zhang, "The IoT-based heart disease monitoring system for pervasive healthcare service", ELSEVIER Procedia Computer Science, Volume 112 Issue C, Pages 2328-2334 , September 2017.

[7] Nicholas Constant,Tanya Wang, KunalMankodiya, "Pulseband: A hands- on tutorial on how to design a smart wristband to monitor heart-rate", IEEE Virtual Conference on Applications of Commercial Sensors (VCACS), April 2016.

[8] Grace Wusk, Hampton Gabler "Non-Invasive Detection of Respiration and Heart Rate with a Vehicle Seat Sensor", doi: 10.3390/s18051463, May 2018.

[9] D. H. PHAN, S. BONNET, R. GUILLEMAUD, E. CASTELLI, N. Y. PHAM THI, "Estimation of respiratory waveform and heart rate using an accelerometer", 30th Annual International IEEE EMBS Conference Vancouver, British Columbia, Canada, August 20-24, 2008.

[10] Online Web Resource referred on 5-08-2018 https://www.livemint.com/Science/6j9LCCGcUiOVgG0gHCkr0O/25-million- died-due-to-pollution-in-India-in-2015-Lancet-s.html

Career Counselling using Machine Learning Algorithms

Pranav Shah, Shohan Shetty, Lakshit Ukani,Priyank Jethva
Department of Information Technology
St. Francis Institute of Technology
Mumbai, India
shahpranavpks@gmail.com, shettyshohan55@gmail.com , lakshitukani77@gmail.com, jethvapriyank@gmail.com

***Abstract*- In this paper, we propose the design of an efficient and comfortable option to predict career using machine learning algorithms. Problems like increased dropouts, poor performances, surprised college, degree changes and costly readmissions impacts one's role in job and real-world industry. According to the survey conducted by the Council of Scientific and Industrial Research (CSIR), about 40% of students are confused about their career options [1]. This leads the students to select a wrong career and may result them working in a field which was not meant for them. With machine learning algorithms, we can analyze the current trends and statistics of various students and utilize them to predict an appropriate career for freshmen. Hence, in the proposed system, we will build a machine learning model using alumni data and the model that gives out the highest accuracy between Naive Bayes and Random Forest classifier will be used to predict career for students.**

Keywords- career; counselling; machine learning; naïve bayes; random forest; prediction; classification; student

I. INTRODUCTION

Educational institutions are important part of our society playing a vital role for growth and development of the nation. They not only nourish student academically but also help them excel in extracurricular activities. How to predict a students' career is always a question of concern by teachers and parents. With the increase in research in a number of domains, there are various career opportunities available to students today. This creates even more confusion to students to select a single career option, resulting into consulting multiple counsellors. At times, people tend to get deceived by fraudulent counsellors who admit students into colleges and help them select a stream which gets them a commission. Also, visiting a number of counsellors is very time consuming and involves a lot of capital. Counsellors suggest a career to students based on their knowledge and perspective only. Hence, to be certain, students tend to take guidance from various people. Career counselling is a process that guides a person to know and understand his capabilities to pursue a fitting career in order to excel in life. The proposed system aims to provide the students with a platform that will predict their career using Machine learning classifier – Naïve Bayes or Random Forest.

II. LITERATURE SURVEY

While searching for details on predicting a student's career using machine learning algorithms, we came across methodologies and researches using wide range of algorithms; but each one has some limitations. We have found some papers published which relate to predicting student characteristics. Most of these papers predict career by focusing on just academics or test scores but do not involve non-academic factors such as personality, family financial status, hobbies or interests. While interacting with many graduates we came to know that these factors do play an important role in deciding a career.

As stated by Ying Cao, Lei Zhang [2], their system consists of the basic intelligence career counseling, career counseling solutions management, auxiliary decision-making and information management. The system is based on the Agent technology to achieve the Java development of Environment. This system does not involve predicting a career but rather just provides solution to professional problems based on the relevant information obtained from the visitors.

According to a paper published by Roshani Ade and P. R. Deshmukh [3], training set for the supervised data mining algorithms can be formed using students' marks and results. Against to the classical batch learning algorithm, incremental learning algorithm tries to forget unrelated information while training new instances. The ensemble technique proposed in this paper is compared with the incremental algorithms, without any ensemble concept, for the student's data set and it was found that the proposed algorithm gives better accuracy. But this approach does not consider non-academic factors involved in predicting a career.

As stated by Ezenkwu, C.P, Johnson, E.H., Jerome, O.B. [4], a majority of tertiary education aspirants are faced with the challenge of selecting courses that will best match their academic capabilities. This paper is aimed at developing an Automated Career Guidance Expert System (AC-GES) using case-based reasoning (CBR) technique. But the CBR algorithm gives excessive classification error on the test-case. Also, this algorithm involves over-fitting problem.

According to Nikita Gorad [5], who created a web application that would help students studying in high schools to select a course for their career. The system would recommend the student, a career option based on their personality trait, interest and their capacity to take up the course. But this C5.0 algorithm gave only 66% accuracy which is quite low.

Our paper presents the concept of prediction using two most efficient and easy to use algorithms and considering all attributes of student ranging from academic to non-academic which affect the decision making process. The accuracy for both algorithms – naïve Bayes and Random Forest- would be measured and the most efficient algorithm will be used to predict career.

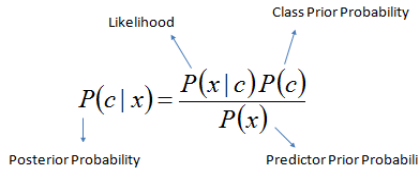
III. METHODOLOGY

MACHINE LEARNING:

The Machine Learning gives computer the ability to learn without being explicitly programmed.[6]. The process of learning uses a dataset to look for patterns in the data and make better decisions in the future based on the input data [7]. The two algorithms that we will use in the proposed system are Naïve Bayes and Random Forest classifiers.

A. Naïve Bayes Classifier:

Naive bayes classifier is a probabilistic classifier model based on Bayes' theorem. Bayes theorem provides a way of calculating the posterior probability, P(c|x), from P(c), P(x), and P(x|c). Naive Bayes classifier assume that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors.



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Fig. 1. Posterior Probability formula

- P(c|x) is the posterior probability of class (target) given predictor (attribute).
- P(c) is the prior probability of class.
- P(x|c) is the likelihood which is the probability of predictor given class.
- P(x) is the prior probability of predictor.

Naive bayes classifier works well when all the predictors are independent. In our case the attributes used to make prediction such as marks, hobbies, financial status etc are all independent of each other. Hence Naive Bayes classifier would perform well when compared to other classifiers due to its simplicity and reduced complexity.

B. Random Forest Classifier:

Random forest is a supervised learning algorithm which ensembles Decision tree. It works as a classifier and also as a regressor. In simple terms, random forest builds multiple decision trees and merges them to generate a more accurate result.

In our case, the random forest will consider all attributes and find the best feature among a random subset of attributes. This would result in a wide diversity that generally provides better results. The disadvantage of Random Forest algorithm is that it tends to become slow with the increase in number of trees and thus becomes ineffective for real-time predictions. Hence, this algorithm is fast to train, but quite slow to create predictions once they are traine. An accurate prediction requires more trees at the cost of a slower model [8].

IV. PROPOSED SYSTEM

According to a survey conducted by us among 70 students of the age group 14 to 25 years, we can see in fig. 2 that around 40% of students do not make a complete research before selecting a career which may be due to various reasons such as lack of resources or incomplete knowledge. These people are then confused with many career options and tend to visit counsellors.

Have you made a complete research before selecting your career?

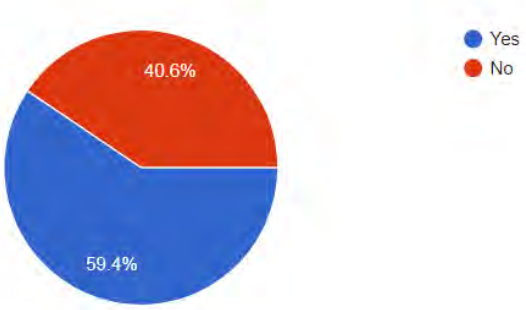


Fig. 2. Chart on research

As per the survey, figure 3 depicts that people had an average experience with the counsellors.

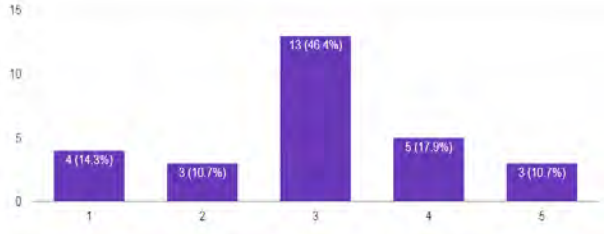


Fig. 3. Experience with counsellors

Also, people tend to take advice from various other friends or relatives which is represented using the graph in the below fig. 4.

Do you take advice from seniors, friends or relatives before selecting a career?

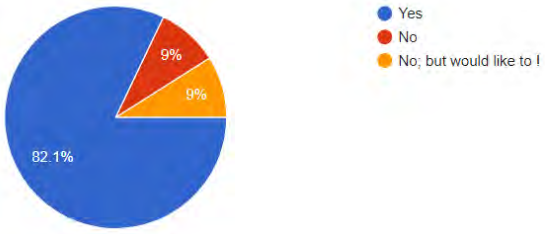


Fig. 4. Chart on advice

The fig. 5 shows that people consider career counselling as a very crucial step in their life since it helps to make correct decisions and develop a bright future.

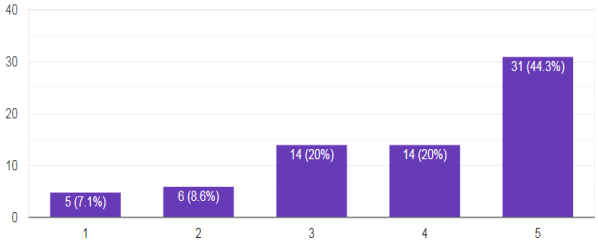


Fig. 5: Importance of career counselling

Hence, this whole process of taking consultancy from people is time consuming and involves a lot of capital. Human involving in making predictions for a student is at times biased and may be towards the overall benefit of the counsellor instead of the student. Hence, to avoid such fraud cases career can be predicted using machine learning algorithms. Predictions can be made on various parameters such as academics, psychology and personal interests. This system helps students to get first-hand information from alumni who have experienced the academics and can provide their views regarding the same.

The proposed system represented in fig. 7 aims to provide Career Counselling using machine learning algorithms such as Naive Bayes and Random forest classifier. The dataset is formed using alumni data from various colleges of different streams. Dataset is then divided into training data and testing data in the ratio 3:1. Thereafter, pre-processing of dataset is performed and later it is used to build and train Naive Bayes and Random Forest classifier models. Then, using testing dataset we evaluate the performance of the models and compare its accuracies. Performance is evaluated on basis of confusion matrix which displays number of correctly classified values and incorrectly classified values. This can be used to calculate classification accuracy using the formula shown in the below fig. 6.

$$\text{Classification Accuracy} = \frac{\text{Total number of correct predictions}}{\text{Total number of predictions}}$$

Fig. 6: Classification Accuracy

The model with the highest accuracy is selected and used to perform predictions. The user inputs its data and a career is predicted based on it using the most efficient model.

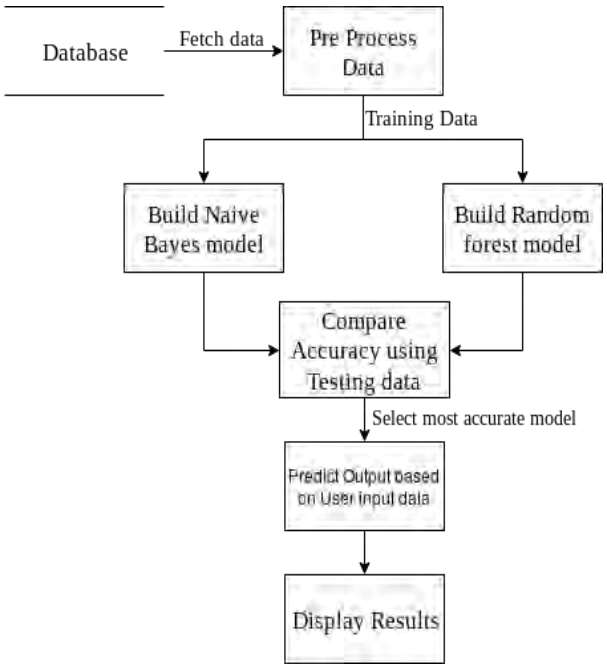


Fig. 7: Proposed System Architecture

V. CONCLUSION

Data mining has provided a great way to extract knowledge from raw data and huge data sets. In addition to it, machine learning - a branch of artificial intelligence, helps us to build models which have the ability to learn and improve without being explicitly programmed. With the help of these, our system tends to help an individual decide a career. This system will provide a personalized analysis, career options, career reviews and prospects of various fields. Thus, enabling each student to evaluate their choices and make an informed decision. Hence, the proposed system successfully addresses the issue of selecting an appropriate career in life.

REFERENCES

[1] https://timesofindia.indiatimes.com/home/education/news/Survey-shows-40-students-confused-about-career/articleshow/12888793.cms
[2] Ying Cao and Lei Zhang, "Research about the college students career counseling expert system based on agent," 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), Dengleng, 2011, pp. 3208-3211. doi: 10.1109/AIMSEC.2011.6010332
[3] R. Ade and P. R. Deshmukh, "An incremental ensemble of classifiers as a technique for prediction of student's career choice," 2014 First International Conference on Networks & Soft Computing (ICNSC2014), Guntur, 2014, pp. 384-387. doi: 10.1109/CNSC.2014.6906655
[4] Ezenkwu, Chinedu & Johnson, Enyenihi & Jerome, Otobong. (2017). Automated Career Guidance Expert System Using Case-Based Reasoning Technique. 8. 81-88.
[5] Nikita Gorad, Ishani Zalte, Aishwarya Nandi and Deepali Nayak," Career Counselling Using Data Mining", Vol. 5, International Journal of Innovative Research in Computer and Communication Engineering, Issue 4, April 2017.
[6] https://www.geeksforgeeks.org/machine-learning/
[7] https://www.expertsystem.com/machine-learning-definition/
[8] https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd

Privacy-Preserving Social Media Data Publishing for Personalized Ranking-Based Recommendation

Snehal Tukaram Patil

P.G. Student

Thakur College of Engineering & Technology , Kandivali(E),Mumbai 400101,Maharashtra ,India

snehal321.patil@gmail.com

Abstract— In most recent couple of years, there has been a hard advancement of fusing comes about beginning organized information source into watchword based web track frameworks, for example, Amazon and Google or any web indexes. In search engines, different users may seek for different information by issuing the similar query. To persuade more clients with fractional query items, output enhancement re-positions the outcomes to coat the same number of client aims as plausible. Most displayed goal mindful broadening calculations separate client aims as subtopics, everything about is regularly a word, an expression, or a bit of elucidation. Web look inquiries are regularly dubious or multi-seek, which makes an easy positioned rundown of results deficient. we proposed PrivRank, a customizable and continuous privacy-preserving social media data publishing framework protecting users against inference attacks while enabling personalized ranking-based recommendations. Its key idea is to continuously obfuscate user activity data such that the privacy leakage of user-specified private data is minimized under a given data distortion budget, which bounds the ranking loss incurred from the data obfuscation process in order to preserve the utility of the data for enabling recommendations. An empirical evaluation on both synthetic and real-world datasets shows that our framework can efficiently provide effective and continuous protection of user-specified private data, while still preserving the utility of the obfuscated data for personalized ranking-based recommendation. Compared to state-of-the-art approaches, PrivRank achieves both a better privacy protection and a higher utility in all the ranking-based recommendation use cases we tested.

Keywords— Information Filtering, Social Computing, Social Network Analysis, Topic Identification, Topic Ranking

I. INTRODUCTION

With the currently growing interest in the Semantic Web, it is reasonable to expect that increasingly more metadata describing domain information about resources on the Web will become available. The idea presented here is to enrich the search process for hypermedia applications with information extracted from the semantic model of the application domain. One of the novelties in the semantic search proposed is the combination of spread activation techniques with traditional search engines techniques to obtain its results. One of the greatest problems of traditional search engines is that they typically are based in keyword processing. Consider the following motivating example for a research institution domain. This domain deals with people, publications and research areas. Notice that "Keyword" is not a concept of the model, but is used in the diagram to repress fact that a keyword occurs inside the textual representation of the associated concept instances. For instance, the keyword "web" occurs inside the concept "The Evolution of Web Services" since it appears in the publication's "title" property. The keyword "ontology" is also related to the same concept since it appears in its "abstract" property.

A query with the keyword "web" would have as results only nodes of type Publication where this word occurs. If the user searches for nodes of type "Professor", the result could well be an empty set, since the keyword "web" may not appear inside the description text (page) of any of the professors. On the other hand, analyzing the semantics of this domain, it seems intuitive to think that if a given professor has many publications that are related to a given keyword, there is a great possibility that the professor himself is indeed related to that same keyword, and should be returned as a result of the query. Therefore, in the previous example, the Professor node "Schwabe" could be returned as a result for the query "web Spread Activation techniques are one of the most used processing frameworks for semantic networks. It has been successfully used in several fields, particularly in Information Retrieval applications. Since it was developed in the Artificial Intelligence area as a processing framework for semantic networks and ontologies, we envision it to be a natural and interesting choice of knowledge processing algorithm in the context of the Semantic Web. An interesting and recent system that uses spread activation to process ontologies is ONTOCOPI [1], which tries to identify communities of practice within an organization. The spread activation algorithm works basically as a concept explorer. Given an initial set of activated concepts and some restrictions, activation flows through the network reaching other concepts which are closely related to the initial concepts. It is very powerful to perform proximity searches, where given an initial set of concepts, the algorithm returns other concepts which are strongly connected to them. An overview of spread activation techniques. Usually spread activation techniques are used either on semantic networks (where each edge in the network has only a label associated to it) or on associative networks (where each edge has only a numeric weight associated to it). In the Semantic Web, we use ontologies as the semantic network used by the spread activation algorithms

II. LITERATURE SURVEY

In this section we refer to the pertinent past writing that uses the different procedures for positioning. Positioning query items is a principal issue in data recovery. Most normal methodologies center around the similitude of inquiry and page and general page quality. Be that as it may, with expanding prominence of web indexes the catching of client practices demands to show up on surface. A considerable measure of strategies have been

done on verifiable measures of client inclination in field of data recovery.

A novel utilization based customized page rank style calculation was introduced by Eirinaki et al [1]. This was utilized for positioning the website pages of the destinations in light of the user‟s past navigational encounters and conduct. In light of this calculation a recurrence determination for use based positioning was created. This calculation can be for the most part used to create customized suggestions.

The different issues in page positioning, for example, dangling joins, recorded qualities, false connections Google Bombing, Google jacking and so forth were advanced by Gupta et al [2]. An idea of trust rank i.e. to put a center vote of trust on an arrangement of seed set of looked into pages was presented. This trust rank can be darted to deliver more effectiveness and high exactness in acquiring the ranked pages.

An Efficient Privacy-Preserving Ranked Keyword Search Method was suggested by Chi Chen at. Al[3]. Proposed system discussed hierarchical approach clusters the documents based on the minimum relevance threshold, and then partitions the resulting clusters into sub-clusters until the constraint on the maximum size of cluster is reached. In the search phase, this approach can reach a linear computational complexity against an exponential size increase of document collection. In order to verify the authenticity of search results, a structure called minimum hash sub-tree is designed in this approach. System also investigated cipher text search in the scenario of cloud storage. System explore the problem of maintaining the semantic relationship between different plain documents over the related encrypted documents and give the design method to enhance the performance of the semantic search.
A method proposed by Feyzania et al [4] to rank the records by building an associated chart of semantic archives. This chart considers just the semantic connections between the archives .Thus a novel approach of substance based positioning was advanced as needs be by removing the verifiable positions .Different weights can be utilized to distinguish the diverse semantic connections.
A powerful way to deal with use based positioning utilizing philosophy was advanced by Junfang et al [5]. The weight count was finished utilizing metaphysics tree. This weight estimation assesses utilization data, examples and structure. The whole cosmology tree was changed over into a weighted chart and afterward avocation was connected to compute the last weight which was then utilized for seeking, positioning and conflict solving.

Automatically Mining Facets for Queries from Their Search Resultsby Zhicheng Dou et al[6] .System proposed a systematic solution, which system refers to as QDMiner, to automatically mine query facets by extracting and grouping frequent lists from free text, HTML tags, and repeat regions within top search results. Experimental results show that a large number of lists do exist and useful query facets can be mined by QDMiner. An effective approach of web service recommendation based on usage based ranking and QOS preferences was presented by Kang et al [7]. This algorithm is effective in classification and extraction of the results.

TEXplorer: keyword-based object search and exploration in multidimensional text databases by Zhao, B., et al [8] The system proposed propose a novel system TEXplorer that integrates keyword based object ranking with the aggregation and exploration power of OLAP in a text database with rich structured attributes available, e.g., a product review database.

In [9] Patil et al has called attention to that the surmising and examination of inquiry objectives can have a considerable measure of focal points in enhancing web crawler pertinence and client encounter by consolidating web utilization and web content mining.it presents a weighted strategy to mine the web content taking into account the client needs.
In [10] Kishor kumar et al proposed an effective calculation for mining the substance through the extraction of web crawler results and development of lexicon for the client question and hence extraction of catchphrases that assume a crucial part in estimation of aggregate quality of the keyword.

III. SYSTEM DESIGN

A. Objectives

- To design and implement a system which provides the privacy of online users profile for media data.
- To design a framework this can automatically classify the unknown labels as private or public parameters.
- Integrate the third party service for raking the parameters of privacy.
- Use page rank algorithm for get the ranked parameter
- Problem Statement

In the proposed research work to design and implement a system than work as classify and re-rank all type of query events along with the current events as well as news. The Google API will provide the third party interface for communicate with search engine the classify the all data using machine learning approach and re-rank with page rank as well as click through algorithms. We will collect the data from Google API, YouTube as well twitter and rank all the news base on current user query

B. Project Overview

The proposed system consolidates and rank the most prevalent topics discussed in both news media and social media during a specific period of time. The system framework can be visualized in Fig. 1. To achieve its goal, the system must undergo four main stages.

Pre-process and features extraction: In this module system extract all the information shared by user and apply similarity base weight calculation using private and public training rules. For each attribute we will generate the specific weight according to training policies.

- *Classification using privacy framework*
  This module will classify all data as private or public according to current similarity weight, the data which is classified by classification algorithm as public or private those attributes which is shared by end user
  .
- *Re-Ranking*
  Here system choose to measure the pair wise ranking loss using a widely known metric, created two private and public vectors to different users using Kendall-T distance algorithm

- *Analysis*
  System provide the comparative analysis using different parameters for accuracy time complexity using different graphs.

C. Development Methodology

The below figure 1 shows a proposed system architecture with execution process, in first phase user uploads the his own information on social media, it contains some private as well as some public information. The proposed work also carried out to identify the potential of private data and provide auto security using proposed given algorithms. Natural Language Processing (NLP) extract the features from user's shared data and apply the weight clustering using cosine similarity algorithm. Once clustering has done it will generate the three different clusters viz public, private and activity etc.

The second phase in system implementation of ranking algorithm according to their weights and apply the privacy policies on those private data attributes. In whole process system automatically maintain the privacy on vulnerable data which is shared by data owner.

D. System Architecture and Algorithms

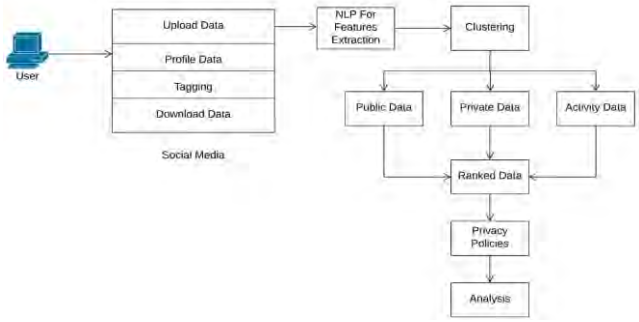


Figure 1: Proposed System Architecture

- System introduced PrivRank, a customizable and continuous privacy-preserving social media data publishing framework.
- System continuously protects user-specified data against inference attacks by releasing obfuscated user activity data.
- System also provide customized protection, the optimal data obfuscation is learned such that the privacy leakage of user-specified private data is minimized.

Algorithm

- *Document retrieval Algorithm*

Input: Users query as Q , Network Connection N;
Output: result from relevancy calculation top k pages base on Q.
Step 1: User provide the Q to system.
Step 2: if (N!=Null)
Process
Read each attribute A from ith Row in incoming document D
Res[i]=Calcsim(Q,A)
Else No connection
Step 3: For each(k to Res)
Step 4: Arraylist Objarray to bind Q to Res[i] or k
Step 5: Return to users Objarray
Step 6: Display Objarray

E. Weight Calculation Algorithm

Input: Query generated from user Q, each retrieved list L from webpage.
Output: Each list with weight.

Here system have to find similarity of two lists: $\bar{a}=(a_1, a_2, a_3, \ldots)$ and $\bar{b}=(b_1, b_2, b_3, \ldots)$ where a_n and b_n are the components of the vector (features of the document, or values for each word of the comment) and the n is the dimension of the vectors:
Step 1: Read each row R from Data List L
Step 2: for each (Column c from R)
Step 3: Apply formula (1) on c and Q
Step 4: Score=Calc(c,Q)
Step 5: calculate relevancy score for attribute list.
Step 6: assign each Row to current weight
Step 7: Categorize all instances
Step 8: end for end procedure

F. Mathematical Model

The system has classified into the different sets like below:

Sys={inp, process, out, analysis}
Inp= {Q1,Q2……Qn}
That is the set of input queries
List={L1,L2,L3..Ln}

$$LD[w] = \sum_{k=0}^{n} \binom{n}{k[D]} \quad (1)$$

G. Extracted list from each documents

L={Wi1, Wi2, Wi3….. Win} weight of each list using below formula

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \quad (2)$$

C={c1,c2….cn} clusters of each list

[C1…Ck]= The finally system work with item or facet rank it can create the set of higher dimensions.

UrlList={URL1(w), URL2(w)…..URLn(w)}

$$UL[k] = \sum_{n=1}^{m} Doc1 + Doc2 \ldots\ldots\ldots Docm \quad (3)$$

Success condition
If(inp != Null)
Failure condition
if(UrlList == Null)

IV. RESULTS AND DISCUSSIONS

We focus on ranking stage that considered as the fundamental commitment of this paper. New positioning calculation is delivered to rank comparative significant information in the wake of ordering stage. Notwithstanding, information recovery process turn out to be quicker, less demanding and more exact. The execution accomplished with 99 percent applicable outcomes in greatest time 60 ms and 1 percent just for immaterial outcomes. The proposed structure and positioning calculation can be additionally produced for later use in distinguishing more exact semantic data from interpersonal organizations in a brief timeframe. The point of the semantic web index has pulled in substantial interests both from industry and research with coming about assortment arrangements in various errands. There is no institutionalized structure that

screens and fortify the advance in this field. In this paper, Four standard undertakings of semantic web crawler are talked about including creeping, ordering, positioning lastly recovering assignment. This survey paper demonstrates that utilizing PrivRank produces an altogether different positioned rundown of news themes, which may imply that depending just on high-recurrence news points gave by the media does not really give understanding into what clients are intrigued on or think about essential. Contemplating all outcomes underlines the point that MF alone is a substandard estimator of what clients find fascinating or think about imperative, and ought to thusly not be utilized as a part of along these lines. PrivRank, then again, turns out to be more equipped for playing out this, thus we infer that the data gave by PrivRank can demonstrate imperative in business based territories where the enthusiasm of clients is paramount.

In experiment 1 system have collected some existing approaches time complexity with proposed. The time required for data extraction for different search engines with this system The below table 11.1 shows the time complexity comparison of keyword base object search, Concept Hierarchy Construction by Combining Spectral Clustering , semantic search using RDF data etc.

Table 1: Comparative analysis between proposed vs Existing approaches (Accuracy)

Method	Accuracy in %	False ratio in %
Keyword Base Object Search	86.15	13.85
Hierarchy base spatial clustering	84.16	15.84
Semantic base RDF search	82.77	17.23
Proposed	95.02	4.98

CONCULSION

System proposed an unsupervised technique Automatic privacy preservation on public as well as private data. The privacy function which will filter all user data and apply anonymity base privacy. Automatic Role Base Access Control (RBAC) to private parameters for user data. It also provide ranked data according to user's perspective for each new session.

REFERENCES

[1] Magdalini Eirinaki, Michalis Vazirgiannis Usage-based page rank for web personalization in proceedings of the fifth IEEE international conference on Data mining (ICDM ¨05)

[2] Samriti Gupta, Alka Jindal Contrast of Link based Web Ranking Techniques at IEEE 2015.

[3] Chi Chen at. Al. proposed An Efficient Privacy-Preserving Ranked Keyword Search Method IEEE 2016.

[4] Jun fang,LeiGuo, Calculation of weight of entities in ontologies by using usage based information in ieee proceedings of 2011

[5] Mercy paul Selvan, ChandraSekar ,A. Priya Darshan Survey on web page ranking algorithms in ijca (International Journal of Computer Applications) proceedings of 2012.

[6] Zhicheng Dou and Zhengbao Jiang, proposed Automatically Mining Facets for Queries from Their Search Results in IEEE transaction and data engineering Vol 28, No. 2 Feb 2017.

[7] Ashlesha Gupta, Ashutosh Dixit, AK Sharma Relevant document crawling with usage pattern and domain profile based page ranking in ieee proceedings of 2013 .

[8] Zhao, B., et al., TEXplorer: keyword-based object search and exploration in multidimensional text databases, in Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM). 2011, ACM: Glasgow, Scotland, UK. p. 1709-1718.

[9] Shital C Patil,RRkeole Content and usage based ranking for enhancing search engine delivery 2014 in volume 3 issue 7International journal of Science and Research (IJSR).

[10] Safaa I. Hajeer, Rasha M. Ismail, Nagwa L. Badr, M. F. Taiba An Efficient Hybrid Usage-Based Ranking Model for Information Retrieval Systems and Web Search Engines . inieee proceedings of 2015 6th International Conference on Information and Communication Systems (ICICS).

[11] AzamFeyznia, mohsin Kahanti A link analysis based ranking method for semantic web documents at ieee proceedings of 2013.

Smart 3-Phase Water Pump Controller

Gauri D. Gaikwad, Nikita S.. Labhade,, Shraddha B. Chavan, Balasaheb J.Pawar4

Department of E&TC NDMVP's KBTCOE, Nashik, SPPU, Pune.

gaurigaikwadg221@gmail.com , labhadenikita6@gmail,shraddhachavan1996@gmail.co,pawar.balasaheb@kbtcoe.org

***Abstract* — The Internet of Things, or IoT, refers to billions of physical devices around the world that are now connected to the internet, collecting and sharing data. Now a days Water management is necessary and an important factor for increasing population and industrialization. Monitoring level of fluids in various areas like domestic and commercial industries, agriculture area etc., plays a key role. The pump motor may get damaged due to dry running or unavailability of proper phase sequence. In some cases like pharmaceutical industries, monitoring liquid level becomes necessary task. This issue can be overcome by using IoT based water level controller.**

Key words — IoT, ESP8266, Starter.

I. INTRODUCTION

Now a day's most of the areas like agriculture, domestic and commercial uses water from the wells, tanks and underground water resources for their application. And for this they need water pumps. To start and control this water pumps user required starter. Right now The starters which are used are operated manually, or controlled by GSM, but in some places the network connection is not available so for that purpose IOT module is used. The purpose of this is to help the user to turn on or turn off the motor by using this user is able to operate the motor from anywhere, And there is another drawback in GSM based starter and automatic starter if the water level is decreased below the required level then also water pumps are not turned off automatically this will damage the water pump, And protection for the motor such as in case any of the phases is failed, so to over this drawback we provide the facility to check all phase availability, in which if anyone phase is absent then motor will remain turn off .[1][2][3]

1 .1 MOTIVATION

Water scarcity is one of the major problems facing many cities of the world and wastage during transmission has been identified as a major culprit; this is one of the motivations for this study, to deploy computing techniques in creating a barrier to wastage in order to not only provide more financial gains and energy saving, but also help the environment and water cycle which in turn ensures that we save water for future.

1 . 2 OBJECTIVES

The tanks used in industries are generally spread across larger areas, for this reason remote monitoring and data collection are most successful tools. To prevent damage of pump by checking phase availability.

To control system through web application.

II. ANALYSIS AND DESIGN APPROACH

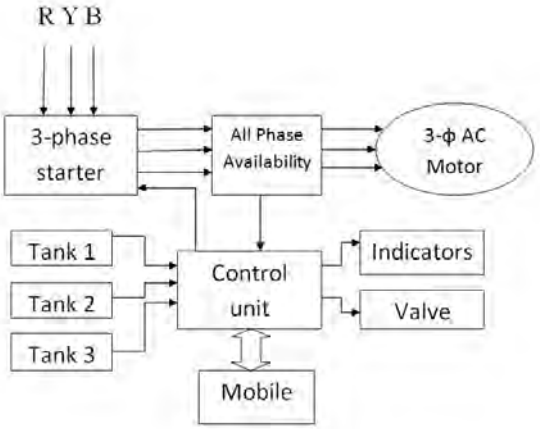


Fig 1: Block Diagram of System

A .Starter:

The primary function of a motor starter is to start and stop the motor to which it is connected. Starter contains overload protection for the motor .Main purpose of the starter is to switch the power automatically or manually to a motor and at the same time protect the motor from overload or faults.

.B. Phase availability:

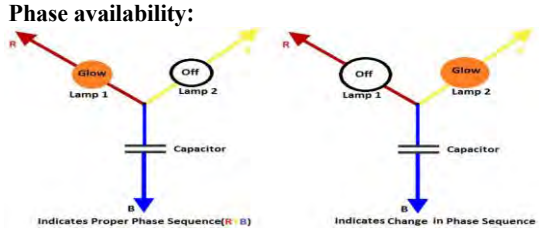


Fig 3: Phase Checking

From the above circuit, with the capacitor, a static-type checker with a capacitor can be obtained, as shown in the above figure, the lamp 1 to R-phase and the lamp 2 to Y-phase are connected. Resistors are connected in series with the lamps for protecting the lamps from over currents and breakdown voltages. From the above circuit, with the capacitor, a static-type checker with a capacitor can be obtained, as shown in the above figure, the lamp 1 to R-phase and the lamp 2 to Y-phase are connected. Resistors are connected in series with the lamps for protecting the lamps

from over currents and breakdown voltages.

C. HC-SR04 Sensor :
The Ultrasonic transmitter transmits an ultrasonic wave, this wave travels in air and when it gets objected by any material it gets reflected back toward the sensor this reflected wave is observed by the Ultrasonic receiver module It measure the distance to the target by measuring the time between the emission and reception.
Distance = Speed × Time
.Operating voltage: +5V.
Theoretical Measuring Distance: 2cm to 450cm.
Practical Measuring Distance: 2cm to 80cm.
Accuracy: 3mm.
Measuring angle covered: <15°.
Operating Current: <15mA.
Operating Frequency: 40Hz

D. ESP8266 :
ESP8266EX (simply referred to as ESP8266) is a system-on-chip (SoC) which integrates a 32-bit Tensilica microcontroller; standard digital peripheral interfaces, antenna switches, RF balun, power amplifier, and low noise receive amplifier, filters and power management modules into a small package. It provides capabilities for 2.4 GHz WI-Fi (802.11 b/g/n, supporting WPA/WPA2), general-purpose input/output (16 GPIO), Inter-Integrated Circuit (I²C), analog-to-digital conversion (10-bit ADC), Serial Peripheral Interface (SPI), I²S interfaces with DMA (sharing pins with GPIO), UART (on dedicated pins, plus a transmit-only UART can be enabled on GPIO2), and pulse-width modulation (PWM). The processor core, called "L106" by Espress if, is based on Tensilica's Diamond Standard 106Micro 32-bit processor controller core and runs at 80 MHz (or over clocked to 160 MHz). It has a 64 KiB boot ROM, 32 KiB instruction RAM, and 80 KiB user data RAM. (Also, 32 KiB instruction cache RAM and 16 KiB ETS system data RAM.) External flash memory can be accessed through SPI. The silicon chip itself is housed within a 5 mm × 5 mm Quad Flat No-Leads package with 33 connection pads — 8 pads along each side and one large thermal/ground pad in the center.

E. Solenoid Valve:
A solenoid valve is an electro mechanical valve for use with liquid or gas controlled by running or stopping an electrical current through a solenoid, which is a coil of wire, thus changing the state of the valve. The operation of a solenoid valve is similar to that of a light switch, but typically controls the flow of air or water, whereas a light switch typically controls the flow of electricity. Solenoid valves may have two or more ports: in the case of a two port valve the flow is switched on or off; in the case of a three port valve, the outflow is switched between the two outlet ports.[7]

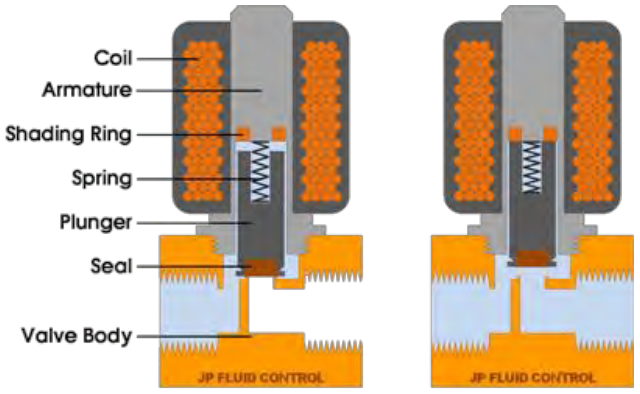


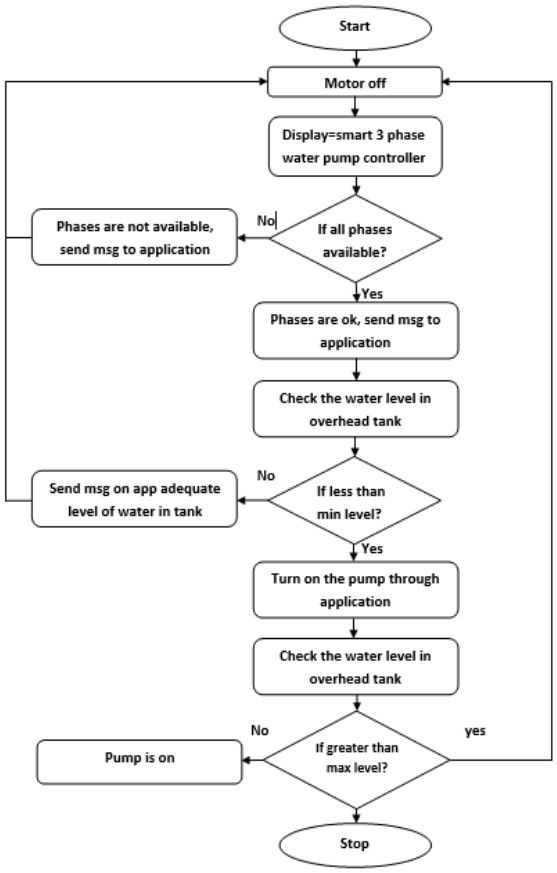
Fig 3: working of solenoid valve

F. Web Application:
ThingSpeak has APIs for collecting data produced by sensors and APIs for reading that data from applications. Think of an IOT project as two parts. One part of the project is where you need to program a thing to send data. And, the second part is where you want to see that data. ThingSpeak sits in the middle and makes it handy to do both.
We will use ThingSpeak in the middle to collect data from sensors and then display the sensor readings on a custom Android app running on a mobile phone. We use the MIT App Inventor to create a custom Android app to see the sensor data and status of the system. This project uses easily accessible hardware to build a proof-of-concept IOT system to monitor level of water in tank or either used for any level measurement applications. surveillance of objects and tracking. GPS uses Trilateration process to compute the position of an object. In our system, GPS is the key component in getting the vehicles coordinates for tracking.[6]

G. Microcontroller :
A microcontroller is a computing device which usually consists of built-in peripherals. It included CPU, USART(Universal synchronous asynchronous Receiver Transmitter), timers, counters and other components. It is a programmable device which takes decision
AT Mega 328P s features :
Flash type Program Memory
32 kb Program Memory
20 CPU Speed (MIPS/DMIPS)
2,048 SRAM Bytes
1024 Data EEPROM/HEF (bytes)
Digital Communication Peripherals
1-UART, 2-SPI, 1-I2C
Capture/Compare/PWM Peripherals
1 Input Capture, 1 CCP, 6PWM
Timers two 8-bit, single 16-bit
1 Comparators
Temperature Range (C) : -40 to 85
Operating Voltage Range (V): 1.8 to 5.5

H. Power Supply :
IC7805 is used to provide supply regulated voltage of 5 Volts to MICROCONTROLLER, and other units in the system.

I. Flow chart



III. CONCLUSION

Many entities and problem solvers have tackled the problem of water wastage in their own way. We have made a serious attempt at solving this problem and saving one of the most precious commodities this world will even have "water". We focused on the problem of water wastage through overflow of water reservoirs or tanks. This project has successfully addressed this problem and will most definitely help control it. The application uses Internet of Things and Android platform to monitor water level changes and also allows user to administer the water tank from different corners of the world.

It has automated the entire process thus reducing energy wastage and manual labor. The centralized architecture of MQTT worked sublimely and proved to be useful in carrying out most of the functionalities of the system. Android is one of the most widely used platforms for deploying applications. Various features of arduino and its ability to utilize sensor data and support for development environment proved beneficial. The system is going to help home owners and organizations to cut back economic losses and provide a reliable way to manage water reservoirs without any hassle. Future developments can improve the system and make it more scalable.

REFERENCES

[1]. Divya Kaur,"IOT based Water Tank Control "[Article-Embedded for You] Jan/Feb 2016.

[2]. N Vijaykumar ,R Ramyas, "The real time monitoring of water quality in IOT environment",IEEE sponsored 2nd international conference on innovations in information, embedded and communication systems (Iciiecs)2015

[3]. IoT based Water Monitoring System:International Journal of Advance Engineering and Research Development (IJAERD) Volume 4, Issue 6, June-2017, e-ISSN: 2348 - 4470, print-ISSN: 2348-6406. [4] International IOT based water level and quality monitoring system in overhead tanks International Journal of Engineering & Technology Conference on Information Processing (ICIP) 2015.

[4]. Asaad Ahmed mohammedahmed ELTAIEB,Zhang Jianmin"Automatic Water Level Control System", international journal of scaince and research (IJSR) 2013.

[5]. Mohd Nor Azni ; A H Zianal ; F A Mohhamad ; N.N. Mohhamad "Home Automation System With Android Application"

[6]. In 2013, Raghavendhra. R,M.Uttara Kumari, S.A. Hariprasad presented a paper on "implemenationof Simulated Water Level Cntroller", International Journal of Advanced Reserch in Computer Science and Software Enginnering.

[7]. Chengtao Xu ; Ying Zhou; Yu Liang "Designing And Research Of A Solenoid Valve Operating Life Test System".

[8]. Mario Essert ; Tihomir Zilie "High Speed Solenoid Valves In Pneumatic Servo Applications".

[9]. Jaytibhatt, jigneshpatoliya,IOT based water quality monitoring system, proceecings of 49th IRF international Conference , 21st February 2016 ,pune, india , ISBN: 978-93-85973-46-8.)". [11]ESP8266 Daasheet

Implementing secure system to visualize business report using data analytics.

Abhishek Anand, Aakash Patel, Abhijeet Saha

Department of Computer Engineering and Technology, Thakur College of Engineering and Technology, Mumbai, India

abhishek.nnd29@gmail.com, patel.aakash19008@gmail.com, abhijeetkrsaha@gmail.com

***Abstract*--This project is an android application which will provide a secure system or platform for the user to generate and visualize business report. There is also security for online transaction in the application. In this project we are using different cryptography hashing algorithm for security of the login, database as well as the transactions and reports which are generated by the user. The users of this application are the CEO, Manager and specific Staff members of the company. Technologies used for front-end and back-end development are Android Studio Java, SQLite and bootstrap. For database security we are using SQL-Cipher. The whole application will be developed on android studio. Purpose of this project is to avail the requirement of the client and prove a secure system for the report generation**.

Keywords--Cryptography, encryption, cascading.

I. INTRODUCTION:

Business's area unit responding to the growing sophistication and range of knowledge of security threats by deploying tools that reach the capabilities of their current security infrastructures. For smaller corporations, this implies deploying deeper network defenses and end point protections. For giant and midsize enterprises, however, it suggests that deploying security analysis tools and analytics software system to gather, filter, integrate and link numerous kinds of security event info so as to achieve an additional comprehensive read of the safety of their infrastructure. Security analytics tools are not to interchange existing security controls and applications; however, they rather complement them. In fact, security analytics tools analyze log and event knowledge from applications, back end controls and network defenses.

The 2013 information Breach Investigations Report from Verizon found that eighty-four of victorious attacks on those infrastructures compromised their targets among hours, whereas seventy-four of attacks weren't discovered for weeks -- and generally months or years. One in every of the explanations it's thus difficult to observe attacks is that they happen quickly. Additionally, information indicating Associate in nursing attack is commonly spread across network devices, servers, application logs and endpoints. This makes it tough to investigate a breach ongoing and even hinders the power to assess its impact. What is more, in line with a Ponenom Institute report, fifty-five of survey respondents that practiced informational knowledge information} loss couldn't determine surely what data was purloined. Up the speed of detection Associate in Nursing analyzing the impact of an attack area unit key driver to adopting security analysis and analytics.

II. LITERATURE SURVEY:

Automated generation of reports from information bases and different data sources like XML files has been a task set to code developers ever since info systems were initial introduced. There square measure varied ways in which of achieving this goal, with variable simplicity of use or flexibility of the results.

One in all the foremost powerful info platforms - Oracle includes a tool named Oracle Reports. The layout models enclosed permit the creation of reports for each paginated output, like printing, still as Web-oriented output. Oracle Reports consists of 2 elements - Reports Builder and Reports Services. Varied information sources will be used - relative databases, text files, XML or OLAP. The tool contains an application program reports editor and templates to style the report. There's conjointly the flexibleness to outline a report layout by putting the fields anyplace in a very page.

Microsoft conjointly includes a less versatile coverage tools in their info platform –MS SQL Server. MSSQL coverage services and SQL Server Business Intelligence development studio embrace options like support for multiple information sources, dynamic end-user sorting, cascading, and multivalve parameters. There's conjointly the tool named Report Builder that is associate ad-hoc coverage tool that permits business users to make their own reports and explore company information. Its question model purportedly let's finish users build reports while not a deep technical understanding of the underlying information sources. A password constitutes a user-memorable secret that consists of some printable characters.

Passwords area unit the common selection in laptop systems for user authentication. The system stores the user's identity and parole for each account. Then, the user logins this knowledge to attest himself and access the service.

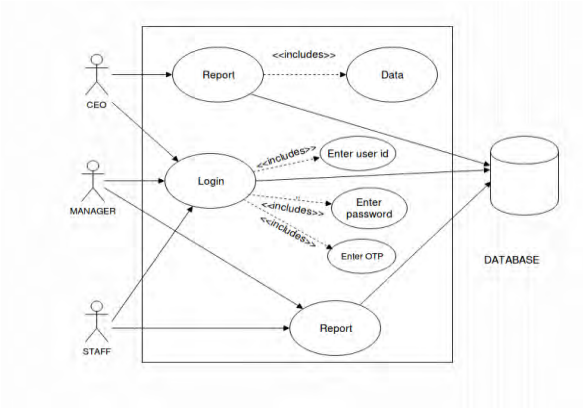
Another utilization of passwords is that the generation of cryptographical keys. The Key-Derivation Functions (KDF) derives one or a lot of cryptographical keys that area unit supported associate degree input parole. Cryptographic keys area unit applied by encryption/decryption functions and KDFs area unit primarily utilized for the coding of session knowledge.

The ordinary choice for user-login services area unit user-originated secrets of eight characters long (8 bytes supported American Standard Code for Information Interchange encoding). These passwords typically exhibit low entropy and area unit prone to attacks. The exhaustive-search technique tests all the various character mixtures and tries to discover the particular parole for associate degree examined account. The offender, then, accesses the account, similar with its legitimate user. The user-

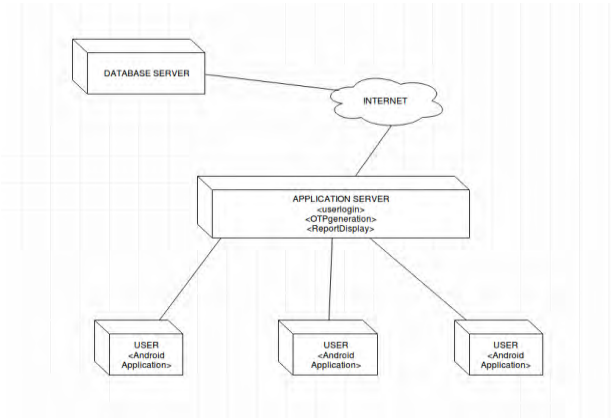
drawn graphical passwords may additionally suffer from low entropy, as they supply 4–5 bytes of security on the average.

The ordinary option to offer protection against such attacks is vital stretching. Cryptographical hash functions represent a security primitive sort that parses input of capricious size and outcomes a digest of fastened length. Within the password-hashing domain, the parole is parsed to a hash function which ends associate degree output of fastened length that then acts because of the password.

A. Use case Diagram:



B. Client Server Architecture:



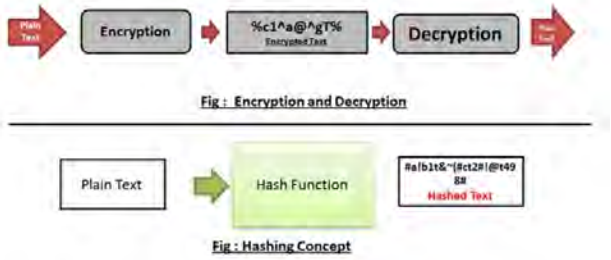
III. ALGORITHM:

For the encryption we are using PBKDF2 algorithm, naturally in programming developers search for the fastest algorithm. But when it comes to hashing, the algorithms have been made slow intentionally to withstand brute force attacks. The super computing resources available today can attempt billions of hashes per second. If the algorithm is slow, the frequency of those attempts can be made slower. The trick is to make the algorithm slower such that it can withstand brute force attacks and still be fast enough not to harm the user experience.

MD5 was once a popular hashing algorithm that was used by many applications. But later many weaknesses were found in MD5. There were collisions where same hash was generated by different inputs. Due to these weaknesses the author himself declared that MD5 is unsuitable for hashing in 2005. Secure Hashing Algorithm (SHA) is another popular hashing algorithm.

There are several versions such as SHA-0, SHA-1, SHA-2 and SHA-3. Among them SHA-0 and SHA-1 are no longer considered as secure. When Linkedin got compromised n 2012 most of their passwords were stolen because, at that time they were still using SHA-1 to hash their passwords. Even though SHA-0 and SHA-1 were unsuccessful, SHA-2 still has some strength left with its different variations of hash lengths. SHA

Family has also released SHA-3 improving its predecessors. If you are a java developer, PBKDF2 is an excellent algorithm to use. In PBKDF2 we can force the algorithm to behave slowly by increasing its iteration count.



System will have 3 options as

1. Input of user name and password.
2. Encryption of filled credentials.
3. Authentication and verification of user.

Every option has its functionality as shown:

A. User name and password will be taken as input.
B. Encryption of the password will be done using PBKDF2.
C. Verification of the user will be done by sending OTP to their number.

IV. CONCLUSION:

The internet has become a major platform in modern business, thus the security on personal as well as professional data has become a one of the most dangerous threat. As per the survey for any kind of business we require a secure platform for our data. Our application provides a secure system for generation and visualizing of business report. We have designed our project to provide the users with a friendly and secure environment.

REFERENCES:

[1] "Smartphone Forensics: Cracking BlackBerry Backup Passwords". Advanced Password Cracking – Insight (ElcomSoft). Retrieved 2015-10-23.

[2] "LastPass Security Notification". The LastPass Blog. Retrieved 2015-10-23.

[3] K. Moriarty; et al. "RFC 8018 - PKCS #5: Password-Based Cryptography Specification, Version 2.1". Tools.ietf.org. Retrieved 2018-01-24.

[4] https://mathiasbynens.be/notes/pbkdf2-hmac

[5] https://crypto.stackexchange.com/questions/26510/why-is-hmac-sha1-still-considered-se

Development of Next Generation Encryption Algorithm for Cloud Security

Siddhant Deshmukh, Aviral Gupta, Sagar Gupta, Viaks Kaul

Department of Information Technology Thakur College of Engineering and Technology, Mumbai India

Abstract: **Cloud Technologies is one of the most emerging technologies in today's world with its applications in several domains. Cloud computing is like a pool of resources which can be used by its users via the internet. Cloud can be used for Storage, Computing, Database Services, Software services and so on. Due to its vast applications, it has become crucial to maintain the confidentiality and integrity of the data that is used in such technologies. Furthermore, the consumption of resources need to be kept minimal while implementing security mechanisms in order to provide better performance when using Cloud Computing systems.**
This project focuses on modifying the AES algorithm in order to provide better resource utilization of the encryption algorithm used in the cloud and improves the protection of resources from unauthorized access or modifications. It further analyzes the proposed system and compares it with the standards in use today.

I. INTRODUCTION

Security is one of the most important issues concerning storage of date on Cloud Systems. In order to maintain the confidentiality and integrity of the resources of the authorized user, it is vital to provide some encryption algorithm to the resources to protect them from attacks against any unauthorized users or attackers. However, it is equally important to analyze the computing resources and requirements of the encryption system in order to enable efficient working environment of the cloud users. AES encryption algorithm is considered to be one of the most secure algorithms which is difficult to crack using traditional attacks such as brute force attack or dictionary attacks. In this project, we compare the existing AES algorithm and modify it in order to enhance the implementation of the same.

The main parameters to determine how efficient an algorithm are the time taken to encrypt the resource from plain text to cipher text, also known as Encryption Time, and to decrypt the encrypted resource from cipher text to plain text, also known as Decryption Time. Other parameters include CPU consumption, Memory consumption, key-size and Avalanche Effect.

The system can be incorporated in various Cloud Systems in order to enhance the efficiency of traditionally used encryption algorithms. It can also be used to offer better alternatives for encryption algorithms in File Transfer Protocols and improved security of communication in local networks.

II. MATERIAL AND METHODS

Materials:

The various technologies used are:

Python: Python is an object oriented programming (OOP) language, similar to PERL that has gained popularity because of its clear syntax and readability. Python is relatively easy to learn and is portable, meaning its statements can be interpreted in a number of operating systems, including UNIX-based systems, Mac OS, MS-DOS, OS/2, and various versions of Microsoft Windows.

A notable feature of Python is its indenting of source statements to make the code easier to read. Python offers dynamic data type, ready-made class, and interfaces to many system calls and libraries. It can be extended, using the C or C++ language.

Eclipse IDE: Eclipse is a free, Java based development platform that is known for its plugins that allow developers to develop and test code written in other programming languages. Eclipse IDE is released under the terms of the Eclipse Public License.

Eclipse got its start in 2001 when IBM donated three million lines of code from its Java tools to develop an open source integrated development environment (IDE). The IDE was initially overseen by a consortium of software vendors seeking to create and foster a new community that would complement Apache's open source community.

CloudSim: CloudSim provides an extensible simulation framework that enables effortless modelling and simulation of app performance. By using CloudSim, developers can focus on specific systems design issues that they want to investigate, without getting concerned about details related to cloud-based infrastructures and services.

Cryptography: Cryptography is the study and practice of various techniques in order to make communication secure in the presence of third parties. It is extensively used in various technologies in order to provide security to various resources. It makes use of encryption algorithms in order to convert a simple plain text message into something that is incomprehensible (cipher text) using a key without which the original message cannot be deciphered.

III. PROBLEM DEFINITION

The use of Cloud Technologies is increasing at a tremendous rate and with the increased use, there is also an increase in the concern of privacy of data that is stored on such technologies.

Therefore, the system should provide security using cryptography techniques.
Also, with the advent of technologies, it is important to consider the reduction in time taken for computation of the encryption technologies. Hence, the system should have better Encryption and Decryption time and should also consume minimal computing resources like memory and CPU.

IV. METHODOLOGY

Extreme Programming-

1. **Planning**: Applying Extreme Programming Methodology for development of encryption algorithm for better usage of resources and time with great performance.

2. **Designing**: Integration of the proposed code and the cloud environment.

3. **Coding**: Coding of AES Encryption Algorithm and developing it further into an enhanced form.

4. **Testing**: Using evaluation software to determine the efficiency of the proposed system.

5. **Deployment & Feedback**: Deploying the proposed system to users and modifying code as per the user's requirement/feedback.

V. WORKFLOW

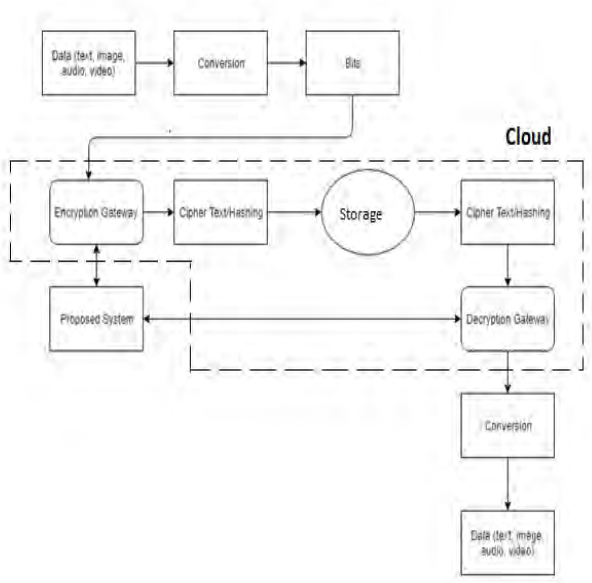


Fig. 1

Steps to be followed:

1. Data in different formats (text, images, audio etc.) is first converted to bits so that encryption can take place.

2. The bits are then converted into blocks of 128 bits and undergo AES encryption using the secret key set by the user at the Encryption Gateway.

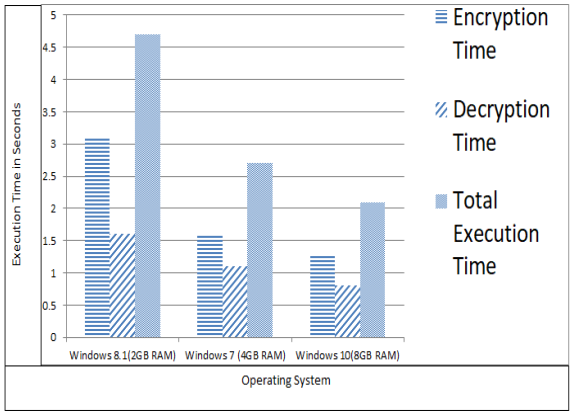
3. The Cipher text obtained from the Encryption process is then transmitted to the cloud over a network and stored.
4. The user then downloads the encrypted data from the cloud and using the secret key, decrypts the data back to the original form.

VI. RESULT

Table I. Result Table

Machine Used	**RAM Size**	**Program Memory**	**Execution Time**
Windows 8.1 (64 bit)	2GB	35000 Kilobytes	4.7 seconds
Windows 7 (64 bit)	4GB	31000 Kilobytes	2.7 seconds
Windows 10 (64 bit)	8GB	35000 Kilobytes	2.1 seconds

Result Histogram



VII. DISCUSSION

In the future, this project can be extended to give better security in various applications of Ubiquitous Computing, which make use of Cloud Technologies for storage and processing of data. It can also be extended in IoT applications like Smart homes, Connected Health etc. for protecting user data from unauthorized access and modification.

VIII. CONCLUSION

With reference to reviewed work we were able to find and define the problem definition we were able to find and define problem definition. We are clear with the idea of problem definition, algorithms, tools (front end, back end) required to implement the same.
We will try our level best to implement the proposed system to achieve the defined objectives.

IX. FUTURE SCOPE

In the future, this project can be extended to give better security in various applications of Ubiquitous Computing, which make use of Cloud Technologies for storage and processing of data. It can also be extended in IoT applications like Smart homes, Connected Health etc. for protecting user data from unauthorized access and modification.

X. ACKNOWLEDGMENTS

We sincerely thank our principal, Dr. B. K. Mishra and HOD, Mr. Rajesh Bansode for always encouraging us to do our best. We are highly indebted to our guide, Mr. Vikas Kaul who supported and constantly supervised us throughout this project and helped us in not only completing this project but also provided us with ample amount of knowledge that was really beneficial to us.

We are thankful and fortunate enough to get constant encouragement, support and guidance from all teaching staff of IT department who helped us in successfully completing the project. Also we would like to extend our sincere gratitude to all staff laboratories for their timely support.

REFERENCES

[1]. Sudhansu Ranjan, Lenka,Biswaranjan Nayak," Enhancing Data Security in Cloud Computing using RSA Encryption and MD5 Algorithm", *International Journal of Computer Science Trends and Technology (IJCST)* – Volume 2 Issue 3, June-2014.

[2]. Prof. Vishwanath S. Mahalle, "Implementing RSA encryption algorithm to enhance the data security of cloud in cloud computing", *International journal of pure & applied research in engineering and technology*, 2013, volume 1(8):220-227, ISSN-2319-507X IJPRET.

[3]. Kartit, Zaid & Azougaghe, Ali & Kamal Idrissi, H & El marraki, Mohamed & Hedabou, M & Belkasmi, Mostafa & Ali, Kartit. (2016). "Applying Encryption Algorithm for Data Security in Cloud Storage"141-154. 10.1007/978-981-287-990-5_12.

[4]. Dimpi Rani, Rajiv Kumar Ranjan, "Enhance data security of private cloud using encryption scheme with RBAC", *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 3, Issue 6, June 2014

[5]. Mrs.Mamatha, Mr.Pradeep Kanchan, "Use of Digital Signature with Diffie Hellman Key Exchange and Hybrid Cryptographic algorithm to Enhance Data Security in Cloud Computing", *International Journal of Scientific and Research Publications*, Volume 5, Issue 6, June 2015

Study of Tubestile

Akash J. Joshi,Shekhar V. Lokare,Urjita R. Jagtap,Dipali V. Patil

NDMVP's KBT CoE ,E&TC Department,SPPU, Pune

itsmeaj137223@gmail.com,shekharlokare16@gmail.com,urjitajagtap27@gmail.com,patil.dipali@kbtcoe.org

***ABSTRACT* — The aim of this project is to provide the door access control and security using tubestile. The proposed system may be implemented by using biometric scanner, password and security question. Remote operating of door access can be done through control rooms. This system can be used in any applications, where security is the critical issue.**

The proposal system can achieve the more safety over RFID System with the 2-step authentication procedures for the enhancement about the security of general RFID systems[2]. After authentication, the proposal system extract the characteristic information of user for acquisition of the additional authentication information. In this project, the system which was proposed more enforce the security of the automatic entrance and exit authentication system. In the result of this project, the system can acquire more exact results as compared with the single authentication system when it uses traditional approach.

Tube stile is ruggedly constructed and has elegant look[4]. It is automatic motorized full height tubestile with two doors controlled each by individual DC motor. It has two stage security checks giving maximum protection. It offers the highest security for environment susceptible to theft, sabotage[5].

***Keywords*—Biometric scanner, RFID, Password, Security questions.**

I. INTRODUCTION

Tubestile is microprocessor based and works on defined parameter. It has LCD display to indicate parameters & operation status and keypad to set required parameters. It has IR sensors and weight measuring device inside the booth to detect presence of user and to avoid multiple people in one transaction. It is built to work with electronic card reader system. The Card reader system can be incorporated to give signals to the tubestile. This System accepts a variety of input release signals to allow entry to restricted areas. For control and security it provides the physical control necessary to compliment manned entry positions. It can be made to be allowing control of entry, or exit or both. The tubestile can be arranged to open from a variety of local or remote input signals according to the user's requirements. The whole system ensures long and trouble free operating life with minimum of the maintenance.

A mantrap i.e. tubestile is essentially just that: a small cylindrical booth designed to "trap" those who would enter a secure area of a facility. This "trap" enables security—whether a system or an employee—to verify the credentials of the entrant and either allow access or trigger alerts indicating an unauthorized entry attempt. Tubestiles, which are sometimes called security vestibules, are small rooms with two or more doors. Authentication procedures may be required either at just the door to the secure area or at all doors.

1.1 .MOTIVATION

1. Physical site security is equally important as that of cyber security.
2.The traditional security system available in market do not provide enough security to their owners. They have only one stage security check..Traditional locks are well known to the thieves and they canbe easily broken by them.
3.These issues are motivation to design the system.

1.2 . OBJECTIVES

1.The main objective behind the project of Tubestile is to provide a safe and secure campus that will create and maintain a feeling of safety within the employees working in an organization.
2.There should be safe environment for important assets of the industries such as important papers and documents or other machinaries.
3.Decrease the amount of actual criminal activity.
4.To give solution for problem of identity theft.

II. LITERATURE REVIEW

Ref No.	Year	Name of Researcher	Name of Paper	Explanation
1	2014	Nurbek Saparkhojayev Department of Information Technology Almaty Management University Almaty, Kazakhstan	Door Access Security System Using Nfc Technology	This project focuses on developing a Near Field Communication (NFC) Smart Lock System that utilizes a smart phones on-board NFC chip as a method to unlock a door. Access is regulated using a NFC Smart Card which runs a simple user interface that allows user to control the access at home or offices.

2	2015	Md. Kishwar Shafin1 , Kazi Lutful Kabir2 , Nazmul Hasan3 , Israt Jahan Mouri4 , Samina Tasnia Islam5 , Lazima Ansari6	RFID Based Access Control System	utilizes the electromagnetic fields for data transfer in order to perform automatic detection and tracking of tags or tags of objects for security aspects. Its a digital access control system that can be employed to a protected area where none but people with authenticated credentials can enter. The system comprises of digital door lock which is unlock able in real time to ensure secured access by bringing the RFID card close to the reader
3	2016	G.Sowjanya M.Tech, Embedded systems Vignan's Lara institute of technology and science, Guntur, A.P, India 1	Design and impleme ntation of door access control and security system based on IOT	Use of innovative technologies will improve the security to a great extent till the technology becomes open to all. Intends to provide the door access control and security by using IoT Server. Its implemented by using biometric scanner, password and security question with IoT. Remote operating of door access can be done by IoT with a smart indication.

3. BLOCK DIAGRAM

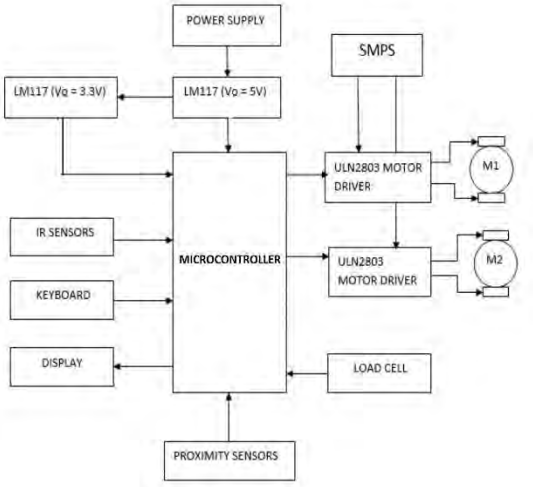


Fig 1: Block Diagram of System

1.3 3.1 PROXIMITY SENSOR

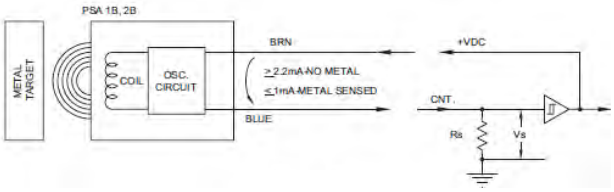


Fig 2: Proximity sensor

Proximity Sensor is used to detect objects and obstacles in front of sensor.Sensor keeps transmitting electromagnetic field and when any object comes near, it is detected by the sensor by monitoring the changes in transmitted electromagnetic waves. It can be used in robots for obstacle avoidance, for automatic doors.

RELAYS

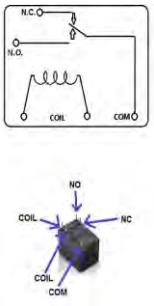


Fig 3: Relay contacts

The Single Pole Double Throw SPDT relay is quite useful in certain applications because of its internal configuration. It has one common terminal and 2 contacts in 2 different configurations, one can be Normally Closed and the other one is opened or it can be Normally Open and the other one closed. So basically you can see the SPDT relay as a way of switching between 2 circuits, when there is no voltage applied to the coil

one circuit receives current, the other one doesn't and when the coil gets energized the opposite is happening.

1.4 3.2 IR SENSOR

An infrared sensor is an electronic device, that emits light in order to sense some aspects of the surroundings. An IR sensor can measure the heat of an object as well as detects the motion. These types of sensors measures only infrared radiation, rather than emitting it that is called as a passive IR sensor. Usually in the infrared spectrum, all the objects radiate some form of thermal radiations. These types of radiations are invisible to our eyes, that can be detected by an infrared sensor. The emitter is simply an IR LED (Light Emitting Diode) and the detector is simply an IR photodiode which is sensitive to IR light of the same wavelength as that emitted by the IR LED. When IR light falls on the photodiode, The resistances and these output voltages, change in proportion to the magnitude of the IR light received.

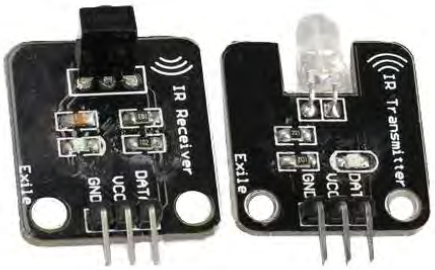


Fig 4: IR Sensor Module

3.3 BUZZER



Fig 5: Buzzer

A buzzer is in the mechanical form of a small rectangular or cylindrical housing, with electrical connection for direct mounting on rigid printed circuit board. In the latter case, the buzzer has two small brackets. The loudness of such a component is about 85 dB / cm. It requires a DC voltage to operate, it should generally in between 3 V and 28 V, depending on the model. A buzzer designed to operate at 6 V generally works very well for any supply voltage between 4 V and 8 V, and a buzzer designed to operate at 12 V can work perfectly at a voltage between 6 V and 28 V. There are also buzzers that work directly on the AC mains 230 V. This type of buzzer is convenient to use, because unlike piezoelectric buzzers (simple piezoelectric transducers without associated electronics), it has no work, except of course the eventualy control stage which will enable it.

1.5 3.4 LOAD CELL

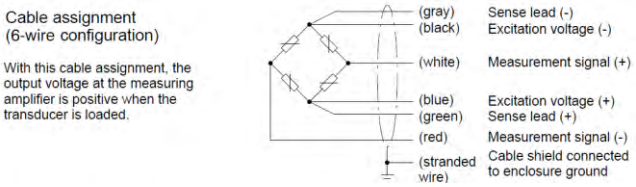


Fig 6: Load cell

Load cell is the combination of four strain guages. They are arranged in the wheat stones bridge manner. Due to this, the sensitivity of the load cell is greater than single strain guage. Hence, we have used the load cell instaed of the strain guage.

1.6 3.5 BLDC MOTOR



Fig 7:BLDC motor

Brushless DC electric motor (BLDC motors, BL motors) also known as electronically commutated motors (ECMs, EC motors), or synchronous DC motors, are synchronous motors powered by DC electricity via an inverter or switching power supply which produces an AC electric current to drive each phase of the motor via a closed loop controller. The controller provides pulses of current to the motor windings that control the speed and torque of the motor.

1.7 3.6 TOUCH SCREEN CONTROLLER

The XPT2046 is a 4-wire resistive touch screen controller that incorporates a 12-bit 125 kHz sampling SAR type A/D converter. The XPT2046 operates down to 2.2V supply voltage and supports digital I/O interface voltage from 1.5V to VCC in order to connect low voltage microprocessor. The XPT2046 can detect the pressed screen location by performing two A/D conversions. In addition to location, the XPT2046 also measures touch screen pressure.On-chip VREF can be utilized for analog auxiliary input, temperature measurement and battery monitoring with the ability to measure voltage from 0V to 5V. The XPT2046 also has an on-chip temperature sensor The XPT2046 is available in 16pin QFN thin package(0.75mm in height) and has the operating temperature range of -40 degree C to +85 degree C.

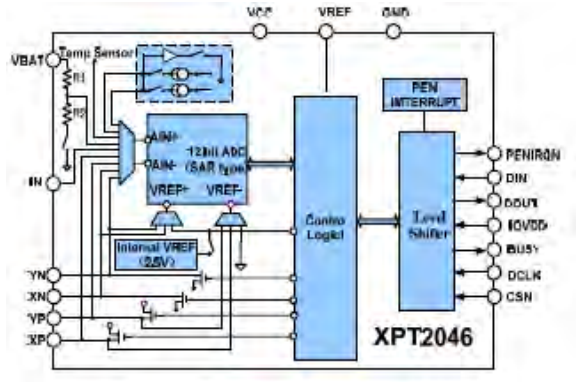


Fig 8: Touch Screen Controller

3.7 LCD MODULE

LCD module i.e. Liquid Crystal Display Module is used along with touch screen controller to use screen as touchpad instead of normal keyboard. 4.3-inch TFT Color LCD Module is a TFT LCD Screen Module , 40pins interface , not just a LCD break but include the Touch , SD card and Flash design. So it's a powerful extension module for your project. The LCD is LB04301. Its a 16M Shaded Display , 480*272 (resolution), 4.3 inch Widescreen TFT LCD screen.The LCD has a wide viewing angle , the contrast is also very suitable. The Screen include a controller SSD1963, its a support 16bit data interface easy to drive by many MCU like STM32 ,AVR and 8051. TFT 4.3" 480*272 With SD Touch Module is designed with a touch controller in it [9]. The touch IC is XPT2046 and touch interface is included in the 40 pins breakout.

2.8 POWER SUPPLY

Here initially 230V ,50Hz power supply is given. This is then converted to 5V ,3.3V,12V etc for ADC and DAC,power for microcontroller BLDC motors etc, respectively with the help of low dropout regulator such as LM1117.

ADVANTAGES

1) Used to alert and detect theft instantly
2) Allow only authorised person inside.
3) No scope for cheating.
4) Battery Backup of 2 to 4 hours

APPLICATION

1) Server rooms
2) Banking
3) Sensitive manufacturing facility
4) Credit card manufacturing units
5) Military
6) Sensitive areas
7) Research centres

ACKNOWLEDGMENT

We thankfully acknowledge Sivananda Electronics, deolali camp, Nashik for providing various resources such as laboratory with all needed software platforms for our project.

CONCLUSION

We are working on the project called tubestile which will now over the highest level security. This project will be one of those security systems which will over multilevel security. i.e. Two stage authentication. Hence the documents, valuables or other credentials will have almost zero percent chances of being misuse or get stolen. The chances of identity theft will be zero in this system which will be the best way of avoiding theft or misuse according to survey. Hence this will fulfill the objectives of our project.

REFERENCES

[1]. Nurbek Saparkho-jayev Department of Information Technology Almaty Management University Almaty, Kazakhstan, Aigul Dauitbayeva Department of Computer Engineering and Information Systems Korkyt Ata Kyzylorda State University Kyzylorda, Kazakhstan, Aybek Nurtayev Department of Computer Science and Software Engineering International IT University Almaty, Kazakhstan, Gulnaz Baimenshina Department of Applied Physics K.I. Satpayev Kazakh National Technical University Almaty, Kazakhstan,"Door access security system using NFC technology ",IEEE,2014.

[2]. G.Sowjanya M.Tech, Embedded systems Vignans Lara institute of technology and science, Guntur, A.P, India, S.Nagaraju Assistant professor, Dept of ECE Vignans Lara institute of technology and science, Guntur, A.P, India, "Design and imlimentation of door access control and security system based on IOT",IEEE,2016.

[3]. Md. Kishwar Sha_n1 , Kazi Lutful Kabir2 , Nazmul Hasan3 , Israt Jahan Mouri4 , Samina Tasnia Islam5 , Lazima Ansari6 , Md. Mahboob Karim7 and Md. Afzal Hossain8 1-3, 5-8Department of Computer Science and Engineering, Military Institute of Science and Technology, 4 Department of Computer Science and Engineering, Stamford University Bangladesh, Dhaka, Bangladesh," Development of RFID based access control system", IEEE, 2015.

[4]. http://www.divyainfotechcorp.net/high-security-tubestile-booth-2077759.html

[5]. https://www.indiamart.com/proddetail/high-security-tubestile-mantrap 16271653333.html

[6]. http://www.sivanandaelectro.com/full-height-turnstiles.htmlhigh-security-tubestile-mantrap

[7]. http://www.datacenterjournal.com/what-is-a-mantrap-and-do-you-need-one/

[8]. https://www.electrocraft.com/products/bldc/

[9]. https://www.emartee.com/product/42175/4.3

IoT Based Weather Station

Sunny Sall, Abhishek Yadav, Akash Rai, Sagar More, Ashish Jaiswar
Computer Engineering SLRTCE Mumbai,India
sunny.sall@slrtce.in, ay7666631050@gmail.com, praksh1509@gmail.com, sam26sky@gmail.com, ashishjaiswa457@gmail.com

***Abstract*— A weather station can be described as an instrument or device, which provides us with the information of the weather in our neighbouring environment. For example it can provide us with details about the surrounding temperature, barometric pressure, humidity, etc. Hence, this device basically senses the temperature, pressure, humidity, light intensity, rain value. There are various types of sensors present in the prototype, using which all the aforementioned parameters can be measured. It can be used to monitor the temperature or humidity of a particular room/place. With the help of temperature and humidity we can calculate other data parameters, such as the dew point. In addition to the above mentioned functionalities, we can monitor the light intensity of the place as well. We have also enabled to monitor the atmospheric pressure of the room. We can also monitor the rain value. The brain of the prototype is the ESP8266 based Wi-fi module Nodemcu (12E). Four sensors are connected to the NodeMCU namely temperature and humidity sensor(DHT11), pressure sensor(BMP180), raindrop module, and light dependent resistor(LDR). Whenever these values exceed a chosen threshold limit for each an SMS, an E -mail and a Tweet post is published alerting the owner of the appliance to take necessary measures.**

I. INTRODUCTION

With the advent of high speed Internet, more and more humans around the globe are interconnected. Internet of Things (IoT) takes this a step further, and connects not only humans but electronic devices which can speak amongst themselves [1]. With falling costs of Wifi enabled devices this trend will only gather more momentum. The main concept behind the Internet of Things(IoT) is to connect various electronic devices through a network and then retrieve the data from these devices (sensors) which can be distributed in any fashion, upload them to any cloud service where one can analyze and process the gathered information. In the cloud service one can utilize these data to alert people by various means such as using a buzzer or sending them an e-mail or sending them an SMS etc.

As mentioned earlier, IoT enables not only Human-Human interaction, but also Human-Device interaction as well as Device-Device interaction. This particular development in the shape of new avenues of interactions will impact essentially every industry such as transportation and logistics, energy, healthcare etc. For example, in the case of energy, IoT is being applied to create Smart Grids which can detect and respond to changes in local and broader level changes in energy consumption, which is going to be an integral part of any nations energy policy.

Looking beyond the aforementioned energy example, there are many areas of interests where IoT can make a meaningful impact such as, Smart Homes, which involve IoT to heighten the degree of automation; Wearable technologies such as smartwatches and fitness bands; One of the biggest areas of potential in IoT is connected healthcare.

Many global electronics behemoths have already invested deeply in the Internet of Things infrastructure. With players like Intel, Rockwell Automation, Siemens, Cisco and General Electric the market is on the cusp of an explosion, with analysts predicting there will be 26 Billion connected devices, more than 4 per human on the planet, and the industry is projected to bring in $19 Trillion, in costs savings and profits with firms like Samsung and Google leading the pack.With this new technological platform however, comes its own set of challenges and obstacles, such as what to do with the enormous amounts of data which is collected.

This project as well measures environmental parameters such as temperature, humidity, pressure, light intensity etc. and uploads these values to a cloud service, IBM Bluemix.[2] In the cloud the data are analyzed and if the retrieved datas are above or below a certain threshold limit, depending on the value, an e-mail, an SMS and a twitter post is published at the exact moment[3].

Earlier people staying in home and busy in their household chores or people busy in their offices workload had no idea about the environmental parameters outside their home or office. They have no idea if the temperature outside is quite high or quite low or normal or if it is raining outside or not or what is the value of the humidity in the outside environment. This device can come in quite a handy in these situations. It will notify us whenever the temperature is too low or too high through an e-mail, an SMS and a twitter post. It will also automatically notify whenever there is a downpour in the surrounding and remind us to carry an umbrella or a raincoat [4]. It will also greet us with good morning and good evening messages as it also has an LDR which measures the light intensity of the surrounding environment[5]. The core of the project is the ESP8266 based Nodemcu which is a low cost wifi module and all the other sensors are connected to this device. The C code is written in arduino IDE and uploaded to the ESP8266 through a serial bus. Once the code is uploaded then the board is connected to a Wifi and the device starts working. The code has to be uploaded only once.

II. IMPLEMENTATION SETUP

A. Components required: Hardware

1) ESP8266 based wifi module Nodemcu[6]
2) Temperature and Humidity Sensor(DHT11)[7]
3) Barometric Pressure Sensor(BMP180)[8]
4) LDR[9]
5) Raindrop Module[10]
6) Mobile phone to recieve email and SMS

B. Components required: Software

1) Arduino IDE[11]
2) Accessible Wifi

3) IBM Bluemix[12]

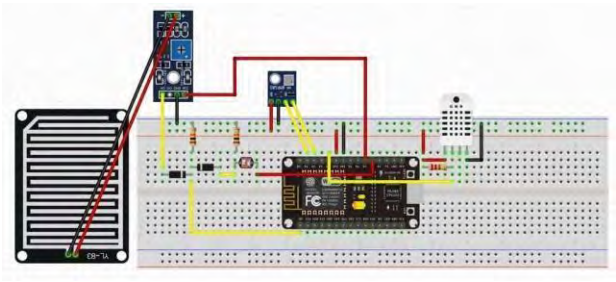


Fig. 1. The complete setup of the device.

III. METHODOLOGY

A. Nodemcu

It is the heart of the device. It provides the platform for IOT. Its a wifi module having esp8266 firmware within. All the other sensors are connected to this micro-controler. They send the measuered values to it and it uploads all the values to the cloud where the values are analyzed. The developer of this board is ESP8266 Opensource Community. It has an operating system called XTOS. The CPU is ESP8266(LX106). It has an in-built memory of 128 KBytes and a storage capacity of 4 Mbytes.



Fig. 2. Nodemcu.

B. DHT-11 (Temperature Sensor)

It senses the temperature of the surrounding. Its a 4-pin device. We should connect a 10k resistor between pin 1 and pin2. Pin 1 is connected to the 3.3V. Pin 4 is connected to GND. Pin 2 is the output pin which gives input to the nodemcu pin D4. Pin 3 is left empty.



Fig. 3.
DHT11.

1) BMP 180(Pressure Sensor): It senses the barometric pressure from the surrounding. BMP180 is an I2C standard device. Its a 4-pin device, viz, SDA, SCL, VIN, GND. Vin and GND are connected to 3.3V and GND respectively. SDA is connected to D2 pin of nodemcu and SCL is connected to D3 pin of nodemcu.

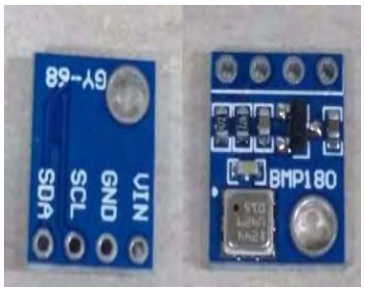


Fig. 4. BMP180.

C. Light Dependent Resistor(LDR)

An LDR is a variable resistor controlled by light. The increasing light intensity falling on it decreases the resistance of the LDR. It has an analog output which is an input to the A0 pin of the nodemcu.



Fig. 5. LDR.

D. Raindrop Module



It is used for the detection of rain. It can also be used for measuring the intensity of the rain. It has both digital output as well as analog output. This module measures the moisture through analog output pin and when the threshold of moisture exceeds too much it provides a digital output. The more water or the lower resistance means lower output voltage. Where as, the less water means higher resistance,i.e, high output voltage on the analog pin. For example a completely dry board will cause the module to output five volts. The analog output of the module is connected to the A0 pin of the nodemcu. Fig. 6. Raindrop Module.

E. Working of the analog pin(A0)

The Nodemcu board has only 1 analog pin, but in this project two analog output devices,viz,LDR and Raindrop Module, are multiplexed to the A0 using two diodes. The multiplexing circuit is shown in the Fig.7 below. Here the Vcc

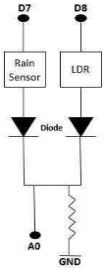


Fig. 7. Multiplexing circuit.

of Raindrop sensor is connected to the D7 of the nodemcu and the input of LDR is connected to the D8 of nodemcu. When D7 is High, D8 is Low making LDR off and raindrop module on. Hence the output of raindrop sensor reaches the A0 of nodemcu through the diode.

Similarly, when D8 is High and D7 is Low, LDR is on and raindrop module is off making a path for the LDR output to reach to the A0 of the nodemcu through the second diode. The 10k resistance is used to reduce the voltage drop across raindrop module and LDR. Hence, we are accomodating 2 analog devices in the nodemcu having just one analog pin.

ALGORITHM 1: Analog Pin Multiplexing

Input : LDR D8, raindrop D7;

Output : Analog pin A0
loop {

if (digitalRead(D7) == HIGH)

{

A0=raindrop sensor value;

Digital Write(D7,LOW);

}

else if (digital Read(D8) == HIGH)

{

A0=LDR value;

digital Write(D8,LOW);

}

}

IV. RESULTS

After the sensor measurements are uploaded to the cloud, IBM Bluemix, the values are analyzed there and then an email, an SMS and a tweeter post is published whenever the threshold limit exceeds. Some of the sample results are as following:



Fig. 8. Email received.

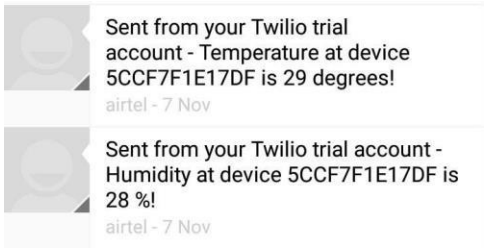


Fig. 9. Message received.



Fig. 10. Tweeter post published.

ALGORITHM 2:E-mail, SMS & Tweeter post

Input: temp, humidity, press, LDR, rain;
Output : email, sms, tweet
loop {

temp=temperature value measured;

humidity=humidity value measured;

press=pressure value measured;

LDR=light intensity measured;

rain=rain value measured;

if (temp && humidity && press && LDR && rain) {

publish all the measured value to the cloud bluemix;
serial.println(temp);

serial.println(humidity);

serial.println(press);

serial.println(LDR);

serial.println(rain);

}

else

{

serial.println("error! check the sensors.");

}

if(temp>=40)

{

email="The current temperature is" + temp;

sms="The current temperature is" + temp;

tweet="The current temperature is" + temp;

```
}
if(humidity>=50)
{
email=”The current humidity is” + humidity;
sms=”The current humidity is” + humidity;
tweet=”The current humidity is” + humidity;
}
if(rain>=200)
{
email=”It’s raining outside. Bring your umbrella”;
sms=”It’s raining outside. Bring your umbrella”;
tweet=”It’s raining outside. Bring your umbrella”;
}
if(LDR>=150)
{
post only once in a day
{
email=”Good Morning”;
sms=”Good Morning”;
tweet=”Good Morning”;
}
}
}
```

V. FUTURE SCOPE

The proposed IoT based weather station can be modified To incorporate many more features.We can add OLED display to display the surrounding parameters into it. We can also add a GPS module in the design so that location of the surrounding also messaged to the user along with temperature, humidity, pressure etc.

VI. REFERENCES

[1] M. H. Asghar, A. Negi, and N. Mohammadzadeh, “Principle application and vision in internet of things (iot),” in *International Conference on Computing, Communication Automation*, May 2015, pp. 427–431.

[2] A. Gheith, R. Rajamony, P. Bohrer, K. Agarwal, M. Kistler, B. L. W. Eagle, C. A. Hambridge, J. B. Carter, and T. Kaplinger, “Ibm bluemix mobile cloud services,” *IBM Journal of Research and Development*, vol. 60, no. 2-3, pp. 7:1–7:12, March 2016.

[3] S. Gangopadhyay and M. K. Mondal, “A wireless framework for environmental monitoring and instant response alert,” in *2016 Interna- tional Conference on Microelectronics, Computing and Communications (MicroCom)*, Jan 2016, pp. 1–6.

[4] H. Saini, A. Thakur, S. Ahuja, N. Sabharwal, and N. Kumar, “Ar-duino based automatic wireless weather station with remote graphical application and alerts,” in *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, Feb 2016, pp. 605–609.

[5] A. Lage and J. C. Correa, “Weather station with cellular communication network,” in *2015 XVI Workshop on Information Processing and Control (RPIC)*, Oct 2015, pp. 1–5.

[6] T. Thaker, “Esp8266 based implementation of wireless sensor network with linux based web-server,” March 2016.

First Person Perspective (FPP) Facade for purchase of Estates

Nuruddin Warsi, Gulamwaris Warsi, Komal Bhatnagar, Hetansh Shah, Anil Chaturvedi

Information Technology Shree LR Tiwari College of Engineering Mumbai, India
nuruddinwarsi89@gmail.com, gwarsi70@gmail.com, komalbhatnagar123@gmail.com, hetanshshah1996@gmail.com, chaturvedi.anilcs03@gmail.com

***Abstract*—Property buying is a sector which there are still some difficulties. When Buying properties, customers have no way of getting any idea on how their future property would look like. Augmented reality AR is a technological field which helps to visualize various real world entities in a combination of real and virtual environment. FPP Facade uses Augmented reality to help potential customers help visualize how their property would look like in a virtual environment. This would help them to make better decisions in their endeavor. This project also helps to provide accurate and real time measurements of the environment.**

Keywords—Augmented Reality, Virtual Reality, Cloud Computing, Voice Controlled

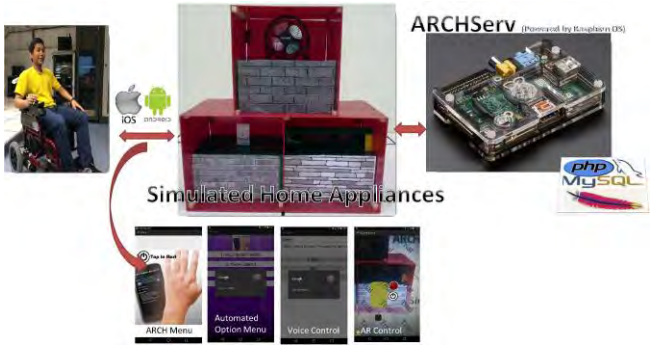
I. INTRODUCTION

With the rapid development of computer science, Human-computer interaction technology has been widely used, besides which the VR and AR technology are part of the important application directions. Virtual Reality(VR) refers that observers immerse straight into the three-dimensional world created by computer, rather than using the traditional human machine interface or windows, VR integrates the three-dimensional computer graphics technology, sensing technology, artificial intelligence technology and so on. Augmented Reality(AR) develops on the base of VR technology, it can add as well as locate virtual object or information by the use of computer graphics and display technology, then place" the virtual objects accurately into the real environment by using the sensing technology, which can fuse the virtual and real objects successfully through some related equipment, so that achieving the information integration of real world with virtual world, bringing a kind of real-time interaction to observers. The main difference between AR and VR is that, Augmented reality (AR) adds digital elements to a live view often by using the camera on a smart phone. Examples of augmented reality experiences include Snapchat lenses and the game Pokémon Go. Virtual reality (VR) implies a complete immersion experience that shuts out the physical world. Meanwhile, the information of all users and the images of the respective models will be stored in cloud database.

II. RELATED WORK

A. Augmented Reality Control Home(ARCH) for elderly and disabled

In this project, authors used Augmented Reality (AR) using Wikitude SDK [5] to allow a virtual object to pop out on the screen of the mobile device when the user points their mobile device's camera to a physical object or an image. This allowed the user to interactively and conveniently control several applications remotely. Besides, AR uses the concept of image tracking, processing and communicating with backend server.



B. Combining Augmented Reality and speech technologies to help deaf and hard of hearing people.

In this paper, authors have used Augmented Reality(AR), Audio to Speech Recognition(ASR) and Text to Speech Synthesis to help people who are deaf. The System takes the narrator Speech and converts it into text and finally displays it on the AR screen. This can help deaf people to understand by looking and understanding through the AR screen. For better accuracy in noisy environment Audio Visual Speech Recognition (ASVR) is used.

Today, using new technologies to help people with disabilities is highly regarded and much research in this area. The important issue is how to combine and use these technologies together to make them more applicable. Augmented Reality AR is a new technology that has a lot of interest recently. Several studies have shown that AR and Virtual Reality VR can be used to help people with disabilities.



III. PROPOSED SYSTEM

A. Cloud Computing

Cloud computing aims at allowing access to large amounts of computing power in a fully virtualized manner, by collecting resources and offering a single system view. It makes computer system resources, especially storage and computing power, available on demand without direct active management by the user. The term is generally used to describe data centres available to many users over the Internet. Cloud providers typically use a "pay-as-you-go" model. Clouds may be limited to a single organization (enterprise clouds,) be available to many organizations (public cloud,) or a combination of both (hybrid cloud). Cloud computing is the use of various services, such as software development platforms, servers, storage and software, over the internet, which is referred as “cloud”. A model is denoted in which a computing infrastructure is viewed as a “cloud” from which individuals and organizations access applications from anywhere in the world on demand

B. Augmented Reality

Augmented reality (AR) is a type of interactive, reality-based display environment that takes the capabilities of computer generated display, sound, text and effects to enhance the user's real-world experience. Augmented reality combines real and computer-based scenes and images to deliver a unified but enhanced view of the world. From social media filters, to surgical procedures, AR is rapidly growing in popularity because it brings elements of the virtual world, into our real world, thus enhancing the things we see, hear, and feel.

Augmented reality is used to enhance natural environments or situations and offer perceptually enriched experiences. With the help of advanced AR technologies (e.g. adding computer vision and object recognition) the information about the surrounding real world of the user becomes interactive.

C. Voice Commands

User will provide command using microphone. The application will use some APIs for converting speech to text. After that, the application will perform the action accordingly.

[1] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

Handover Security Solution for HetNet in Millimeter Range Communication

Devidas Chikhale, Shankar Deosarkar, Mahesh Munde

Electronics and Telecommunication, Lokmanya Tilak College of Engineering Koparkhairane, India
Electronics and Telecommunication, Dr. Babasaheb Ambedkar Technological University Lonere, India
Electronics and Telecommunication, St. Xavier Technical Institute Mahim, Mumbai, India.
devidas.chikhale@gmail.com, sbdeosarkar@yahoo.com, mahesh79@gmail.com

***Abstract*— Heterogeneous network security is a prime concern in the era where billions of human beings, and devices will get connected through next generation technologies like millimeter wave wireless, internet of things and machine to machine communication. Huge bandwidth in millimeter range will help to achieve seamless connectivity, and tens of gigabits per second data rate. In the dense network of small cells handover security has to play vital role to avoid call drops by reducing number of handovers and probability of handoff. This paper basically deals with how handover security can be achieved in femtocell network using UMTS network architecture when user moves from femtocell to macrocell and from macrocell to femtocell. Number of handovers and probability of call drops are less when user moves from femtocell to macrocell. Handoff mechanism is an important aspect in deciding the quality of communication. Novel handover policy reduces the outage probability and increase signal to interference plus noise ratio and power required at the mobile terminal. It directly affects revenue of the operator and call quality. Handoff security is also an important aspect to avoid eavesdropping and spoofing of an information or data due to delayed handoffs. In mobile networks base stations along with mobile switching centers or the mobile station alone handle the task of handoff.**

Keywords— handoff, call admission control, millimeter wave, femtocell, hetnet.

I. INTRODUCTION

When user is mobile concept of handoff plays an important role. When user moves out of coverage area of registered base station and enters into new base station area the call control gets transferred to new base station and the process is called handoff. When user moves from macrocell to small cells and vice versa it is called crosstier handover [9]. Visitor location register keep details of such user and then transfer it to home location register. In mobile network, mobile switching center along with base station and mobile station handles the task of handoff. If handoff will not get completed properly call will gets dropped. Femtocells are gaining popularity these days because of low transmit power, capacity and coverage can be enhanced [4]. Today because of large number of cells in small cell network, mobile switching centers may get burdened so mobile assisted handoffs are used. Handoff security is essential to avoid delayed handoffs. Hetnet is a network containing node with different characteristics such as transmission power and radio frequency coverage area[5]. Physical layer security may help authentication process in complex HetNets in 5G networks [6]. Integration of millimeter wave and massive mimo with HetNet is essential to understand security requirement for next generation heterogeneous network[7]. Protocol usage like 802.11r will reduce handover authentication delay by using handover key [3].Handoffs are basically classified into various types based on link types and network. First is hard handoff which is break before make kind of handoff where existing base station providing service discontinued first and later call gets connected to new base station. Second is soft handoff which is make before break kind of handoff where existing base station gets disconnected after call gets connected to new base station or channel. Third is intersystem type of handoff where call gets transferred from one mobile switching center to another mobile switching center. Fourth is mobile assisted handoff in which call gets transferred from one mobile station to another mobile station without involvement of base station and mobile switching center. Fifth is intra system handoff in which under one mobile switching center call gets transferred from one base station to another base station. In ultra dense networks millimeter wave communication is excellent choice as it is well suited for small distance and line of sight communication. But frequency of handover and handover probability are the important parameters for handover security of heterogeneous network. Call drops results due to handover failure can reduce revenue of operator upto greater extent. This paper basically focuses on this aspect. Proposed handover security approach reduces number of handoffs and probability of handover in dense environment and helps to improve handover security. The paper is divided into three sections. Related work, simulation results and conclusion.

II. RELATED WORK

In small cells network handover takes place from macrocell to femtocell and from femtocell to macrocell. Call handling becomes challenging when user moves from macrocell to femtocell. Handoff security is improved when numbers of handovers are reduced and user moves from macrocell to femtocell. Hetnet security is prime concern to us. This approach gives optimum solution when compared to existing. Unnecessary handoffs remain serious problem in femtocell network. The network can be formed using UMTS technology. It consists of FemtoGateway (FMG) and Femto Access Point (FAP) as main components.

Femto access point consists of femto management system. Handover procedure is basically divided into two phases. Handover preparation phase and execution phase. In handover preparation phase information is gathered and handover decision is taken. Mobile equipment collects information about number of handover candidates to whom authentication is required for security purpose. Best handover candidate is determined and mobile station initiate request to connect to new access point. Radio resources are required in this approach but number of handovers are reduced to get optimized

solution. In femtocell to macrocell handover, time required for handover should be small but it doesn't require complex interference calculation and authorization check in this type of handover. Frequent and unnecessary handoffs reduce quality of service level and also decrease capacity of a system. This problem arises when high speed user moves from macrocell to femtocell and vice-versa. Three parameters have been considered for proposed call admission control (CAC). Those are received signal level, time for which mobile station maintains minimum required signal level, and signal to interference level. Proper threshold is selected for handover decision. It is a minimum signal level required so that handover decision can be taken for movement of mobile station from macrocell to femtocell. Threshold time is decided by the operator. It is playing important role along with interference level in determining handoff decision in proposed call admission control. Larger the threshold time lesser is the number of handoffs. Simulations are done for approaches without call admission control and with call admission control for thresholds ten seconds and twenty seconds respectively. System architecture used is 3GPP UMTS network based macrocell and femtocell integrated network[1]. Femtocells are attracting attentation to increase coverage, capacity, low transmit power and also in the process of handoff [4].To improve coverage, capacity and handoff security frequent and unnecessary handoffs in femtocell network must be reduced. Femtocell measurement of signal to interference ratio parameter is considered for handover between macrocell to femtocell network. Handover optimization scheme is used to validate the performance. Femtocell Gateway (FGW), Femtocell Management System (FMS) and Femtocell Access Point (FAP) acts as a concentrator to core network connectivity. Traffic flows from femtocell access point to femtocell gateway to radio network controller to target femtocell. Femtocell access point has preregistered users. It reduces number of target femtocell for macrocell to femtocell handover. Communication signals among different femtocell access point are optimised. It is one of the main issues in the deployment.

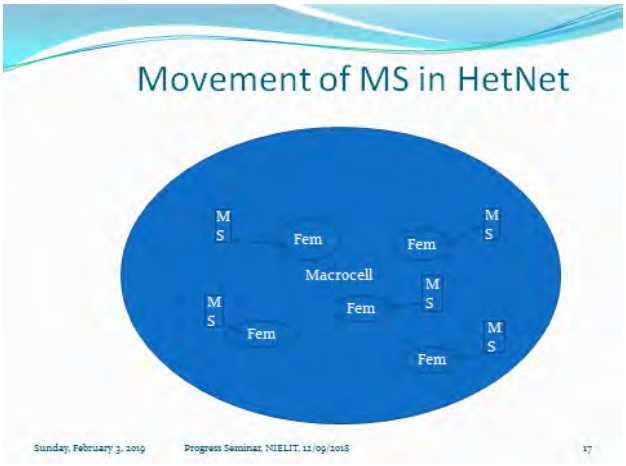


Fig. 1 Movement of mobile station in HetNet

Movement of mobile station in heterogeneous network is shown in figure one. HetNet is formed by macrocell, femtocell and picocell.

From fig. 2 shown below, it is femtocell system architecture. It consists of core network, access network and mobile equipment. Iu is the interface between core network and access network. Uu is the interface between femtocell access network and mobile equipment.

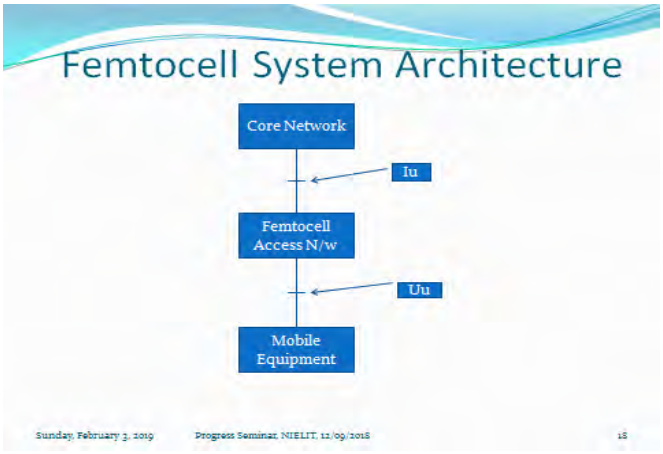


Fig.2 Femtocell System Architecture

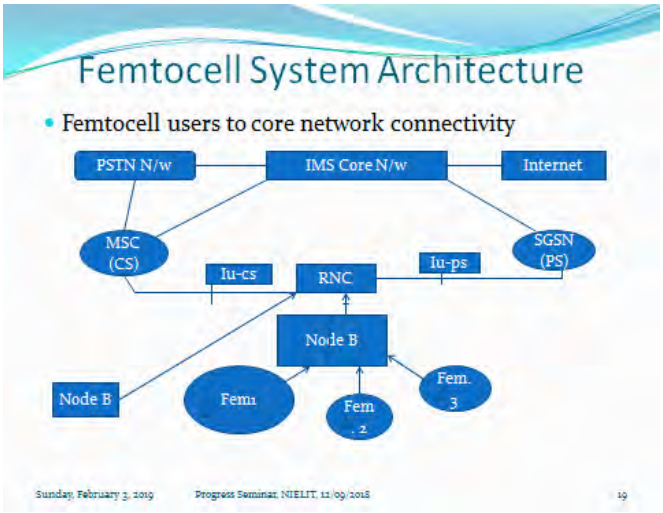


Fig. 3 Femtocell Users to Core Network Connectivity

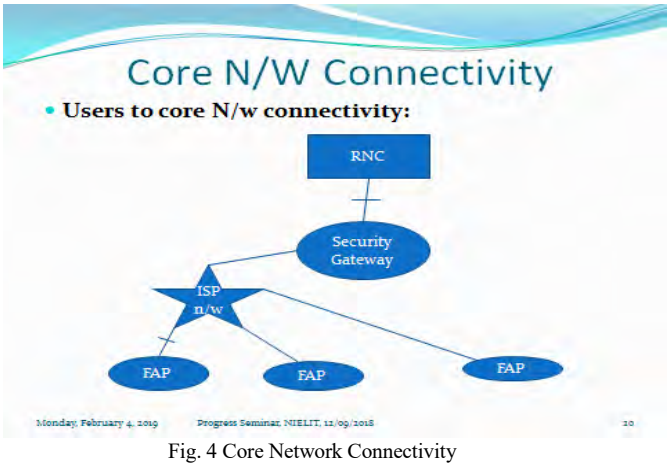


Fig. 4 Core Network Connectivity

Figure 3 also shows femtocell system architecture. It consist of circuitswitching and packetswitching network elements like mobile switching center, serving general packet radio service support node(SGSN) , interfaces Iu-cs and Pu-cs, radio network controller(RNC), nodeB, femtocells, PSTN network, IMS core network and internet.

Figure 4 shows users to core network connectivity. It consist of RNC, Security gateway, ISP network, and femtocell access

points. When call gets transferred from macrocell to femtocell, there are two handover phases. First is Preparation phase in which information gathering, handover decision, and authorisation check takes place. Second is Execution phase in which actual handoff takes place. Optimization is done on the basis of selection, reselection, and RRC management. Femtocell access point has radio resource control while nodeB has no radio resource control. Macrocell to femtocell handover is most challenging and complex. In femtocell to macrocell handover no complex interference calculation and authorization check is there. Measurement report send by mobile station to femtocell access point should consist interference level otherwise quality of service and capacity can be affected and unnecessary handoffs may increase. Modified handover policy is needed.

Call Admission Control (CAC) is a practice or process of regulating traffic volume in voice communications. In voice over IP or mobile networks to maintain certain quality of voice which is required. It works by regulating total unlimited bandwidth, total number of calls, and total number of packets or data bits passing through specific point per unit time. Actual content of a call may not be known to descriptor. It uses different algorithims used in handover process to accept handover call by femtocell access point.

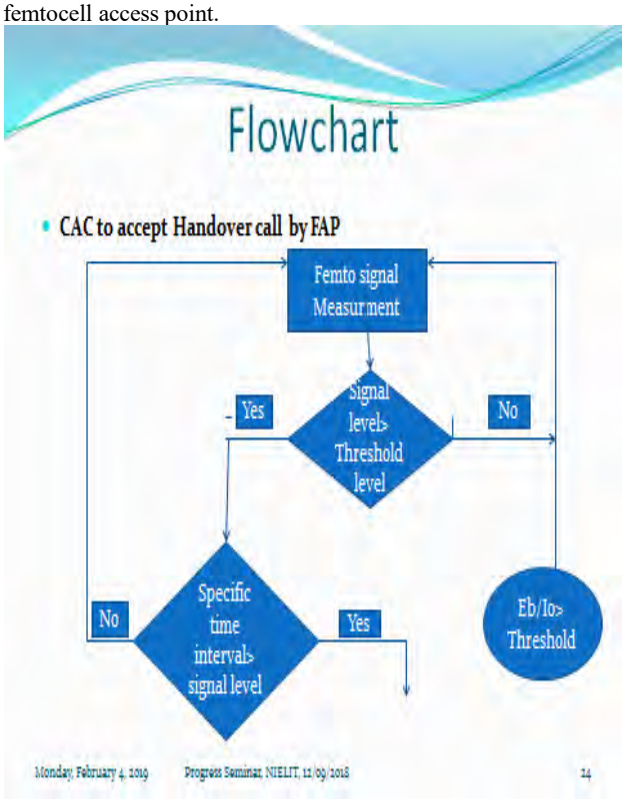


Fig. 5 Flowchart: CAC to accept handover call by FAP

Figure 5 shows flowchart call admission control to accept handover call by femtocell access point. Femtocell signal level is checked by comparing with threshold level for specific time interval. If signal level is not greater than threshold level, Eb/Io will be checked against threshold.

Figure 6 shows if Eb/Io of femtocell is greater than Eb/Io of macrocell.

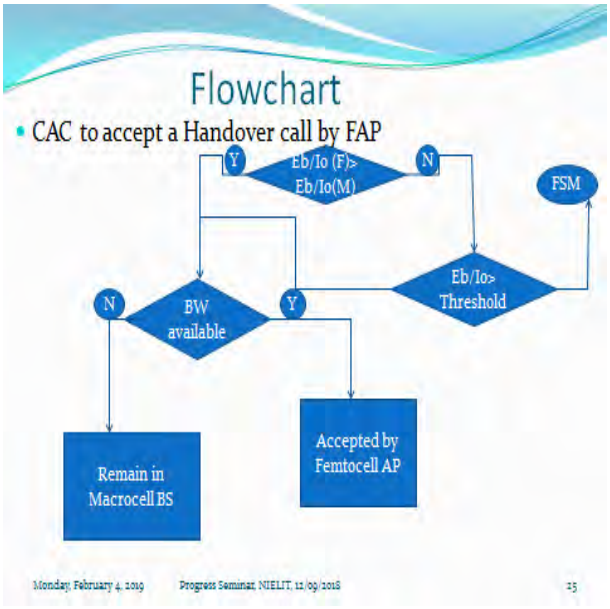


Fig. 6 Flowchart: CAC to accept handover call by FAP

Macrocell bandwidth availability will be checked and if it is available accepted by femtocell access point otherwise remain in macrocell base station. If Eb/Io of femtocell is not greater it will be checked against threshold and control is transferred to femtocell management system. Information flows from mobile station to femtocell access point to femtocellgateway. From gateway it goes to core network consist of mobile switching center cs-core, serving general packet radio service support node ps-core, internet multimedia system core network, public switching telephone network and internet. Interfaces used Iu-cs, Iu-ps, Iub (between femtocell gateway and radio network controller), and Iuh (between femtocell gateway and ISP).

In macrocell to femtocell handover some modification in network architecture and protocol architecture is needed to change signal flow for handover procedure. Larger the value of threshold time lower the number of handovers. Unnecessary handover optimization using proper call admission control (CAC) and radio resource management is there. Threshold level is the minimum level of signal that must be needed to handover a mobile station from macrocell to femtocell. Threshold Time (T) specified by the operator must be maintained when mobile station moves from macrocell to femtocell for that Tmin. Assumption made is hundred femtocell access points within a macrocell coverage area. Then we request for a single femtocell from the average result of hundred femtocells. Parameters need to be considered are shape of femtocell coverage, radius, average velocity of mobile station in femtocell coverage area, average call lifetime after handoff from macrocell to femtocell, call life time and user velocity, number of femtocell access points within macrocell. Simulation results obtained using parameter call admission control with time period ten seconds and twenty seconds respectively. Figure 7 shows cell of 800*800sq.met. consist of small cells obtained is shown below.

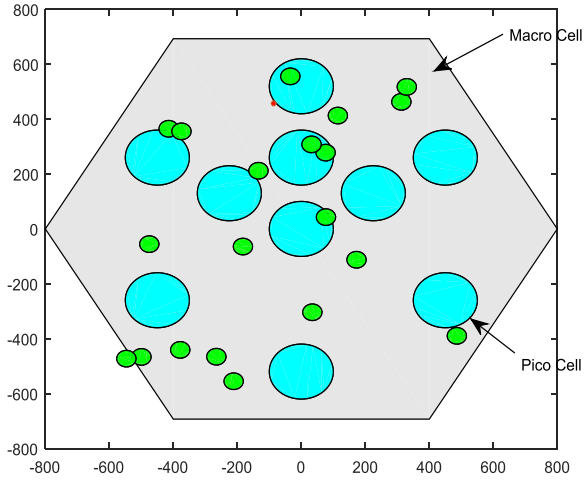


Fig. 7 Heterogeneous Network

It shows movement of users from macrocell to femtocell or picocell and from femtocell to macrocell. Handovers are more when user moves to femtocell from macrocell because handoff period is short and user can be high speed user.While there is smooth handover when user moves from femtocell to macrocell.

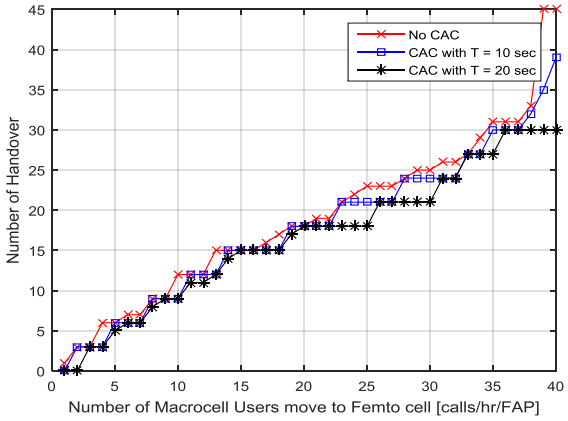


Fig. 8 Number of handovers vs users

From figure 8, it is clear that without call admission control number of handovers from macrocell to femtocell are 45. Number of handovers with threshold equal to 10 seconds reduces to 38. Number of handovers with threshold equal to 20 seconds are equal to 30. As threshold time increases number of handovers reduced and security improved.

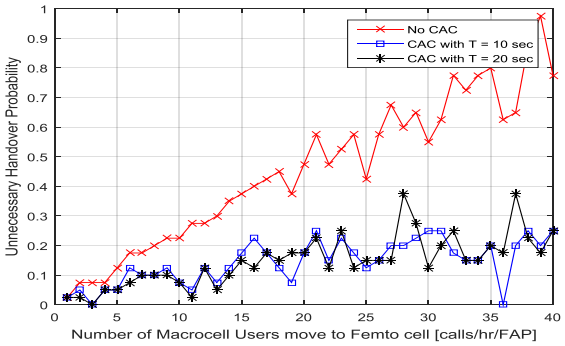


Fig. 9 Probability of unnecessary handovers vs number of users

From figure 9, it is clear that unnecessary handover probability without call admission control is 0.98. It goes on increasing when number of users interested in handover increases upto 38. Probability of call admission control with handoff threshold time period 10 seconds is 0.2 on an average and and on 20 seconds it is 0.15 on an average. These optimum results are obtained with implementation of this approach when compared with existing solutions.

III. CONCLUSION

Hetnet security is of prime concern in next generation networks which can be achieved in different ways. Results obtained with call admission control reduces number of handovers hence reduces probability of unnecessary handovers and call drop. Increase in threshold time also reduces number of handovers and handover security can be improved. Parameters like signal to noise ratio, signal to interference plus noise ratio and channel quality indicator can also be used in ultra dense network [2]. In future if handoff security task achieved for heterogeneous network along with securities in millimeter range, reliable, ultra high speed, secure and hundreds of gigabits per second communication is possible in fifth generation networks. Thus hetnet with femtocell is one of the good options in near future which can be implemented with other than UMTS network and is still open area of research.

ACKNOWLEDGMENT

Authors are grateful to Dr. Babsaheb Ambedkar Marathwada University Aurangabad affiliated NIELIT Research Center Aurangabad, Lokmanya Tilak college of Engineering Navi Mumbai for their kind support.

REFERENCES

[1] Mostafa Zaman Chaowdhury, Won Ryu, Eunjun Rhee, and Yeong Min Jang, “Handover between Macrocell and Femtocell for UMTS based Networks,” Kookmin Research Institute Korea, ICACT 2009, Feb. 15-18 2009.

[2] Garima Chopra, Sanjeev Jain, Rakeshkumar Jha, “Possible Security Attack Modeling in Ultra Dense Networks using High Speed Handover Management,” IEEE Transactions on Vehicular Technology, 2017.

[3] Xiaozheng, Behcet Sarikaya, “Handover Keying and It’s Uses,” IEEE Network, March/April2009, 0890-8044.

[4] Dionysis Xenakis, Nikos Passes, Christos Verikoukis, “A Novel Handover Decision Policy for Reducing Power Transmissions in the two-tier LTE Network ,” IEEE ICC 2012-Communication QoS, Reliability and Modelling Symposium, PP. 1352-1356.

[5] Rose Qingyang Hu, Yi Qian, Sastri Kota, Giovanni Giambene “HetNets A New Paradigm For Increasing Cellular Capacity and Coverage,” IEEE Wireless Communications, June 2011.

[6] Xianbin Wang, Peng Hao, Lajos Hanzo, “Physical-Layer Authentication for Wireless Security Enhancement:Current Challenges and Future Developments,” IEEE Communication Magazine, 2016.

[7] Tadilo Endeshaw Bogale, Long Bao, “Massive MIMO and Millimeter Wave for 5G Wireless HetNet: Potentials and Challenges” IEEE Vehicular Technology Magazine , October 2015.

[8] Julien Stephan, Mathieu Brau, Yoann Corre, Yves Lostanlen, “On the Effect of Realistic Traffic Demand Rise on LTE-A HetNet Performance” IEEE 2014.

[9] Yateng Hong, Xiaodong Xu, Mingliang Tao, Gingya Li, Tommy Svensson , “ Cross-tier Handover Analyses in Small Cell Networks: A Stochastic Geometry Approach” IEEE ICC Mobile and Wireless Networking Symposium 2015.PP. 3429-3434.

Ad-hoc Network and Database Object Analytics

Aniket Pathak, Sagar Parmar, Pankti Parikh, Vikas Kaul, Ranganathan Iyer, Manoj Nambiar

Department of Information Technology Thakur College of Engineering and Technology Mumbai, India
Manager Information Services and BI Bhakti Vedanta Hospital and Research Institute Mumbai, India

***Abstract* - Ad-hoc Network and Database Object Analytics provides a way of solving the problems faced by the businesses by saving time and effort of analysing the data. This paper discusses about creating dynamic dashboards which allows the decision makers to analyse the data of networks, database objects and other important data of healthcare unit. Here a business intelligence tool named Tableau is used as a front end for displaying the dashboards. Data is obtained from SQL Server through stored procedure. The job for the stored procedure is scheduled on the daily basis and is fetched to the Tableau management console. This job would run every day on fixed time. The dashboard will be stored on management console site of the healthcare unit whose link will be sent to the IT- Department so that they can directly access the dashboards and do their analysis on daily basis. It will take around 2-3 minutes to fetch the data from database.**

Keywords – analysis; business intelligence; networks; database object; database; data integration; dashboard; Tableau; healthcare unit database; schema; SQL server; management; IT.

I. INTRODUCTION

Data Analytics is a complicated subject, historically only tackled by well-trained and experienced experts. However, as more and more data is generated at various companies and corporate world, an increasing number of people need to understand the basics of data analysis in a corporate world. The objective of the object analytics system is to make analysis easier to understand, monitor, track and resolve, and to provide the business with an effective tool to identify and target problem areas, monitor records, handling performance and make improvements in the manual process [1]. in them timely, as well as provide an outlook of business opportunity [2].

Database Object Analytics is a management technique for assessing, analysing and responding to data. This software is used to record object data and respond to changes taking place

II. BACKGROUND

A Database Object is any defined object in a database that is used to store or reference data. Some examples of database objects include tables, views, clusters, sequences, triggers, stored procedures, indexes, and synonyms. The table is our focus because it is the primary and simplest form of data storage in a relational database. Also the data objects shall be connected using star schema. Data Analytics is the science of analyzing data to convert information to useful knowledge [3].

This knowledge helps us understand our current status and prospect better, and in many contexts enable us to make better, well-informed decisions that result in the benefit of the organization. Data Integration or database integration implies the ability to access any source of data within or outside your enterprise and use this information in virtually any application or mission-critical system.

Data may be stored in relational or non-relational/hierarchical database formats such as Oracle, DB2, SQL Server, MySQL, Essbase, FOCUS db and VSAM. Information may also be stored in ERP systems such as SAP. Access to the data is usually in the form of a query, reporting system or dashboard, but can also include update capabilities, and can happen in bulk, real-time, or near real-time [3].

III. IMPORTANCE

In many environments, the processing and data storage demands of advanced analytics applications have limited their adoption - but those barriers are beginning to fall.

The growing availability of data platforms and data analytics tools has enabled environments in which predictive and prescriptive analytics applications can scale to handle massive data volumes originating from a wide variety of sources [4].

In healthcare unit, analytics can be used to provide information on ways a healthcare unit can modify and make it facilities efficient. The dashboards we will produce through Tableau tool is actually to analyse the data of trusts. These dashboards give various information like number of patients visiting, number of patients doctor is seeing each day, the earning of the doctor, number of employees in each trust etc. All this will help the healthcare unit and the trusts to take some initiatives in making their services efficient. This project will also help the healthcare unit in giving the old manual way of doing the same work in Microsoft Excel by dashboards in Tableau which will be updated automatically.

Through the dashboards and data visualizations one can detect efficiency of employees working in the healthcare unit, details with respect to any specific patient and employee, upcoming problems and their solutions, unfocused areas that needs the focus of healthcare unit ,areas that are over focused and are not that important for growth of healthcare unit, etc.

IV. LITERATURE SURVEY

The study of comprehensive survey on big data analytics tools [1] shows how big data tools are very widely being used for analytics of data. Hadoop, Storm, HPCC, HBase and Grid Grin are some of the most widely used tools. We are also using a data analytics tool Tableau which provides user friendly dashboards.

MultiBase: a heterogeneous multibase management system [5] gives information on how SQL interface can be used to modify data objects and integrating local schemas using multibase data interpretation language. Also our project will be using multibase and SQL interface to deal with data objects in backend. We will use SSIS for connection with Tableau in the frontend.

The study of Generating Data Mart Schema from OLAP Requirements [6] gives information on the schema structure and its decomposition into set of categories which are: fact, dimension, measure, attribute, parameter, and hierarchy. Our system will also provide a schema comparison method along with the schema similarity which will remove the redundancy.

V. PROBLEM DEFINITION

This project would explore the field of Business Intelligence with the broad aim of designing dashboards capable of providing multidimensional view of trusts details and thus helping the healthcare unit analyse their reports more easily. In this project we will create a centralized database which allows healthcare unit to fetch and access the dynamic dashboards which gets scheduled and updated daily, so that they can evaluate and analyse each and every aspect of the trusts data and take decision based on data accordingly.

However, there are useful strategies and tools to make data from trusts easier to interpret and work with. One such tool that we will be using is Tableau which is a popular and well known BI analytical tool for creating dynamic dashboards. Tableau provides many facilities such as in memory analytics, direct/live query, on demand app generation and API on demand. Also Tableau tool is less costly and more advanced as it provides user friendly dashboards which is helpful to analyse data effectively [7].

Data is the lifeblood of the healthcare unit management business, giving insight to an organization's current status and thus helping actuaries and other analysts make informed decisions and projections on a daily basis. Because of the complexity and size of trusts data, working with it can be frustrating and challenging as each new data set has unique issues and problems. In addition, trusts data is a rapidly evolving and difficult-to-understand breed of business data.

VI. MODULES OF THE SYSTEM

a) User Module:

In this module we will have the dashboards for the management team. These dashboards will include all the graphs, pie charts, lists, tables and other criterias and outputs so that the management team will be able to make effective decisions for their business. The statistics will be distinct and helpful for the staff to use it precisely.

b) Database Module:

The proposed system uses SQL server in the backend for simplicity and flexibility. This module stores all the information about the healthcare unit staff, patients, and network related details such as active switches and machines and model their data on specified operations. These operations can be importing data from the trusts and updating them timely.

VII. METHODOLOGY

1. A data warehouse will be created which will contain the data objects of all the 4 trusts all in connection with the star schema [8].
2. A data dictionary will be used to retrieve the data objects, size of the objects, objects last modified and creating pivot table.
3. In latter part Tableau business tool will be used to create a dashboard to analyze and get precise information timely.
4. This process will continue everyday as the data gets updated in the data warehouse. Thus a regular and periodic analysis of the data of the healthcare unit and its trusts will be maintained.

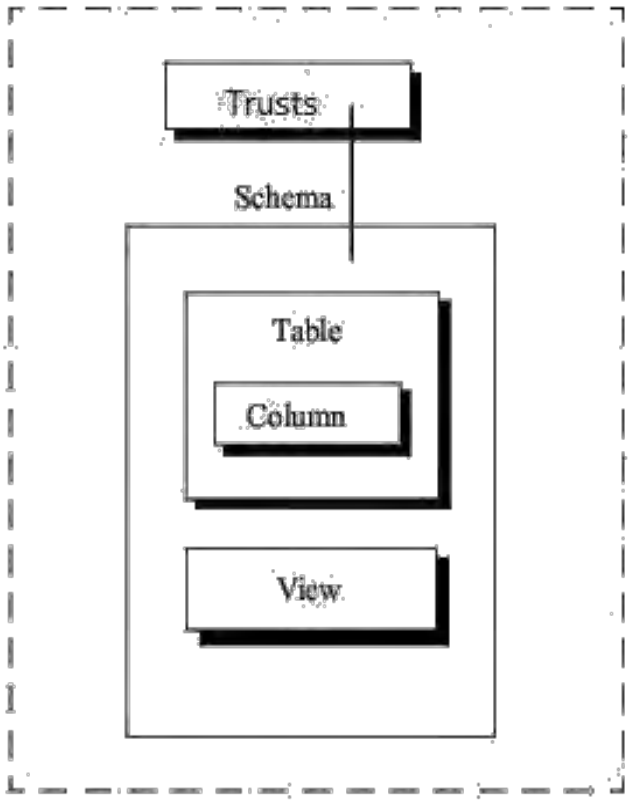
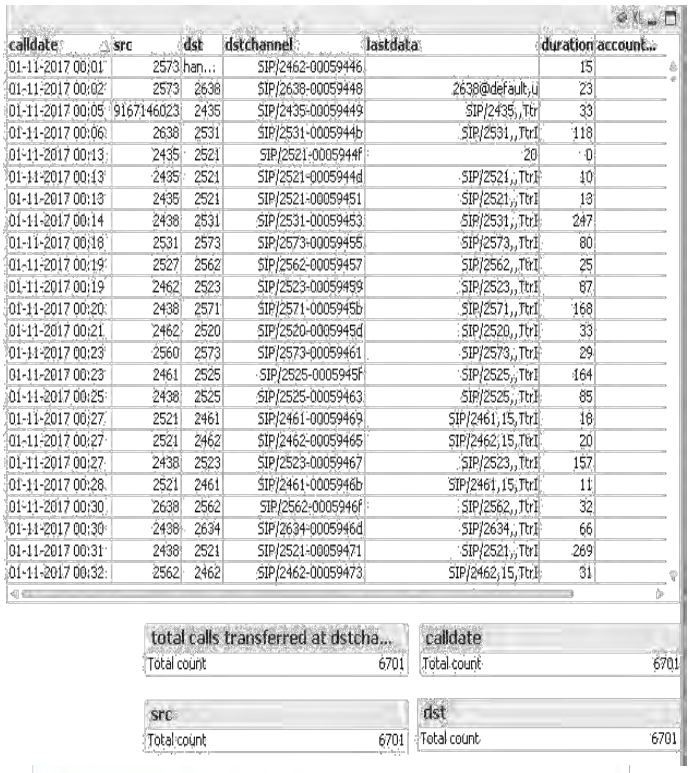


Fig 1: Project flow



calldate	src	dst	dstchannel	lastdata	duration	account...
01-11-2017 00:01	2573	han...	SIP/2462-00059446		15	
01-11-2017 00:02	2573	2638	SIP/2638-00059448	2638@default,u	23	
01-11-2017 00:05	9167146023	2435	SIP/2435-00059449	SIP/2435,,Ttr	33	
01-11-2017 00:06	2638	2531	SIP/2531-0005944b	SIP/2531,,TtrI	118	
01-11-2017 00:13	2435	2521	SIP/2521-0005944f	20	0	
01-11-2017 00:13	2435	2521	SIP/2521-0005944d	SIP/2521,,TtrI	10	
01-11-2017 00:13	2435	2521	SIP/2521-00059451	SIP/2521,,TtrI	13	
01-11-2017 00:14	2438	2531	SIP/2531-00059453	SIP/2531,,TtrI	247	
01-11-2017 00:18	2531	2573	SIP/2573-00059455	SIP/2573,,TtrI	80	
01-11-2017 00:19	2527	2562	SIP/2562-00059457	SIP/2562,,TtrI	25	
01-11-2017 00:19	2462	2523	SIP/2523-00059459	SIP/2523,,TtrI	87	
01-11-2017 00:20	2438	2571	SIP/2571-0005945b	SIP/2571,,TtrI	168	
01-11-2017 00:21	2462	2520	SIP/2520-0005945d	SIP/2520,,TtrI	33	
01-11-2017 00:23	2560	2573	SIP/2573-00059461	SIP/2573,,TtrI	29	
01-11-2017 00:23	2461	2525	SIP/2525-0005945f	SIP/2525,,TtrI	164	
01-11-2017 00:25	2438	2525	SIP/2525-00059463	SIP/2525,,TtrI	85	
01-11-2017 00:27	2521	2461	SIP/2461-00059469	SIP/2461,15,TtrI	18	
01-11-2017 00:27	2521	2462	SIP/2462-00059465	SIP/2462,15,TtrI	20	
01-11-2017 00:27	2438	2523	SIP/2523-00059467	SIP/2523,,TtrI	157	
01-11-2017 00:28	2521	2461	SIP/2461-0005946b	SIP/2461,15,TtrI	11	
01-11-2017 00:30	2638	2562	SIP/2562-0005946f	SIP/2562,,TtrI	32	
01-11-2017 00:30	2438	2634	SIP/2634-0005946d	SIP/2634,,TtrI	66	
01-11-2017 00:31	2438	2521	SIP/2521-00059471	SIP/2521,,TtrI	269	
01-11-2017 00:32	2562	2462	SIP/2462-00059473	SIP/2462,15,TtrI	31	

total calls transferred at dstcha...	
Total count	6701

calldate	
Total count	6701

src	
Total count	6701

dst	
Total count	6701

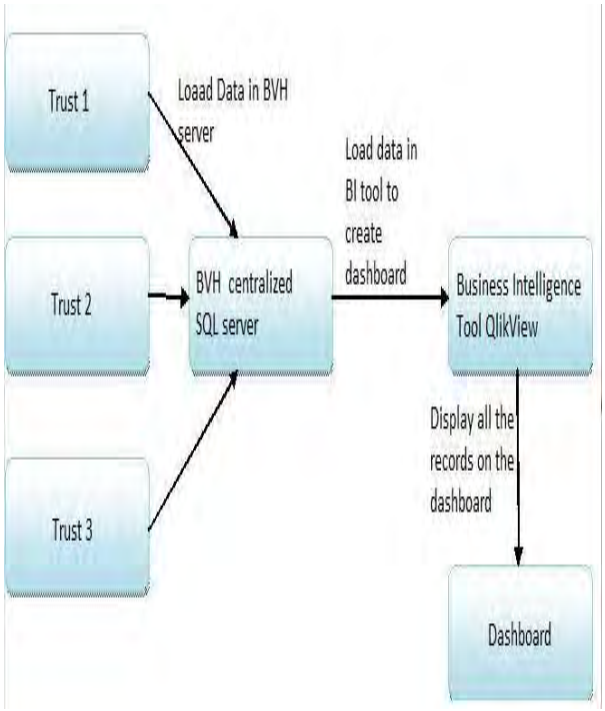


Figure 2: Data Flow Diagram

VIII. RESULTS

The Tableau dashboard we are currently working upon for the healthcare unit is on telephony data which is a part of ad-hoc network in the healthcare unit. At present the healthcare unit management is not able to decide on how many staffs are needed at the reception to attend calls. After the dashboard has been deployed in the healthcare unit the management staff is now studying and making mature and effective decisions on how many more or less staffs are needed at reception at a particular time.

We shifted from the QlikView dashboard to the Tableau dashboard, which is a faster and effective processing business intelligence tool, in the latter stage. The purpose of changing the tool was to help the healthcare unit with much better service such as quick decisions, more analytical studies, user friendly dashboards, etc. With all the process done properly in

the completion of the project the healthcare unit has been deployed with a full-fledged solution for their problem.

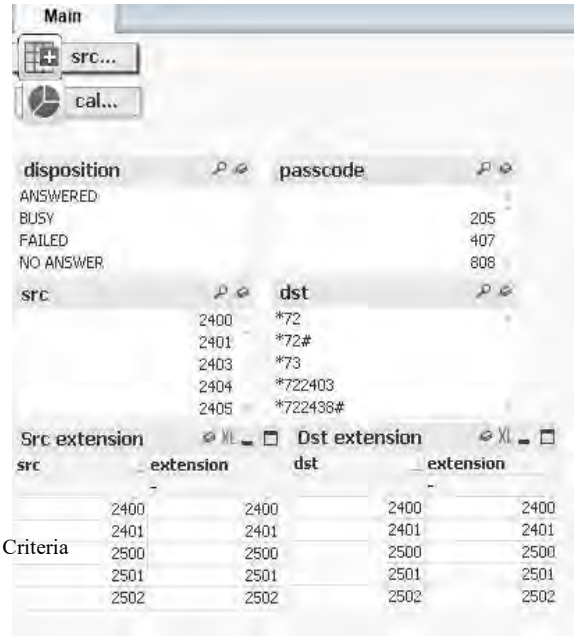


Figure 3: Criteria

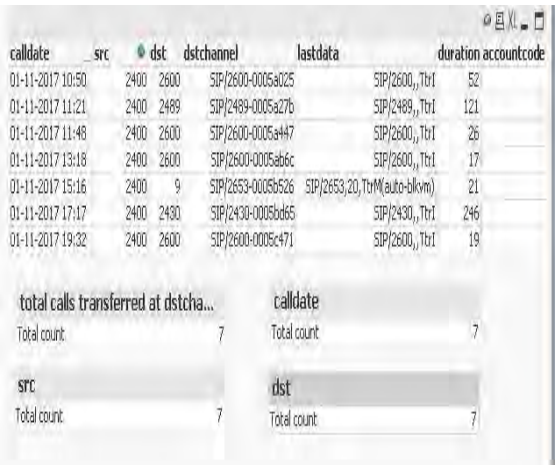


Figure 4: Output and Statistics



Figure 5: Selected Criteria

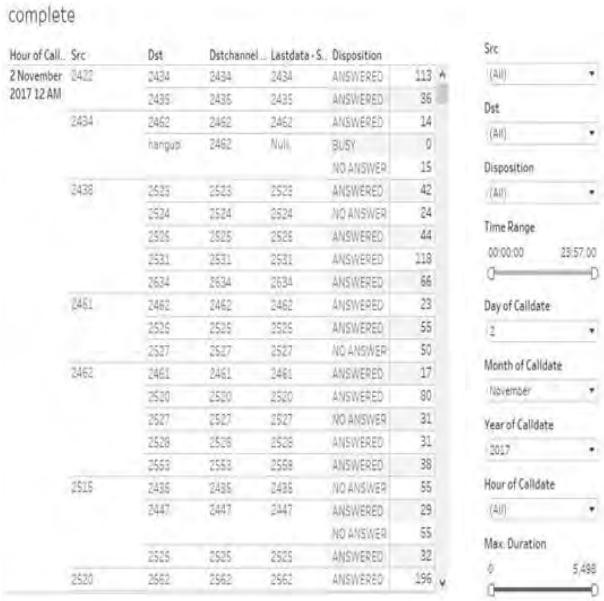


Figure 7: Complete Tableau dashboard with criteria

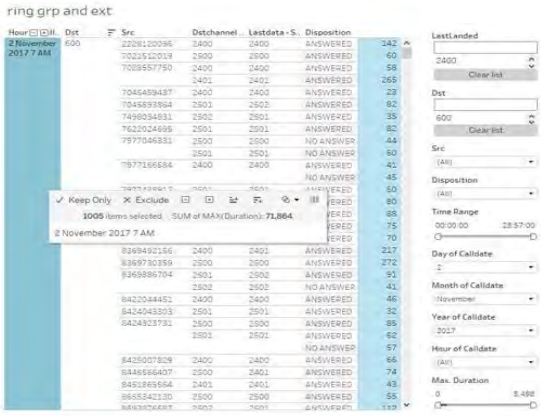


Figure 8: Ring group and extension Tableau dashboard with criteria

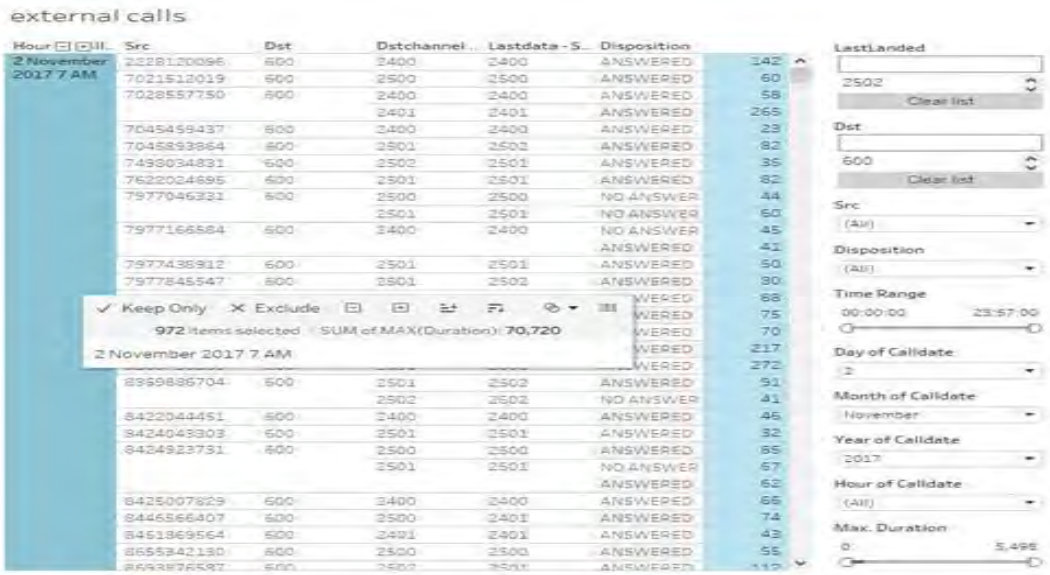


Figure 9: External calls Tableau dashboard with criteria

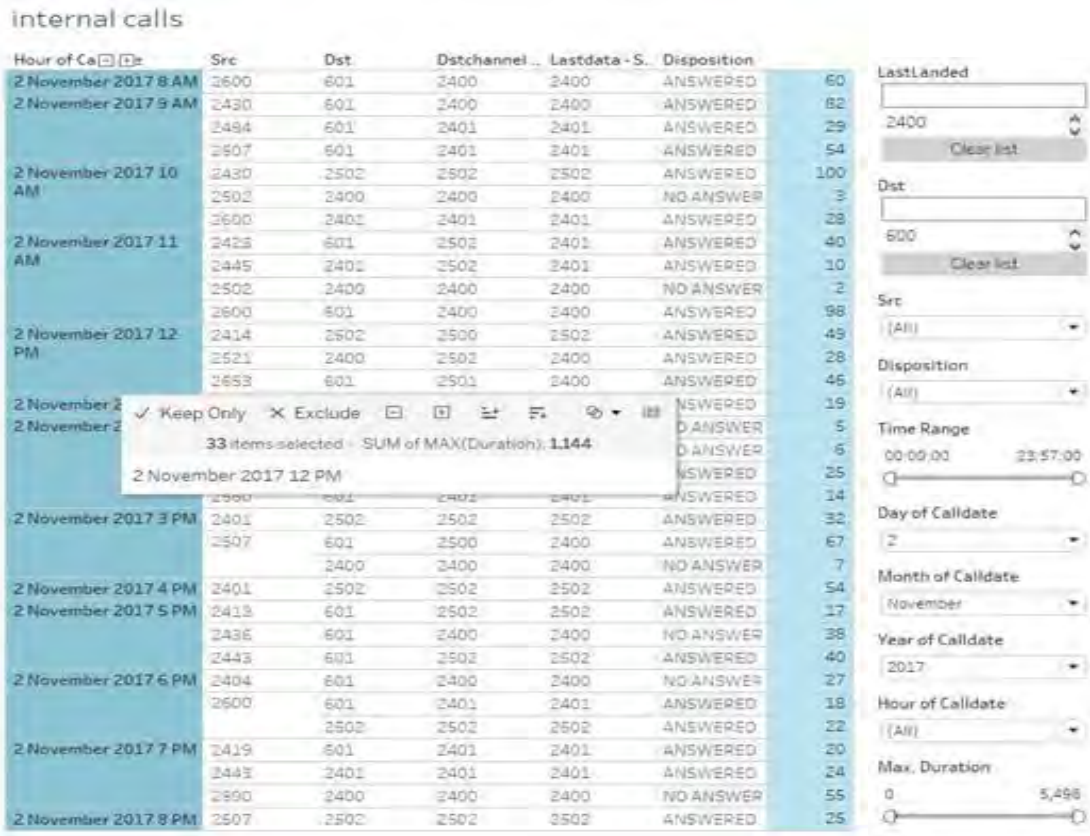


Fig 10: Internal calls Tableau dashboard with criteria

IX. DISCUSSIONS

In the above discussion, we have seen about the Tableau tool, how it works and advantages of using this tool over other tools. We also saw the problems being faced by the healthcare unit and how effectively it can be solved with the help of the Tableau tool. Data today is the most important asset for any organization to grow in their respective field. Thus analysis of data is very important so as to extract the right information from the huge amount of data. This analysis should be in a way that it is clear and the management is able to make decisions out of it, which can be done effectively with Tableau. This tool can be effectively used in different fields where large and complex number of data is being analyzed.

X. CONCLUSION

We learnt that actually creating dashboard incorporates all the data retrieval concepts and the parameters that we have studied so far we were able to create dashboards for the telephony data of the healthcare unit. We are looking forward to the other part of our project that is working with database objects which includes changing the ETL process of the architecture. Currently used ETL process is stored procedure which leads to a full load and hence we are changing it from full load to Incremental load (day wise load) using SSIS.

While working with the database objects, integrating all the data from various tables, it will be easy for the healthcare unit to access all the data at one place, saving them the time to find the table in which the data belongs. Also, the dashboard gives a visual analysis of the data which is much easier than analyzing the star schema. Thus, we can conclude that by the proposed system for data analysis can be used effectively by the organization for better decision making.

XI. Future Scope

The IT department using our software will be able to access precise data of all the related trusts of the Bhaktivedanta Healthcare unit and use it intelligently to help all the trusts in maintaining the database efficiently. This software would be useful for any business having related branches like Bhaktivedanta healthcare unit as it does the work of connecting the data objects of many different databases together into a centralized database and keep record of all the data objects that are growing with time and provide many business opportunities.

Tableau dashboard can be used by any organization so as to keep track of their data time to time. In today's time data is said to be the biggest asset for any company or organization. Thus, it becomes of immense importance to organize the storage of data and keep a track of it. With proper and timely tracking of this data a business organization can find ample of growth opportunities and expand their business.

Organizations having multiple branches can increase their efficiency of work by centralizing their database and keeping record of every single branch's collection and other important stuff's like employee details, clients, services etc. In today's time a company having well managed database with the tracking of the data timely is the one that can grow at a faster pace. Centralized database helps in assessing and accessing the data from multiple databases at one single point. This helps in effective utilization of the data and also tracking the same at one central point.

XI. ACKNOWLEDGMENT

We are doing this project for BhaktiVedanta hospital. The hospital staff is always concerned and very generous and helpful for the project work. Our special thanks to Mr. Ranganathan Iyer and Mr. Manoj Nambier for their collaborative efforts and time to provide us with all the necessary documents, place and invaluable support in doing the project.

REFERENCES

[1]. J. Vijayaraj, R. Saravanan, P. Victer Paul and R. Raju, "A comprehensive survey on big data analytics tools," 2016 Online International Conference on Green Engineering and Technologies (IC-GET), Coimbatore, 2016, pp. 1-6.

[2]. https://docs.oracle.com/cd/B12037_01/server.101/b10759 /statem ents_4005.htm

[3]. https://docs.oracle.com/cd/B28359_01/server.111/b283 10/general002.htm# ADMIN11524

[4].]http://searchbusinessanalytics.techtarget.com/feature /Alteryx-advanced-analytics-tools-simplify-data-analysis

[5]. Jer-Wen Huang, " MultiBase: a heterogeneous multibase management system", Proceedings Eighteenth Annual International Computer Software and Applications Conference (COMPSCA 94), Taipei, 1994, pp. 332-339.

[6]. Nouha Arfaoui and Jalel Akaichi "Generating Data Mart Schema from OLAP Requirements", 2013 Third International conference on information technology and e-services(ICITeS),Sousse,2013, pp.1-

[7]. https://community.qlik.com/blogs/qlikviewdesignblog/20 16/08/02/how-do-you-handle-big-data

[8].]http://www.cs.toronto.edu/~nn/csc309/guide/pointbase /docs/html/htmlfiles/ dev_dataobjectsFIN.html

Vehicle Tracking/Parking System

Mrs. Vandana Munde, Shreyas Snail, Yash Mujumdar, Shweta Panchal
Information Technology
Thakur College of Engineering & Technology
Mumbai, India

***Abstract* – In recent times, the concept of smart cities has gained great popularity. Looking at the current scenario the problems such as, traffic congestion, limited car parking and tracking facilities and road safety are being addressed by IoT. In this paper, we present an IoT based System integrated reverse parking and vehicle tracking system. The proposed Reversed Parking and vehicle tracking system consists of an on-site deployment of an IoT module that is used to monitor and signalize the tracking path and the location of the vehicle. Proposed System is a tracking device focusing on vehicles that GSM Module and GPS module for the location of the vehicle with the help of longitude and altitude including obstacle detection with the help of the ultrasonic sensors. Proposed System will give information of location coordinate to mobile phone whenever there is a request for it through the Notification and the sensors will beep and alert the user or the driver. The integrated Global Positioning System and Global System for Mobile communication (GPS- GSM) is used to track vehicle using application such as Google Maps or Google Location API.**

Keywords – Tracking; Parking Assistance IoT; Ultrasonic Sensors, GSM Modem, Google Location API.

I. INTRODUCTION

Due to growth in population in society the common problem everyone faces is parking and tracking a vehicle irrespective of the circumstances at social places such as shopping-malls, local areas, commercial and industrial areas. For matter will need to create these components, incorporating the applicable criteria that follow Vehicle tracking system is a system used to determine the location of a vehicle

using method like Global Positioning System (GPS) and operating system through satellites and device. (Reference 1) By using the vehicle tracking system, it is easy to calculate and get the accurate location of the vehicle. Assisted driving technologies which use display system for driver (partial automation). A system and method of controlling an automotive vehicle includes determining a steering wheel angle, determining a steering wheel direction. Instead of completely relying on human intervention, use of device to take charge in critical situations. Reducing the time consumed to park in difficult situations through the help of this aid. Intelligent driver assistance system provides the full assistance considering the traffic intensity & white lane detection. Reverse parking assistance provided by using Ultrasonic Sensor.

II. BACKGROUND

Vehicle tracking system is a system used to determine the location of a vehicle using method like Global Positioning System (GPS) and operating system through satellites and device. By using the vehicle tracking system, it is easy to calculate and get the accurate location of the vehicle. Assisted driving technologies which use display system for driver (partial automation). A system and method of controlling an automotive vehicle includes determining a steering wheel angle, determining a steering wheel direction. Instead of completely relying on human intervention, use of device to take charge in critical situations. Reducing the time consumed to park in difficult situations through the help of this aid. Intelligent driver assistance system provides the full assistance considering the traffic intensity & white lane detection. Reverse parking assistance provided by using Ultrasonic Sensor.

III. MOTIVATION

The common problem everyone faces is parking and tracking a vehicle irrespective of the circumstances at social places such as shopping-malls, local areas, commercial and industrial areas. Vehicle tracking system main aim is to give Security to all vehicles; as well as Locals can easily track/find the vehicle, where talking about reverse parking assistance is introduced after looking at the current scenario where it is being difficult for locals to park in commercial and industrial not only the system is applicable for reverse parking but as well as for safety purpose on hill side areas.

IV. PROBLEM DEFINITION

Due to growth in population in society the common problem everyone faces is parking and tracking a vehicle irrespective of the circumstances at social places such as shopping-malls, local areas, commercial and industrial areas. Vehicle tracking system main aim is to give Security to all vehicles. This is improved security systems for vehicles. After going through some research papers “Pham Hoang Oat, Micheal Drieberg and Nguyen Chi Cuong Development of Vehicle Tracking System using GPS and GSM Modem”. And understanding the current scenario we have been able to develop a prototype, where with the association of the GSM Modem and GPS Modem it’s been easy to build a system which easily tracks the location of the vehicle with help of the network. And it has added value to it with the help of its features such as: Obstacle Detection that ensures safety of vehicle and smooth experience. Security through tracking system that used location tracking It can be used for trucks carrying valuable goods, to keep track of the status of delivery and location of the truck at all times You can also use it to keep tab on your driver. It reduces vehicle abuse and ultimately results in significant cost-savings for individuals, fleet owners and the like. You can locate your stolen vehicle easily using your mobile without any extra cost. Semi-Automatic Parking Experience that helps in parking our vehicles in difficult situation.

Phase-1: - Market Research:

Due to growth in population in society the common problem everyone faces is parking and tracking a vehicle irrespective of the circumstances at social places such as shopping-malls, local areas, commercial and industrial areas. However, there are not much parking spaces available and as per the current condition, even the available parking spaces are being congestion for parking of multiple vehicles due to manual management.

Secondly, there are also no smart applications and features available at the moment for the management and parking of vehicles which make the management of parking spaces more difficult. Smart parking systems enable individual to save time and reduce traffic problems as well as improves the way people commute. Increasing number of automobile and lack of efficient parking management systems are some of the factors driving the demand for smart parking across the globe.

These systems help in making the cities greener, cleaner, and technologically advanced by offering several opportunities for the reuse of commercial and municipal parking areas and garages. Moreover, cell phone-enabled automated solutions and low-cost sensors boosting the market for s mart parking and Tracking.

Phase-2: Requirement Gathering:

For the development of the system, certain requirements need to be gathered. The first step in the systems engineering process is understanding the needs of the relevant stakeholders and ascertaining the functional requirements of the consumer approach. To that end, conducted an extensive stakeholder assessment. The goal of the developer assessment is to identify the reverse parking challenges and tracking of the vehicle in the city and to illicit the role of a technology solution to address these challenges and understanding the hardware and software requirement.

Hardware:

- Ultrasonic Sensors
- GSM Modem
- GPS Modem
- Raspberry Pi b+
- Wires
- Mechanical Tools

Software:

- Python Idle

Phase-3:- Development of Base Model:

After the market research and requirement gathering, we begin with the development of base model.

The development procedure will include developing of a mobile connectivity and configuration of server using Raspberry Pi. The Mobile device will be used for live location detection and navigation for the end user for tracking the vehicle and while reverse parking the vehicle selected parking space.

The Following is the flow of the proposed system:



Fig .1 Work flow of proposed model

The Application will provide the detection of obstacle while reverse parking and tracking of vehicle in the public area and transportation and navigate the user towards the destination where the user car is parked.

The server will be configured on the raspberry pi and will manage the processing of obstacle detection and tracking the location entry from the user end. The Vehicle latitude and altitude will be shown on LCD Screen or the end user using device.

The Camera used for obstacle detection and GPS GSM Module for tracking the vehicle location will be a module of the raspberry pi programmed to detect objects and track the vehicle.

Phase-4: Testing of Base Model/ Prototype

Once the base model and prototype is ready, the system will be tested in real time at a limited parking location. The testing phase would be named VTPS and will include testing in following aspects:

i. Obstacle detection efficiency.
ii. Simulation of parking procedure through the Camera module.
iii. Testing of navigation to the selected vehicle.
iv. Testing of server efficiency.
v. The testing will check these aspects and feedback will be recorded. These feedbacks will be used to perform improvements in the next phase and release the final product.

Phase-5: Deployment of Project

In the implementation phase of the project, we will include the bug fixes and changes that we find in the testing phase of the project.

V. FEATURES

- Semi-Automatic Parking Experience that helps in parking our vehicles in difficult situation.
- Obstacle Detection that ensures safety of vehicle and smooth experience.
- Vehicle Tracking allows the user the track their vehicle at a national/global level.
- Connectivity with mobile devices makes it easier for users to operate the proposed module.
- Transporting System
- Security through tracking system that used location tracking.
- Safety while public transportation.
- You can locate your stolen vehicle easily using your mobile without any extra cost.

The below block diagram explains the working of the system. Controlling of the other supporting devices those are GPS, GSM, ultrasonic sensors, mobile, Raspberry pi, LCD screen. Actually, this gives the practical working of a vehicle tracking & parking system which can do reverse assisting for parking, tracking of moving vehicles in area. It consists of two sections, first which will be built on the prototype which is having GPS in it and as the car moves the location of the car goes on changing continuously, the GPS find the location in terms of two co-ordinates that are Longitude and latitude. These co-ordinates are communicated to another section by GSM device, as shown in the block diagram both the part consist of GSM for communication. This GSM is connected to the computer through RS232 which should have internet connection in it. So, that as soon as the co-ordinates are received, it can be located in the Maps & the direction of vehicle.

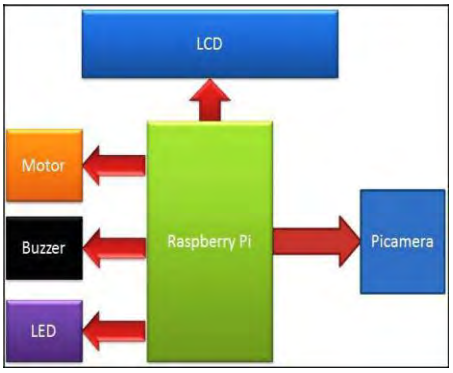


Fig. 2 Work Flow of Reverse Parking

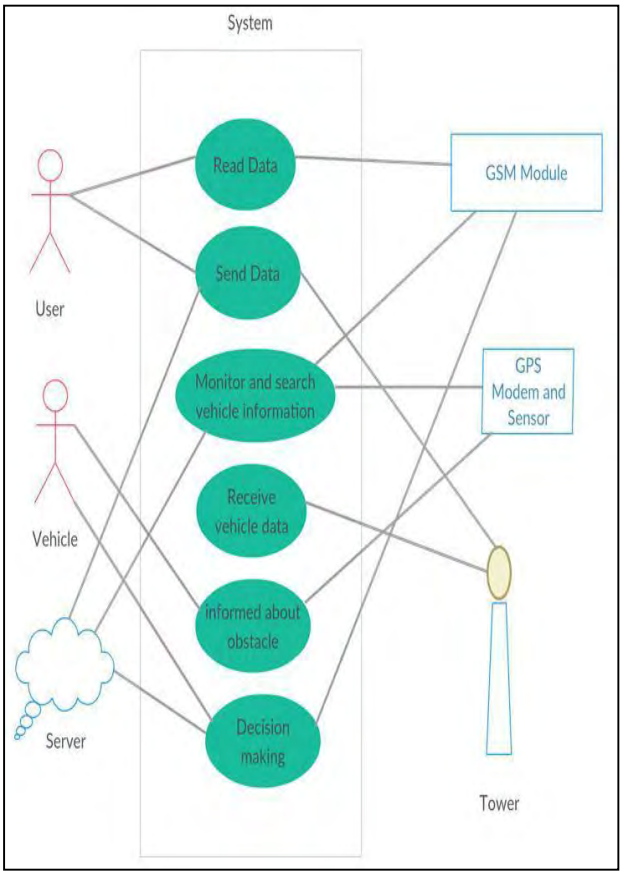


Fig. 3 Use Case Diagram

VI. METHODOLOGY

This vehicle tracking system takes input from GPS and sends it through the GSM module to desired mobile/ laptop using mobile communication. Vehicle Tracking System is one of the biggest technological advancements to track the activities of the vehicle. The security system uses Global Positioning System GPS, to find the location of the monitored or tracked vehicle and then uses satellite or radio systems to send to send the coordinates and the location data to the monitoring center. At monitoring center, software is used to plot the Vehicle on a map. In this way the Vehicle owners are able to track their vehicle on a real-time basis. Due to real-time tracking facility, vehicle tracking systems are becoming increasingly popular among owners of expensive vehicles. The below block diagram explains the working of the system. Controlling of the other supporting devices those are GPS, GSM, ultrasonic sensors, mobile, Raspberry pi, LCD screen. Actually, this gives the practical working of a vehicle tracking & parking system which can do reverse assisting for parking, tracking of moving vehicles in area. It consists of two sections, first which will be built on the prototype which is having GPS in it and as the car moves the location of the car goes on changing continuously, the GPS find the location in terms of two co-ordinates

that are Longitude and latitude. These co-ordinates are communicated to another section by GSM device, as shown in the block diagram both the part consist of GSM for

communication. This GSM is connected to the computer through RS232 which should have internet connection in it. So, that as soon as the co-ordinates are received, it can be located.

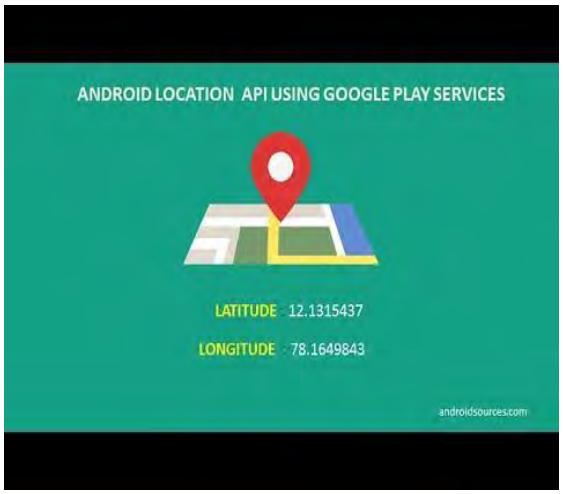


Fig. 4 Proposed System Output

VII. RESULTS & CONCLUSION

This is the final phase of development. In this phase, the improvements suggested and recorded in the testing phase are implemented and a final product is built. This final product is then deployed in the said project deployment area.

Phase-1:

The market research phase has enabled us to learn more and create a report on the current infrastructure of the tracking and parking. Based on the report and study of the current systems, a new smart infrastructure was developed by us that will improve on the currently present iterations and overcome their limitations.

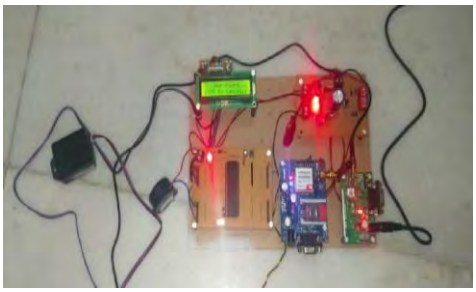
Phase-2:

In the requirement gathering phase, a complete report on the stakeholder requirements was collectively prepared by us. Secondly, a study on the existing smart systems helped us in developing the desired outcome based on our understanding of the modules present in the market. Based on the learnings we gathered, requirements and processes for our system was designed and developed by us.

Phase-3:

In this phase, the prototype of the model was developed that was based on the learning and requirements obtained from phase-5. The development of the project was divided into 3 phases mentioned below:

1. Development of parking system using a RP module.
2. Development of safety protocols that stop the vehicle on the detection of obstacle on it's path using ultrasonic sensors.
3. Development of tracking system with the help of MC that uses of GPS and GSM modules.
4. The next stage was creating a client-server relationship using a database on the server side with the help of two GSM modules and the GPS module for indicating the location.
5. With the help of pressure sensors on the occasion of an accident the GSM module on the client side will send the location of the totaled vehicle to the hospital acting as the server side.



Phase-4:

In this phase, the development of the prepared prototype was thoroughly tested and the remaining features of the project were tested. Based on the outcome of the test, the improvements were implemented and a better final product was built by us.



Phase-5:

This was the final phase of development. In this phase, the improvements suggested and recorded in the testing phase were implemented and a final product was built. This final product can now be deployed in real time.

FUTURE SCOPE

More and more businesses are turning to GPS technology to manage vehicles, their employees, and their assets. Businesses use GPS Vehicle Tracking and reverse parking assistance to assess driver performance and to keep track and parking of

users while they are travelling. As the devices become smaller and more accurate, it becomes a safe bet that even more businesses will begin to utilize this powerful tool. It is mainly benefit for the companies which are based on transport system. Since it can show the position of all vehicles in real time, so that they can create the expected data accordingly. This tracking system can store the whole data where the vehicle had gone, where did it stop, how much time it take at every stop and can create whole data analysis. It is also used in buses and trains, to estimate how far are they, how much time it takes for them to come to a particular stop.

REFERENCES

[1]. Abhirup Khanna *"IoT based Smart Parking System* 978-1-5090-0044-9/ 16/$31.00©2016IEEE Pune, India.

[2]. Pham Hoang Oat*, Micheal Drieberg and Nguyen Chi Cuong Development of Vehicle Tracking System using GPS and GSM Modem* 978 -1-4799-0285-9/13/$31.00 ©2013 IEEE Malaysia

[3]. Fatin Balkis *Binti A lzahri Vehicle Tracking Device* 978-1-5090-1636-5/ 16/$31.00 ©2016 IEEE Pu lau Pinang, Malaysia.

Authentication and Access Control

Akanksha Jain, Shivani Rege, Pooja Jain, Mr. Sudhir Dhekane
Department of Information Technology,Thakur College of Engineering,Mumbai

Abstract **- Authentication and access control of computer applications is today's need in the technical education institutes. It is common practice that students performing various subject labs where they are utilizing computer applications actually required to perform task rarely, rather they are busy utilizing the application irreverent w.r.t lab task. So a system is needed to be designed which takes care of the authentication and access control. This can be achieved by allowing authenticated users to use the computer labs and actual application and blocking unnecessary application to valid users. After the process of authentication is done it is necessary to give access rights to the authenticated users according to their lab requirement.In this work, the system designed would be tested on the college labs. This system will have a login page which would test the authenticity of the user and then allow access to the user if the credentials are valid. The students would be allowed to access applications that are necessary for that particular lab session. This access of applications would be dependent on the timetable of the labs.The system would be implemented in college labs. There would be three type of login i.e. teacher login, student login, guest login. Teachers would be allowed to access all applications as well as change the timetable fed in server computer. The student would only be able to access applications according to the timetable of the labs. The applications required for that particular lab would only be accessible. This would ensure that all the students are performing the intended task.**

Keywords: Authentication, Access control.

1. I. INTRODUCTION

Access Control and Authentication is basically allowing access rights to particular users after they are validated with the user after they are validated with the user identity database. Nowadays, Security of data is of utmost important as the data present in the computer can be misused by unauthorized intrusions. An intruder can modify and change the program source codes and can also use your pictures or email accounts to create derogatory content such as, fake misleading and offensive social accounts. Thus, access to data, accounts etc should be given to only authenticated users. After authentication process, the access to resources should be limited too. Giving access to all resources to each authenticated user would lead to misuse and harm to data. Therefore access rights play an important role in any company, educational institutions etc which classify the authenticated users and grant access to resources accordingly.

2. II. RELATED WORK

According to the author the system is used to give all access to admin and then it also defined the user groups that can be characterized by security levels, roles or any partition of users. A user can access any data that has lower or equal security level and i.e. accessible by the roles to which the user is assigned. All type of access mode are controlled in this phase [3].

In this paper, the Author has specified that the system is having different access right to different level of users. It species the need of access control in particular task. It also specifies that usage of different access control which help the user to give different right to different user module[4]. Here the author has written about the system specifies the need of new access control as Content Based Access Control where each user is allowed by RBAC rule to access a large set of data objects. In CBAC, each user is allowed by an MLS or RBAC rule to access a large set of data objects, while the CBAC rule imposes an additional layer of restrictions that the user could only access "a subset" of the designated records.[5]

3. III. PROPOSED SYSTEM

Access to applications should be given to authenticated students and professors. After authentication is done, only the applications that are required for the practical could be accessed and others should be blocked .This access and blocking of applications would be according to the timetable of that particular lab schedule, which would be fed in the server computer. This server machine would accordingly adjust the access rights of the computers connected via LAN setup in a particular lab.

4. IV. METHODLOGY

The users will be authenticated by validating details that are stored in the database .After authentication work is done then access is granted according to the user role modules. i) Student Module:-The student would be given access to applications that are required for implementation of the practical session of a particular subject. Other applications would be blocked. If the student requires a certain application

then the teacher module would grant access accordingly ii) Teacher Module:- Teacher has all rights to access all the Software application available on the computer. If in case any student wants another software application other than the allotted application, teacher would grant access by modifying the server iii) Guest Module:- This module would be granted access to all the applications but the session would expire in 30 minutes and again the user would be redirected to the login page and if again access required ,should be requested to teacher module. The below figure 1 is showing workflow of proposed system. The coding language that will be used to implement this work is java (session handling).

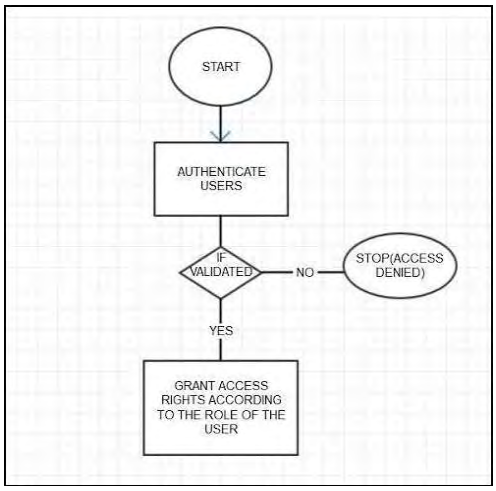


Fig1.-Flow Chart for authentication and access control system

5. V. RESULTS

Student should get access of all needy application with respect to the lab which he is going to perform unnecessary applications must be denied. Teacher should assign necessary other (except assigned) application access to student on his demand if it's relevant.Teacher should continue the authentication and access of all application to guest user as per his demand and necessary once limited session period expires. No application or computer usage to any invalid i.e. non authorization user during lab session. Let us see the execution environment in background with different cases as follow.

Table1.case 1**:** java lab

SUBJECT/ ACCESS	JAVA	MYSQL	MATLAB	VB
STUDENT	yes	no	no	no
TEACHER	yes	yes	yes	yes
GUEST	yes	yes	yes	yes

This access control table lists the access rights of different modules. Here the example taken is the java lab and hence the applications needed to execute java code are accessible to students, others are blocked. Next access table gives the access rights of database lab.

Table2.case 2: Database Lab

SUBJECT/ ACCESS	JAVA	MYSQL	MATLAB	VB
STUDENT	no	yes	no	no
TEACHER	yes	yes	yes	yes
GUEST	yes	yes	yes	yes

***Yes denotes access granted**
***No-access denied**



Fig2. Input Form For Blocked Application's

In the above image, all the applications will be blocked apart from the schedule for a particular lab.

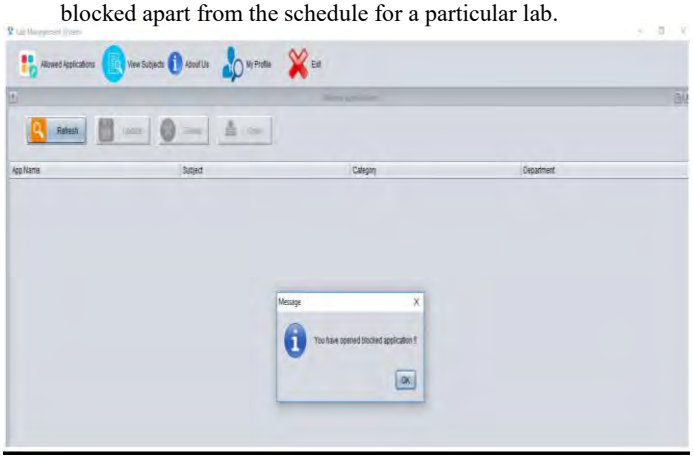


Fig3. Displays Blocked Application Alert Message

In the above image, if a student try to open a blocked application then the student will get a pop up notification that "You have opened blocked application".

6. VI. CONCLUSION

The system will be providing following features: Authenticated users and provide access application. Allocate access rights to different users according to the role i.e. student and teacher would be given different access rights. The access rights granted to students would be set according to the lab subject. If any change in access rights is required that could be modified & allocated by the teacher module. This system will avoid the use of unnecessary applications by the students while performing

his/her task in lab session i.e. Student cannot use any other software other than the application allocated to him.



REFERENCES

[1] G.A Champine"Project Athena as a distributed computer system", IEEE computer, vol. 23, no. 9, pp. 40-51, Sept. 1990.

[2] R.M Needham "Using encryption for authentication in large networks of computers".

[3] Min-A Jeong, Jung-Ja Kim , Yonggwan Won "database security system as a Parallel and Distributed Computing, Applications and Technologies" IEEE computer, 20 October 2003.

[4] Qing-hai Bai, "Cross Strait Quad-Regional Radio Science and Wireless Technology Conference" IEEE computer, Harbin, China,10 October 2011.

[5] Wenrong Zeng,Yuhao Yang, andBo Luo " Big Data (Big Data)," 2014 IEEE International Conference, Washington, DC, USA, 08 January 2015

[6] Jeff Voas, Piero P. Bonissone, Yoshihiro Ohba "IEEE Consumer Communications & Networking Conference" IEEE Communications Society , Las Vegas, USA, 12-15 January 2018.

[7] Mohamed Fathy , Marianne Azer, Moustapha Bahgat, Ayman Yehia "Security access control research trends" Wireless and Mobile Networking Conference (WMNC), 27 June 2013.

Gabor-Canny Algorithm for Under Vehicle Object Detection Application

[1]P.G.Salunke, [2]Dr.S.T.Gandhe, [3]J.S.Soni,
[1]Professor, [2]Principal,[3]PG Scholar
Department of Electronics and Telecommunication Engineering
Sandip Institute of Technology and Research Center, Nashik
Savitribai Phule Pune University, Pune
[1]janhavi.deore14@gmail.com, [2]sanjay.gandhe@sitrc.org, [3]prashant.salunke@sitrc.org

Abstract: In this work a novel strategy is intended to distinguish risky object under the frame of vehicle utilizing Edge discovery procedure. The strategy depends on watchful edge locator and surface examination is finished with the Gabor Filter. The framework is identified with the directed characterization procedure. Object recognition depends on machine learning strategies which require a learning stage. Preparing calculation used to recognize the bomb object from an arrangement of pictures.

Keywords: Object Detection, Image Classification, Under-Vehicle Inspection, Quad copter, Edge Detection Technique, Texture Analysis, Normalized Distance Classifier

I. Introduction:

With national security raised to a noteworthy issue in our lives, the examination on under vehicle review has drawn more consideration than previously. The under case is the most widely recognized zone to put any risky thing like bomb, explosives, medications, weapons or booty things. Since under case is the main place of the vehicle which can't be bolted or anchored and continually being opened. So it is vital to take this marvels in thought. Consequently the need of under vehicle observing framework arrives in a situation. Anyway such kind of framework is these days accessible in market however because of its staggering expense, isn't moderate for everybody. The current framework is manual moreover. In Early framework, Underside of the vehicles was checked by monitor using mirror dish yet this framework experience the protect in threat and is less productive and tedious because of its manual activity. For this situation, proper reconnaissance is unrealistic with this strategy. Another way to deal with pass on this evaluation by using three to four cameras settled under the hard steel and when the vehicle will disregard this course of action at an explicit confined speed, so that underneath of the vehicle can be researched. Nevertheless, this will be costly. So here we go with our examination. This paper gives an exceptionally valuable strategy to recognize the bombs in under undercarriage of the vehicle dependent on watchful edge identification method. The procedure depends on the surface examination with standardized separation classifier. Question discovery depends on machine learning techniques which require a learning stage i.e. preparing and testing of the pictures. The calculation should be prepared to identify the bomb protest from an arrangement of pictures. At that point it tends to be utilized to recognize the protest.

Related Work:

CH. Manoj, N. SaiTeja, Ajay Kumar.R, M. Jasmine Pemeena Priyadarsini, K.Murugesan, Srinivasa Rao Inabathin [1] has proposed a framework in which they have actualized a framework to distinguish the bomb dependent on Quad copter procedure. Quad copter only UAV (U. It is an air vehicle which has different rotors (Around 4 rotors) and is controlled by RC transmitter (Joystick). The bomb is recognized with the assistance of Geiger counter module. The data with respect to the area of bomb is sent to the client by making utilization of GPS and GSM modules. Combination of Geiger module, GPS module, GSM module, Arduino-Uno functions as a payload to the Quad copter. The framework needs to pass the quad copter on the region on which the administrator needs to inspect and on the off chance that any bomb is distinguished, and, Geiger tube recognizes it. At that point instantly bomb area is sent to the administrator's portable by GSM and GPS module. This sort of method decreases the labor and recognition process time contrasted with manual location.

The Study paper [2] has examined about the Under Vehicle Observation Framework which principally engaged to use in stopping zones of open spots, government workplaces, and shopping centers. This paper demonstrates one of the prominent utilization of microcontroller 8051, which clarifies about how it can assess the undercarriage of a vehicle. The task comprises of remote automated vehicle remote camera mounted on it whose video yield is seen on the PC. The remote automated vehicle on which remote camera is mounted can be controlled from the PC utilizing graphical UI application. This automated vehicle can be moved under the vehicle and in this manner the assessment of the underside of that vehicle can be happened.

Chenlu Qiu, Huiying Xu and Weixiang Liu [3] have proposed a framework as Supervision on Irregular Exercises in Vehicle Examination Administration by Abnormality Discovery in Bipartite Chart. Inconsistency location in bipartite diagram details the supervision on unusual exercises in vehicle review stations as a peculiarity recognition issue in weighted bipartite chart. Pertinence scores and typicality scores are registered for enrolment regions and assessment stations. The doubt of an assessment station including anomalous practices is assessed by the dispersion of the ordinariness scores. Exploratory outcomes on genuine datasets are given, demonstrating the viability of the proposed technique. This work plans the supervision on irregular exercises in vehicle assessment stations as an abnormality location issue in weighted bipartite chart.

In the overview paper [4] Sanngoen Wanayuth, Taichi Yamada, Akihisa Ohya and Takashi Tsubouchi have dealt with a mechanized under-vehicle audit in ceasing place using a convenient robot. To get the underside of vehicle, laser go discoverer (LRF) is utilized to anchor information of dissent with facilitate measure under vehicle. highlights of the methodology join the strategy for under vehicle recognition, pose estimation work, data arrangement using iterative nearest point (ICP calculation), underside vehicle data connection with find the contrasts between as of now recorded state and current state. Normal stature and recognized size of the protest were controlled by thing distinguishing proof technique. This procedure is displayed to examine and perceive underside of the vehicle change. Trials are led to exhibit the productivity of framework's way to deal with review and perceive protests under vehicles.

Erick E. Ruiz, K. Larry Head has presented a vehicle review at ports of section [5]. One a player in the vehicle screening process includes traditions and outskirt assurance (CBP) faculty playing out a starter assessment of the underside of irregular vehicles by looking under the vehicle through a mirror mounted on a stick, scanning for abnormalities or remote articles present on the undercarriage structure and parts of a vehicle. On the off chance that any deviances are recognized, the vehicle is coordinated to an optional and more careful investigation. This paper shows a venture that goes for mechanizing this fundamental undercarriage examination by utilizing a programmed under-vehicle review framework (AUVIS) and picture preparing calculations to help staff in recognizing vehicles for auxiliary investigation. The investigation framework will picture the undercarriage of each vehicle in a path as they approach the CBP specialist and utilize a novel change identification calculation to contrast the caught picture with a relating reference picture of the vehicle's make and model. The product will distinguish and feature any contrasts between the two pictures to give a quick and target proposal for optional investigation. Another strategy for change recognition and another technique for producing consistent, high goals, composite pictures of a vehicle's undercarriage have been exhibited.

II. Proposed Work:

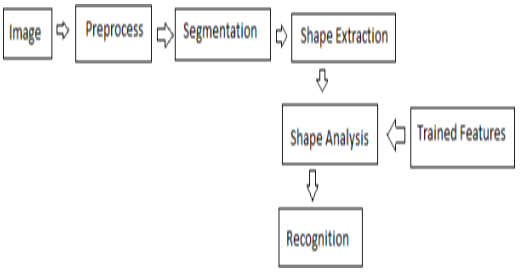


Figure 1 System Block Diagram

Images are manufactured pictures created by the superimposition of closer view and foundation picture.

Preprocessing is the arrangement of tasks connected for the enhancement of nature of picture. This enhancement in nature of picture builds the exactness of further advances associated with handling without losing data.

Image Segmentation is the way toward parceling an advanced picture into various portions i.e. set of pixels.

Image Segmentation is an arrangement of sections that on the whole cover the whole picture, or an arrangement of shapes separated from the picture.

--Various methods for segmentation of image[9]

- Thresholding Method
- Edge based method
- Region based method
- Clustering based method
- Watershed based method

Feature Extraction: Feature implies comparative qualities and Extraction implies precisely recovery of those features. The acknowledgment proportion can get expanded on account of appropriate feature extraction. This phase can be based on following types of features: Global features, Local features and Transition features [5].

Shape Analysis strategies assume an imperative job in frameworks for object acknowledgment, coordinating, registration, and examination. Shape analysis techniques break down the articles in a scene. Shape representation strategies result in a non-numeric portrayal of the first shape (e.g. a diagram) so the essential qualities of the shape are saved.

Prepared features are the feature given to the calculation which is then coordinated to the standard picture.

The calculation depends on the Gabor filterl where vigilant edge finder is utilized to get clear underside and object picture with no additional clamor or additional data. The algorithm goes through two basic phase: Training and testing phase.

- **Steps for training of images:**
  - Read the image
  - Preprocessing done with converting RGB image into gray image.
  - Extract texture using gabor filter.
  - Apply canny to remove extra or unwanted part taken from Gabor filter (These are final features)
  - Now store it as a local variable.
- **Steps for testing of images:**
  - Read the image
  - Preprocessing
  - Apply Gabor filter
  - Scale decomposition carried out of achieved image
  - Apply canny filter
  - Output image compared with dataset of training phase.

IV Result

For bomb object identification application in underside of the vehicle, canny edge recognition procedure is presented alongside the surface examination did by the Gabor filter in this proposed framework. For that the framework diagram is proposed as above. Proposed Software will be utilized for proposed framework is Opencv

Results are shown as below. *Figure 2* is the original image where the bomb object is implanted.*Figure3* gives the result where the canny filter gives the positions of tracked intensity discontinuities and figure 4 gives the Gabor output.

.

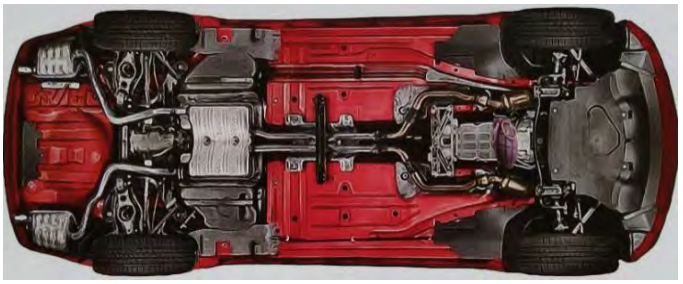


Figure 2: Original Image

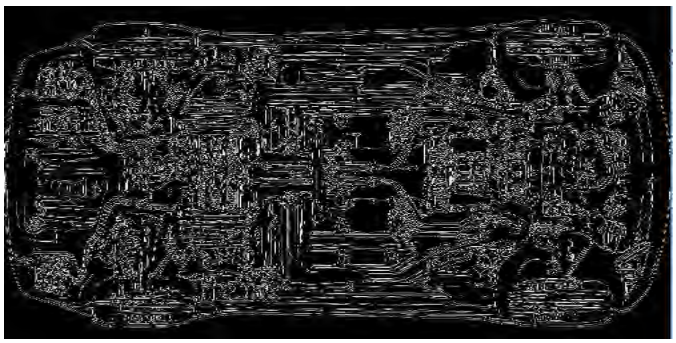


Figure 3: Canny Image

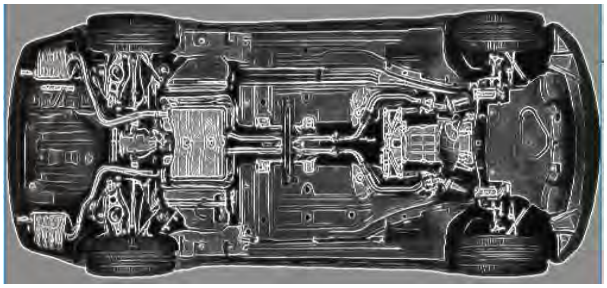


Figure 4: Gabor Image

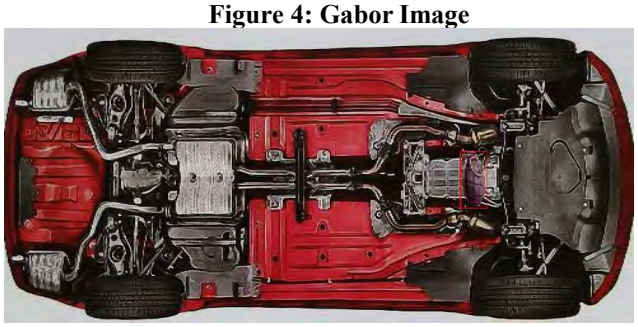


Figure 5: Output image

V. Performance Parameter

True positives (TP) =Number of correct detections.
True negatives (TN) =Number of correct rejections.
False positive (FP) = Number of false detections.
False Negatives (FN) = Number of missed detections.
True Positive Rate (TPR) = TP/ (TP+ FN)
False Positive Rate (FPR) = FA/ (FA+ TN)

VI. CONCLUSION

The above depiction gives thought regarding the novel framework presented here. Results clears the image of proposed calculation that which procedure is utilized for specific. This framework is qualification for security territories and gives 75% preferable outcomes over other under vehicle checking framework. Additionally in future we can build up the programmed number plate acknowledgment framework to the propose framework.

References

[1] CH. Manoj, N. SaiTeja, Ajay Kumar.R, M. Jasmine Pemeena Priyadarsini, K.Murugesan, Srinivasa Rao Inabathin,"Unmanned Aerial Vehicle based Bomb Detection", IEEE 2016.

[2] Sudarshan B. Divekar," Under Vehicle Surveillance System",IJCTA 2014

[3] Chenlu Qiu, Huiying Xu and Weixiang Liu," Supervision on Abnormal Activities in Vehicle Inspection Service by Anomaly Detection in Bipartite Graph",IEEE 2016

[4] Sanngoen Wanayuth, Taichi Yamada, Akihisa Ohya and Takashi Tsubouchi,"Under-Vehicle Inspection Utilizing a Mobile Robot with a LRF Sensor", IEEE 2013

[5] Chaitali Dhaware, K. H. Wanjale,"Survey on Image Classification Methods In Image Pocessing", International Journal of Computer Science Trends and Technology (IJCST)-Volume 4 Issue 3,May – Jun 2016

[6] E. Ruiz, K. Larry Head," Use of an automatic under- vehicle inspection system as a tool to streamline vehicle screening at ports of entry and security checkpoints", 2012 European Intelligence and Security Informatics Conference.

[7] Asmaa Kafuo,Samera Diaf,Aisha Gonifeda,Zamzam Alshatouri."A Literature Survey on Gabor filter and its application",2017 ResearchGate open access

[8] Anjum Sheikh, R. N Mandavgane, D. M. Khatri," REVIEW ON CANNY EDGE DETECTION" , 2014 IJAICT Volume 1, Issue 9, January 2015.

[9] Dilpreet Kaur , Yadwinder Kaur."Various Image Segmentation Techniques:A Review", *IJCSMC, Vol. 3, Issue. 5, May 2014, pg.809 – 814*

Smart Signalling and Emergency Vehicle Management System

Vidya Prashant More
School of Computer Science and Engineering
MIT World Peace University, Pune, India
(vid5089@gmail.com)

Prof. Sharmishtha desai
School of Computer Science and Engineering
MIT World Peace University, Pune, India
(sharmishtha.desai@mitcoe.edu.in)

Abstract — In recent year's traffic surveillance system becomes important as it gives the important information about traffic density, traffic jams, length of the queue and it helps to analyse the causes of traffic conjunction. The aim is to handle the traffic based on the current density of vehicles on road and the main focus is on providing fast and efficient service to the emergency vehicles such as ambulance, fire bridged, government vehicles, etc.

In this paper, the image pre-processing, object detection and analysis techniques are reviewed and analysed to evaluate a more robust and efficient signalling system with more accuracy.

Index Terms—Smart signalling, Object detection, Object analysis, Background subtraction, Template Matching, Emergency vehicle detection, Rule violation.

I.INTRODUCTION

Study of traffic signalling system becomes important because it provides meaningful information and study of some new methodologies for the improvement of performance.

It will provide the information about the current traffic density on road i.e. count of vehicles present on the road as well as it provides the facility to detect the emergency vehicle such as ambulance, fire brigade, and some government vehicles so whenever and wherever the emergency vehicle is detected the signal switching and servicing is performed accordingly.

In the working of the system we are firstly accepting the input videos from the signals view then the pre-processing module performs some pre-processing techniques on it after the pre-processing stage we get the filter input which is completely free from the unwanted and noisy contents therefore, it provides better input for further processing. Once we get the pre-processed video then we apply an algorithm module on it which consist of Background subtraction, Blob detection, Blob analysis, Vehicle count and template matching which provides the results helpful for traffic analysis. It will also help to provide low cost and efficient implementation of a system for better performance in traffic analysis.

II. HISTORY AND OVERVIEW

A. History

All the developing countries leads to the increase in number of vehicles which is somewhere responsible for the traffic conjunction and if we consider our current signalling system it works on the countdown timers which has some statically allocated time for that time the signal remains in the same state because of this system we have to face traffic related problem which can be handled in normal traffic scenarios but it can't be handled in case of emergency vehicles such as ambulance. This system is about providing signal switching based on the current density of vehicles. The system also mainly focuses on the detection of an emergency vehicle and fast and efficient service for that vehicle.

B. Overview

Our existing system has some drawbacks like consider one signalling system which has a timer of 60 sec. and there is a condition in which there is no single vehicle present on the road after the first 10 sec but still, the signal shows the greenway to that road until it completes its 60 sec. which is not feasible traffic solution and at the same time suppose signal shows the red signal to the road where more number of vehicles are present on the road then also signal supposed to be in the same state which leads to the conjunction on that road. To avoid such a problem, we can use the various image pre-processing techniques and algorithms.

III. RELATED WORK

In [1] different image pre-processing techniques suitable for the density based traffic control and emergency vehicle detection are explained.

This paper represents vehicle counter-classifier which is based on various video processing methods such as object detection, edge detection, frame differentiation, and the Kalman filter. It improves the accuracy of the vehicle classification.

This paper mainly focuses on software-based vehicle detection techniques. In this process, it first identifies the vehicle in the video then assigned each vehicle with some classifiers based on some features in particular vehicle type then count all the detected vehicles.

Background information:
It consists of :

1. Video processing: In the computers to achieve the goal of video analysis.
2. RGB to Grayscale conversion: It is necessary for efficient processing.
In video processing technique the capturing rate of a video should be changed from RGB to 0 to 255 gray level.
3. Power-low transformation: enhanced image can provide better contrast and detailed images as compare to non-enhances one.
There are various methods available for performing this transformation such as linear method and logarithmic method.
4.Soble edge detection: For object detection, we require image matching function and edge detection. Edges are the points in the image where the brightness changes suddenly with some amount. Soble edge detection is a very easy method.
5. The Kalman filter: The unwanted part and noise should be removed from the image using some filtration. It provides an improved estimate based on series and noisy elements.
6. Detection zone: A region should be defined in which the moving objects edges are defined. This zone is in the middle of the screen and has some threshold values based in term of height and width to avoid perspective challenges and wrong vehicle counting.
7. Counting and classification function

Vehicles are typically classified as:
Type1: bicycle, motorcycle
Type 2: motorcars
Type 3: pickups and minibuses
Type 4: buses, trucks, trailers.

It is important to have knowledge about height and width of the vehicle to determine it belongs to which vehicle type. It just count the moving objects. The type of vehicle is also helpful for the analysis of traffic.

Vehicle detection tracking process can be performed using some methods such as:

1. Matching
2. Threshold and segmentation
3. Point detection
4. Edge detection
5. Frame differentiation

This paper uses a combination of both the tech such as ‘frame differentiation’ and ‘edge detection’ to provide better quality and accuracy. Use of Kalman filter correctly tracks and estimate the position of each vehicle. It is also used to classify detected vehicles into particular groups and count them separately and results in some useful information for traffic flow analysis.

Experimental results and conclusion [1]:

AVI format and 29 frames per second frame rate with resolution 960*540 and it also consist 24 passing vehicles have been collected and real-condition test has performed on it.
This method reduces the cost and provides better quality and accuracy it also classifies the vehicles using some image processing and Kalman filter algorithm. It provides 95% accuracy for classification [1].

In [2] paper it consists of the intelligent vehicle counting method which is based on blob analysis it is composed of moving object segmentation, blob analysis, and tracking. It results in meaningful information for traffic surveillance. It performs greying, binarization, denoising and moving target detection.

The common approach is the use of background subtraction which identifies the moving objects. To reduce the implementation cost a new moving object detection algorithm is used which is known as Boundary block detection algorithm. It consists of Image acquisition, Image enhancement, image segmentation, image analysis, and counting.
It has a methodology to perform background subtraction after performing background subtraction then it applies image binarization and then performs object detection. For object detection it first extracts the background and then tracks the moving objects.

Boundary block detection algorithm: here we detect the blocks of moving objects by using the information of motion vector field which is generated by the video compression pixels in the same object should satisfy the condition of interpolation.

Flow:

1. Reading of frame from a video sequence
2. Background reference image
3. Normalization of image
4. Background subtraction
5. Graying and binarization
6. Boundary detection
7. Finding objects
8. Counting vehicles

In this paper the efficient object detection algorithm is used. Background reference image tech is used to construct the background image as well as various image pre-processing techniques are used which results in the accurate information [2].

In [3] the processing methods to provide reliable and robust traffic surveillance system are explained.

It uses the algorithm which consists of five elements such as Background subtraction, Blob detection, Blob analysis, Blob tracing, and Vehicle counting. Here meaningful features are extracted then by using these features moving objects are tracked.

Generally, the traffic surveillance system requires more sensors which include a push button, loop detectors, radar sensors etc.

The evaluation of the traffic conditions can be represented using some parameters such as traffic flow rate, average traffic speed, the length of the queue.

This paper describe the computer vision system to count the vehicles moving on the road.

It uses approaches such as:

1. Background subtraction
2. Blob detection
3. Blob analysis
4. Blob tracking
5. Vehicle Counting

Here the first focus is on statistics of the background scenes such as road, trees, buildings, etc. then we classify the objects i.e. (vehicles) from the surroundings when we capture the foreground objects then they are grouped together by using neighbourhood analysis and the filtering process is applied on it to remove the noise and misdetection.

Background subtraction:

In our case, the position of the camera is fixed above the road, therefore, the background captured is always static in nature. This algorithm is robust and it has high detection performance and computational requirements than other algorithms. It first builds a statistical model of video image background for every pixel based on background subtraction model. Then it collects the colour co-occurrences in the video and applies the quantization to decrease the computational load. The colour co-occurrences improve the accuracy while capturing the moving objects. Bayes decision rules are then implemented for the classification of the foreground and background image. The important thing for the algorithm is it should adopt the background model to the changing conditions in the scene. After classification new features are captured and used to update the model. Therefore the sudden, as well as gradual changes in the scene, are noticed and use to update the statistics. Then the algorithm maintains the resulted images as a referenced background image.

Blob Detection:

The background subtraction model provides the pixel as foreground. The blob detection model groups these pixels in the current frame by using the counter detection algorithm. It groups individual pixel in different classes and then finds the counter surrounding to it. These classes are known as candidate blob (CB). Then the algorithm checks the size of each CB and removes the CB's which are small in size to reduce the wrong detection.

Blob Analysis:

It accepts the CB's with their size as an input and provides the new blob in the current video frame. This gives which CB belongs to which vehicle and these correspondences are computed by positions. For this it uses the k-Mean clustering. The output of this algorithm is the blobs that identify the vehicles in each frame.

Blob tracking:

It tracks each blob within successive image frame after performing blob analysis each blob is extracted with its bounding boxes and centroids then Euclidean distance method is used to get the distance between centroids and area of a vehicle is also consider for tracking.

Vehicle Counting:

This counts and registers the vehicle flow for each lane. For counting all the passing vehicles, it uses a method like the moving vehicle is counted while it passes the baseline. The system will also calculate the velocity of a vehicle.

In this paper, the virtual loop based system is used to detect the vehicles. Long-term test is performed on actual traffic so that the system shows that the system is reliable to estimate real-time traffic flow [3].

In [4] the drawbacks of existing countdown based static signalling system and the importance of density based vehicle counting are explained.

The main purpose of this paper is to study the effectiveness of countdown timers. The opinion of the countdown system is not consistent rather it is contradictory. Our existing signalling system is based on countdown timers and its working efficiently but doesn't mean that it is a perfect solution. They cause so many financial and social losses and also important aspects is of intersections with traffic lights relates to risks associated with the requirement of taking rapid decisions while the signal changes from one state to other. In the situation of dilemma zone driver supposed to take the correct decision. It is also not possible to stop safely before the stop line. The range of this dilemma zone changes depending on various factors like vehicle speed, delay, and acceleration, road conditions, a reaction time of the driver.

Some study results that collisions and accidents at the intersections are about 30% and it happens because of the decision-making problems related to the dilemma zone of drivers. It is more occurring in case of large vehicles. Sometimes the driver has lack of information about the sudden changes in transmission of signals.
Conclusion: percentage of stopping vehicles after yellow dropped from 39.1% to 19.6% and crossing the red light increased from 7.2% to 13.5%. For the countdown case, the drivers cross the stop line during red signal is 24 cycles out of 88 and for non-countdown timers, it is 22 out of 90[4].

In[5] The usefulness of the edge detection method is explain for the system to control traffic by measuring the real-time vehicle density which result in efficient response time, vehicle management, automation over an existing system edge detection is very useful for extracting the required traffic information from the video. Canny edge detection algorithm provides better and improved results.

The criteria for edge detection always includes:

1 Detection of edges should have occurred with low error rate means all the possible edges should be cached in the detection
2. Detected edge points should be accurately localizing at the centre of the edges
3. Edges in the image should be marked once and image noise should not create the false image [5].

In [6] paper explains that the moving vehicle detection is one of the important task in (ITS).

This paper presents the smart counting method based on the blob analysis.

The algorithm consists of moving object segmentation, blob analysis, and tracking.
By performing the analysis on the blob important features are extracted from the image and also the speed of vehicle and vehicle flow in particular defined area is calculated. It also uses the forward boundary block detection with noise reduction for detecting the moving vehicle [6].

[7] Consist of the dynamic system that solves all the drawbacks of the current system.

The system uses the camera fixed at signal position performs image pre-processing on it then compute the volume of real-time traffic and sets the timer accordingly. It consists of methods such as image acquisition, edge detection, and image analysis and density calculation.

The system provides a real-time solution to the critical traffic conjunction problem [7].

[8] Shows the importance of fast template matching. If we consider our case the main objective of the system is to provide quick and reliable service to emergency vehicles as soon as the emergency vehicle is detected. For the detection of these vehicles template matching algorithm is very useful where we provide the features that we want to extract from the source image.

To improve the performance of this algorithm this paper uses the matching based on the K-gray-degree clustering in which the given template is divided into the blocks of a particular size. In all the process of matching it only checks whether the corresponding blocks between the template image and sub-image are in the same degree. The experimental results show that the algorithm is robust, and fast under the condition of non-linear shapes, complex background etc. [8].

The anomaly is nothing but the something which has different behaviour which makes it different from others. This paper consists of the unsupervised clustering technique (Novel anomaly detection density-based spatial clustering of application with noise. This is mainly used to identify the abnormal behaviour of the object.

In traffic signalling case we can consider a situation where the signal is in a red state so ideally, all the vehicles supposed to stop there for that time

but if any vehicles crossing the signal is considered as the anomaly because it doesn't follow the expected behaviour.

This kind of algorithm can be very useful for the detection of the vehicles that break the rules and useful to control rule violation [9].

IV. CONCLUSION

In this study signalling system and emergency, vehicle tracking is reviewed. We have studied the different approaches for object detection and tracking. Various image pre-processing techniques are studied for efficient and noise free output for further processing also the template matching algorithm is studied for the detection of an emergency vehicle.

V. REFERENCES

[1] Ali Tourani, Asadollah Shahbahrami" Vehicle Counting Method Based on Digital Image. Processing Algorithms" 2015 2nd International Conference on Pattern Recognition and Image Analysis (IPRIA 2015) March 11-12, 2015.

[2] Yoginee B. Bramhe (Pethe), P.S. Kulkarni "An implementation of Moving Object Detection, Tracking and Counting Objects for Traffic Surveillance System" 2011 International Conference on Computational Intelligence and Communication Systems.

[3] G. Salvi," An Automated Vehicle Counting System Based on Blob Analysis for Traffic Surveillance".

[4] Milaszewicz Barbara," The issue of using countdown timers at intersections with traffic lights: a literature review" 978-1-5386-4578-9/18/$31.00 ©2018 European Union.

[5] Taqi Tahmid and Eklas Hossain Khulna University of Engineering & Technology, Dept. of EEE, Khulna" Density Based Smart Traffic Control System Using Canny Edge Detection Algorithm for Congregating Traffic Information" 2017 3rd International Conference on Electrical Information and Communication Technology (EICT), 7-9 December 2017, Khulna, Bangladesh.

[6] Pranav Maheshwari, Deepanshu Suneja, Praneet Singh, Yogeshwar Mutneja, "Smart Traffic Optimization Using Image Processing" 978-1-4673-6747 -9/15/$31.00 ©20 15 IEEE.

[7] Sha Sha, Chen Jianer, Luo Sanding, "Fast matching algorithm based on K-Degree Template" 978-1-4244-3521-0/09/$25.00 ©2009 IEEE.

[8] R. Ranjith, J. Joshan Athanesious, V. Vaidehi "Anomaly Detection using DBSCAN Clustering technique for Traffic Video Surveillance".

Blockchain Based Automated Trading Platform

Neil Naidu
UG Student
CMPN, TCET
Mumbai
neilvarunnaidu@gmail.com

Sidddhesh Bhatgaonkar
UG Student
CMPN, TCET
Mumbai
bhatgaonkar.siddhesh@gmail.com

Abhishek Delwadia
UG Student
CMPN, TCET
Mumbai
abhhishekdelwadia@gmail.com

Ms. Megharani Patil
Lecturer
CMPN, TCET
Mumbai
@gmail.com

I. Abstract

There has been a complete change in the way a person prefers to shop for even the most basic of things. Consumers now prefer to place their orders online to purchase goods/services and ordered goods are dispatched directly from regional warehouses (or distribution/fulfillment centers) to a customer's doorstep.Future warehouses might be designed and implemented as centralised distribution centers because it is the most cost-effective way to run businesses for any large manufacturers, suppliers or retailers. Retailers are trying to find better methods for storage, pick-up and dispatching goods in distribution centers; hence, a good design of a trading platform is required.

Our objective is to design an NFC/RFID-based management system and mobile client as part of a proposed future generation warehouse software system through application and integration of fast-growing IT technologies (radio frequency identification/RFID or near field communication/NFC tags, wireless sensors and communication networks, an Android client, a PWA client, an RPA bot for back end operations and optimisation using machine learning algorithms) to maximize the warehouse capacity and have real-time visibility of inventory control and data handling management.

The differentiating factor is the Blockchain Ledger being used, as well as the PWA front end. The technologies being used here have never before been combined on a platform. The PWA front end is being built using guidelines from Google regarding its PWA specifications for Chrome. The Blockchain Ledger is being built as a singular data structure without the use of mining and consensus algorithms that might have driven up complexity with no discernable gain. UiPath is being used to automate and autonomously manage ledger operations.

II. Keywords

Inventory, RFID,PWA,Warehouses, UiPath, NFC, trading, warehousing, automation, RPA, Automated, Process automation, Robotic Process Automation.

1. Introduction

Over the last 2 decades, India's retail sector has witnessed a tectonic shift. From an unorganized market, primarily ruled by Kirana shops, it now boasts of large multi format stores and superstores that offer global experiences to local consumers. Tier 1 cities(metro cities) have been the first to benefit from the boom of organized retail. Rising interest of reputed real estate developers, growth of premium high streets and entry of foreign brands have made the marketplace exciting for consumers. And this trend has gradually permeated across Tier-II & III cities on account of the rising demand from an evolving consumer base. This consumer is also demanding a high quality shopping experience, similar to his metro counterpart. With increased land prices and high lease rentals in big cities, retailers too are seeking options beyond the metros. Popular brands are looking at Tier II cities for expansion. Due to rising urban population, increasing per capita income and the success of brand expansions in Tier II cities, this transformation will continue to be seen in the coming years.

There has been a complete change in the way a person prefers to shop for even the most basic things. Customers prefer to place their orders online to purchase goods and ordered goods are dispatched directly from regional warehouses (or distribution centers) to customers' doorsteps .Future warehouses might be designed and implemented as centralized distribution centers as this is the most cost-effective way to run businesses for many large manufacturers, suppliers or retailers. Retailers are trying to find better methods for storing, picking and dispatching goods in distribution centers; hence, a good design of warehousing systems is desired. Our objective is to design an RFID/NFC-based management system and mobile client as part of a proposed future generation warehouse software system through application and integration of fast-growing IT technologies (radio frequency identification/RFID/NFC tags, wireless sensors and communication networks, an Android client, Machine learning algorithms) to maximize the warehouse capacity and have real-time visibility of inventory control and data handling management.

2. Theory

The Authoritatively mandating Person submits the goods acceptance order document to the Contractor in an electronic form only, directly from his/her IT system, at least 1 working day afore the scheduled distribution. Every goods acceptance order shall contain:

- Unequivocal description and quantity of the goods to be accepted;
- Unequivocal description of the distribution method (own denotes of convey, courier or postal consignment etc.), in accordance with the distribution forms concurred on by the Parties;
- Unequivocally verbalized distribution date;
- Adscititious distribution markings (distribution number, shipping note number etc.);
- Adscititious distribution markings sanctioning the Contractor to identify the distribution facilely (distribution symbol, deliverer's name etc., conveyance number).
- The deliverer (courier, driver) shall have a printed-out distribution label (on the parcel) or ken its number.

After reviewing the distribution note, the Contractor substantiates (accepting the distribution note) the possibility of unloading the distribution on the betokened date or states the most proximate possible date of unloading the distribution.

he Contractor may reluct to attest the distribution note if due to reasons beyond his control on the suggested distribution date it is technically infeasible to receive a distribution (arbitrary events, malfunctions etc.), if on the suggested date there is already another distribution scheduled and notified of at that time, or if receiving the distribution on a given date needs adscititious arrangements made with the Authoritatively Mandating Person such as cognate to reorganising operations at the warehouse or lowering the warehouse efficiency due to accepting the distribution at the designated time. In such an event, the Contractor together with the Injuctively authorizing Person will accede on the most propitious solution which will enable the receipt of distribution on the suggested date or on the most proximate day possible.

The Injuctively Authorizing Person and the Contractor acknowledge that each electronic acceptance order accepted by the Contractor has been submitted by a person sanctioned by the Authoritatively mandating Person to engender such orders and may be processed without any supplemental accedences, predicated solely on the content of the order.LlIff the Contractor relucts to receive the distribution for reasons directly attributable to the carrier (poor technical condition of the conveyance or other circumstances hazardous to the Contractor's employees), then photographic documentation of the factors that avert the

unloading and receipt of distribution will be engendered and submitted to the Injuctively Authorizing Person.

During the acceptance of distribution into stock, the Contractor will verify the correctness of the goods against the distribution note present in the WMS system. Verification consists in counting the quantities of distributed goods

If the Contractor finds any discrepancies between the genuine condition and the distribution note, the Contractor will immediately apprise the Injuctively authorizing Person about such a situation via an electronic mail sent to the designated electronic mail address or by entering the germane data into the WMS system.

In the event of finding any damage to the distributed products, the Contractor will every time notify the Injuctively Authorizing Person by marking the distribution order in the WMS system and will engender photographic documentation of the damage to sanction the Injuctively authorizing Person to make a claim.

3. Research/Teachnical Project work

- The primary purpose of this project is to resolve the problems plaguing the (largely unorganised) small-to-medium sized retail and wholesale business sector.
- Today, India is the second largest consumer market in the world and the seventh largest retail market worldwide.
- India is the fastest growing (large) economy as of 2018 but with a heavy dependence on the service sector.
- As such, retail businesses have the potential to drive further economic growth as well as weaning us off of our IT and Service Sectors Dependence.
- For this, we need to resolve the following major problems.

1. Business Environment:
   - As with many other avenues of the modern Indian ecosystem, the retail sector too is fraught with nepotism and a varying degree of corruption with a desperate need of a tool of regularisation.

2. Unnecessary Overhead (Organising the Unorganised Sector)
   - Precious time and money is wasted in bargaining, bribes, fighting undependable delivery mechanisms all the while dealing with irate customers.
   - Records maintained by individuals, if any, are not dependable or verifiable.
   - With defined, suitable infrastructure tailor made for Indian retailers, we enable the largely unorganised Indian retail sector the opportunity to regularise relatively painlessly.
   - The platform provides accountability at all requisite levels;
     - collection of government revenues is made easier and
     - auditing of suspicious business practices is made easier

We reviewed various existing system and found out various gaps in the previous systems and they are as follows: -

1. Problem of on-time cancellation of orders.
2. Problem of proper management of orders.
3. Problem of proper management of pending orders.
4. Problem of implementing better strategies in case of delay.
5. Problem of customer's demands not fulfilled.
6. Problem of recall of delivered order.
7. Problem of generating automated invoice.

4. Results

The customer should be able to do more for themselves at a time of their choice and through a medium of their choice. There has to be a decrease in calls that go unanswered or end in an unsatisfactory conclusion. There will be a greater emphasis on joined up public services at front end delivery points and an emphasis on a consistent response. The Council will be delivering a modern customer experience that is comparable to private sector benchmarks.

Same-Day Priority Distribution

Retail or online companies can now employ products that enable you to prioritize distribution schedules predicated on customer requisites. Recent innovations in automation have given companies the option to even make one or two-hour distributions. All this is born from a desire to accommodate the customer with immediate and lasting solutions to their everyday requisites. With the difficulty and complexity of processing, the companies can shift the instances of shopping from infrequent to frequent. Conceptually, if an order is distributed in an hour; customer might want to the same on the next day for something else he/she requires. The conception is to satisfy the impulse buy principle of consumer goods sales indirectly by keeping the supply chain replenished.

Authentic-time Distribution Tracking

Last mile distribution has evolved to the caliber that all stakeholders of a shipment get authentic-time updates about the kineticism of the verbally expressed shipment. The customer and the store manager can, concurrently, track the en-route distribution as it reaches the customer. This integrates to the control the store manager (or the logistics management company) has on the last mile distributions. It withal integrates to the contentment level of the customer, kenning when, precisely, the packaged would be distributed. The next phase of this tracking is the electronic proof of distribution (ePoD). Through ePoD, invoicing has become a hassle-free and transparent affair.

Dynamic Distribution Re-Routing

Retail and e-commerce companies homogeneous have faced the trouble of en-route rerouting of distributions due to last minute destination address changes (or preferred distribution time changes) done by the customer. Companies can now re-route their workforce as anon as such an update appears. There would not be any downtime or delays in any distributions due to this shift in schedule. Such dynamic and authentic-time schedule management has become the corner stone of an ever-evolving retail and e-commerce sector.

5.Conclusion

We have conducted research into the need of UiPath and the distinction between Artificial Neural Networks(AI based bots) and RPA; along with their distinctive requirements, advantages and disadvantages.

Using RPA as the core of our back end base we intend to make a completely autonomous platform that is both dependable and independent. We have probed the present use cases where a PWA is deployed in place of a native app (eg. *m.bookmyshow.com*):

1. To be usable on a variety of devices
2. To be usable with satisfactory performance on mid-

to low-tier mobile devicesRPA use case can be to automate the back end machination of blockchain management and server side operations.After consultation with my group members, we reached the conclusion that we need to expand on this idea and patent it as it is a commercially viable platform.

We have successfully identified the sectors for improvement and optimisation regarding our client's operations. We intend to use all underlying technologies to create a single, seamless user experience.

6. Future Scope

There are 2 current approaches intelligent warehouse management. Amazon uses fully automated warehouses while local organisations like Flipkart use manual warehouse management.In a market like India we use human employees as well as some degree of automation to maintain a balance between efficiency and minimization of costs.This also has an added benefit of providing employment to locals, benefit in their economy as well as incentivizing local governments to provide tax benefits.

7. References

[1] Q. Wang, R. McIntosh, and M. Brain, "A new-generation automated warehousing capability," InternationalJournal of Computer Integrated Manufacturing, vol. 23, no.6, pp. 565-573, 2010.

[2] M. K. Lim, W. Bahr, and S. C. H. Leung, "RFID in the warehouse: a literature analysis (1995-2010) of its applications, benefits, challenges and future trends," International Journal of Production Economics, vol. 145, no.1, pp. 409-430, 2013.

[3] H. Wang, S. Chen, and Y. Xie, "An RFID-based digital warehouse management system in the tobacco industry: a case study," International Journal of Production Research, vol. 48, no. 9, pp. 2513-2548, 2010.

[4] J. S. L. Ting and A. H. C. Tsang, "Design of an RFID-based inventory control and management system: A case study," The West Indian Journal of Engineering, vol. 34, no.1-2, pp. 70-79, 2012.

[5] A. D. Ross, D. Twede, R. H. Clarke, and R. H. Ryan, "A framework for developing implementation strategies for a radio frequency identification (RFID) system in a distribution centre environment," Journal of Business
Logistics, vol. 30, no. 1, pp. 157-183, 2009.

[6] C. M. Liu and L. S. Chen, "Applications of RFID Technology for improving production efficiency in an integrated-circuit packaging house," International Journal of Production Research, vol. 47, no. 8, pp. 2203-2216, 2009.

[7] G. Gaukler, R. Seifert, "Applications of RFID in Supply Chains" in Trends in Supply Chain Design and Management: Technologies and Methodologies, New York:Springer, 2007.

[8]https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/using-blockchain-to-improve-data-management-in-the-public-sector(June,19,2018)

[9]https://www.martinstellnberger.co/document-certification-through-the-blockchain/(July,12,2018)

Efficient System to Prevent Session Hijacking Attack on Servers Using Proxy Server

Anuradha Maurya
Computer Engineering
Thakur College of Engineering and Technology
Kandivali,Mumbai,India
anuradha.maurya102@gmail.com

Veena Kulkarni
Computer Engineering
Thakur College of Engineering and Technology
Kandivali,Mumbai,India
veena.kulkarni@thakureducation.org

***Abstract*— HTTP is a stateless protocol therefore we use cookie to maintain session. For Maintaining a session server creates a unique identifier to remember on going session for particular client. This unique identifier is nothing but a randomly generated text which is stored on client browser in form of cookie. These cookies are generally created in the login process, after successful authentication process, server generates unique id for requested client and send it to client. Browser sends this cookie on every request where authentication of user is required. Authentication of cookie become temporary replacement of user password authentication for the entire session.**

Cookies are static in nature, they do not change in the session life time, because of this nature anyone can steal and use this cookie for their benefit. Use of cookies introduces a number of risks in security especially in session authentication. For secure communication use of HTTPS is not easy for those applications that are highly distributed due to performance and financial issues and HTTPS provide security at network only there are many ways were attacker can steal cookie by using different attack like cross site scripting attack, cross site tracing attack, domain related attack etc.

Hence, creating a system using Concept of one-time Dynamic cookie (OTDC) for authentication instead of HTTPS. OTDC and Sanitization will prevent various attacks on servers. A reverse proxy server with OTDC, IP Address, session ID and browser fingerprinting are used to prevent adversary from capturing session credentials. In this system doing sanitization on Reverse proxy server to prevent SQL Injection attack on main server. Setting up HTTPS to HTTP Reverse Proxy Server, Session time out implementation on Proxy server because "The less time you give your account to be cracked, the better for you." Generating log on attack detection and Reporting to Administrator.

Keywords—prevention of session hijacking, cookies, reverse proxy server, one-time cookie (OTC), one-time dynamic cookie (OTDC).

applications security, web application transactions which of those handling sensitive data content, E-commerce transactions, E-business related operations.

II. LITERATURE SURVEY

Use There are so many scientists are designing and developing different detection tools for different attacks and different prevention techniques for remediation of session hijacking. Some of the Detection Tools and Techniques for Session hijacking [2] are Arp-ON, ARP-PING, ANTI-SNIFF, Cookie Monster, Wavelet Based Detection, Cisco Intrusion Detection System (IDS), and Sans Intrusion Prevention System (IPS) are shown in the Table I.

TABLE I. SESSION HIJACKING DETECTION TOOLS

Tool name	**Auth or & year**	**Tool function**	**Type of attack to avoid**
Arp-ON	Darknet, 2000	To secure the Address resolution Protocol	MITM (Man-In-The- Middle) Attacks
ARP-PING	Beyond-Security, 1998	Allows the user to ping MAC address Directly	To detect Sniffer on the network
Anti-sniff	Storm, 2011	Used for packet capturing	Packet sniffer
Cookie-Monster	Pauli, Engebrets on, Ham &Zautke, 2011	Analyzing the strength of the cookie by Archiving & analyzing	Cookie stealing
Wavelet-based-detection	Long &sikdar, 2008	Analysis the signal strength using wavelet Transform	Session hijacking

I. Introduction

Use of web application is increasing continuously and session vulnerabilities are common in web applications. Attacker can take benefits of poorly configured websites or web portals he/she can hijack victim's session and take control over their session identity. In Session management session id plays an important role for maintaining a session. There are different types of session such as active session, passive session and hybrid session. Whatever be type of session, attacker can do session hijacking. The prevention mechanism on session hijacking is very important for web

There are some algorithms and Techniques for preventing Session Hijacking [2], they are

•Ensuring a secure Cookie generation

•Implementing the CIA technology, C-Confidentiality, I-Integrity, A-Authentication.Use encrypted connections or protocols for the transmission of data such as HTTPS, Open SSH protocol suite.

•Locking a particular session to its corresponding user's i.e. Session Lock [5].

A. Ensuring a Secure Cookie Generation

Cookie is the one important component in web-based applications, which can be used to authenticate a client and cookie information is sent to the server system for that user. The cookie information contains user name, password, timestamp, session timeout. Whenever a session is hijacked, the attacker analyses this cookie information and comprises the user's account.

B. Implementing the CIA Technology

This is the low-cost solution to secure session cookies. This also prevents cross-site-scripting. This is achieved by creating a java script on the server side and cookie is available only to the client's browser. It develops the standard plug-in for the Chrome, Mozilla Firefox, etc.

C. Encrypted Connections for The Transmission

The connection was encrypted with HTTPS (Secure Hyper Text Transfer Protocol) and this protocol uses an SSL (Secure Socket Layer) protocol stack. The drawback is cost. Implementing and maintaining this service is cost effective.

D. Locking a Particular Session

This is an effective technique to prevent session hijacking. Even if the session is hijacked, the attacker has no use. A unique HMAC (Hashed Message Authentication Code) algorithm with secret key is shared between client and server. A token is generated from an initial login over the SSL. This is an efficient and low-cost solution to secure the session.

There are different research papers found in literature survey, different session hijacking prevention techniques mentioned in Table 2.

TABLE II. Existing Session Hijacking Prevention Techniques

Author	Prevention Technique	Summary	Gap Identified
Nikiforakis	Session shield	Protect from Xss	Limited to browser protection
Dacosta	OTC (one-time cookie)	Light weight, disposable good performance	Always need to be configured with proxy server and long session live with no session time out and SQL Injection Attack can happen
Asif & Tripathi	Double authentication to user id	Double factor authentication system	Venerable to many threats
Amerah & Mostafa	One-way hash chain (OHC)	effective protection against hijacking	Need high computation storage capacities, Not working on mobile devices
Unknown	Dynamic ID generation Using Session Id and Client Time using MD5	Prevent from session Fixation	From MD5 hash easy to convert and lack of security no longer can be used for Secure system
Annies Minu Sathiyaselan, Vincy Joseph, Anuradha Srinivasaraghavan	A proposed system for preventing session hijacking with modified one-time cookies	Protect from MITM	Just Proposed System, Not implemented yet
Karis D'silva, J Vanajakshi; K N Manjunath; Srikanth Prabhu	Hashing of session Id, browser name, Browser Platform, Browser major and minor version	Dynamic hash ID generation on each time when session is created	On registration time it takes unique finger printing and use the same hash value so MITM attack can happen and hash digest are vulnerable

III. Problem Definition

To Implement a system that will protect web application on server level were no need to implement additional functionality on client side and server side. Proxy server will provide security from session hijacking attack. Design a system that will detect & prevent session hijacking using the concept of OTDC (One-Time-Dynamic-Cookie) with

Reverse Proxy Server. Proposed system will protect main Server from SQL Injection Attack, this will do Sanitization on Reverse Proxy server before it goes to the main server. Setting up HTTPS protocol environment to the HTTP Reverse Proxy server to improve application performance and Implementing Session time out on Proxy server. On detection of malicious activity system will generates log and at the same time it will Report to the Administrative about the malicious activities. For client authentication, system will use Browser Fingerprinting concept which will include client Ip address, client system time and browser information.

Contribution of this thesis will be to make web application more secure and protected from attackers by using One-time Dynamic Cookie and browser fingerprinting Concept, improve robustness of web applications, also doing sanitization for SQL Injection queries which will provide security against SQL Injection attack and real time report updates to administration. Proposed system will protect from various types of session hijacking attack and SQL Injection attack and will provide real time updates about ongoing activities on server to administration of the server.

IV. Methodology

This section will describe the methods used to investigate the work research questions and the reasons for choices made. Here will also take a look at different challenges and the limitations regarding these methods. The main goal of the research to study how to prevent session hijacking attack for web application using reverse proxy server.

In this research, will learn about how the attacker do session hijacking attack and finding out the prevention solution for it, to provide a safe and secure way to guard server by them self at server level using proxy server. This kind of network layout, with one proxy/reverse proxy acting as middle ware between end-machines, is common in organization work-groups and which makes this method appropriate for insider session hijacking prevention.

The below mentioned flow chart diagram is shown in figure 4-1 showing the complete flow of the proposed system which will prevent from different session hijacking attack this system will prevent from man in middle attack, SQL injection attack, will do session fixation attack and auto session time out implementation on proxy server, log generation on malicious activity and real time reporting to the server administration about these ongoing malicious activity.

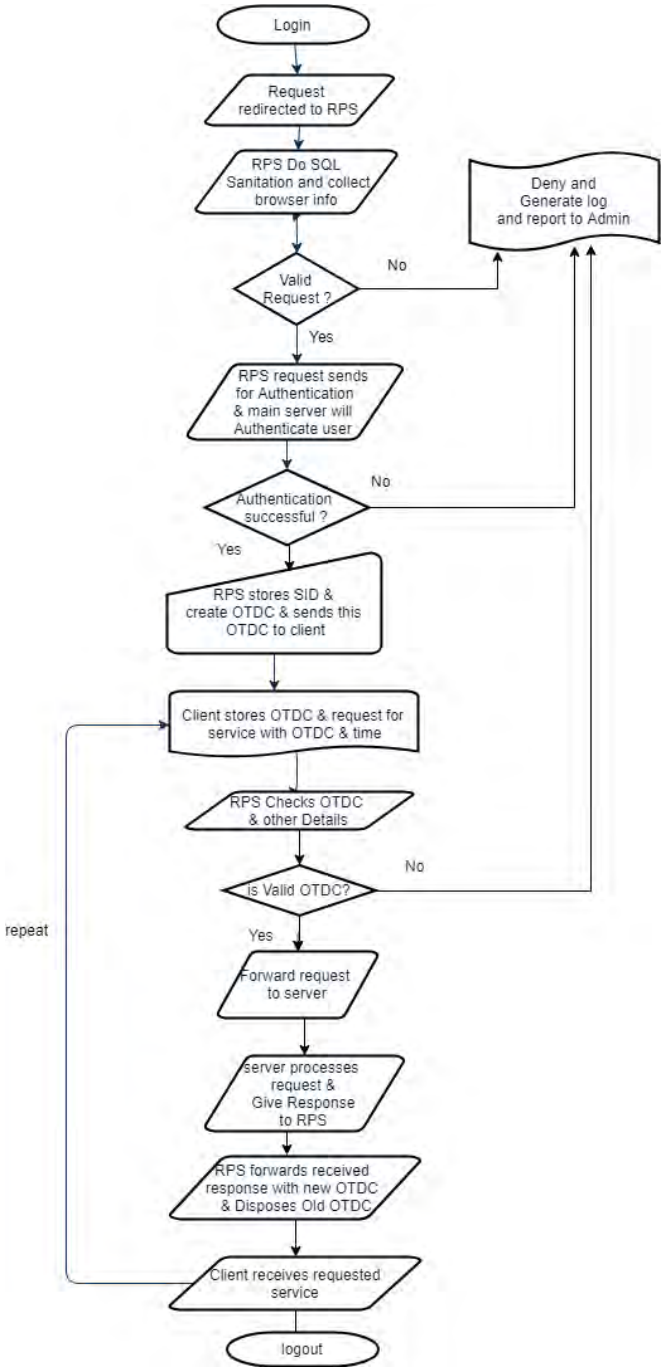


Fig. 1. Flow Chart of Proposed System

V. System Architecture

The session hijacking preventer system will consist of three major components, these components are mentioned below:
1)Client
2)Reverse Proxy Server
3)Main Server

Client: Client or user of the Web Application who initiates a request for resource to server. Consider user want to access some page, he/she will send a request using the login credential to the server. Server will authenticate the user, after successful authentication, user will be given an OTDC by which user will be authenticated for every new request. Every new request will be having an OTDC which will be sent along with the request.

Reverse Proxy Server: Proxy Server is nothing but a computer that acts as intermediary between client and server. Here we are using RPS at server side. Thus, every request from client has to be pass via RPS. The RPS will get IP Address, browser finger print, Set OTDC, session ID and Time for every request, if one of these parameters would change then RPS will redirect to another web page.

Server: This the Actual Main server where request will be processed, server authenticate user and process all the request and give response to all client request.

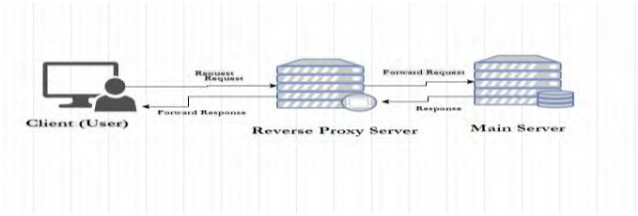


Fig. 2. System Architecture

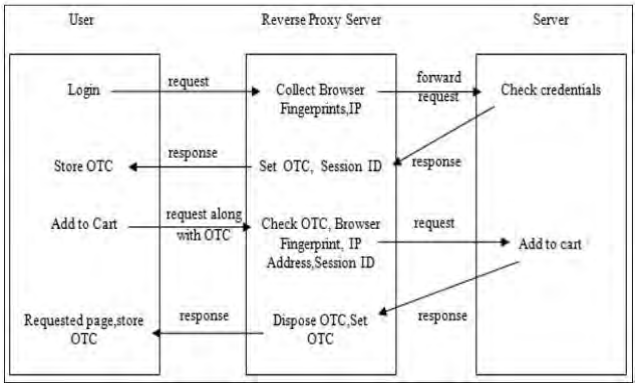


Fig. 3. Block Diagram of Existing System

In above figure 3 shows the block diagram of existing system in which OTC (One-Time-Cookie) concept has been used for prevention of session hijacking attack, In our proposed system modified this concept to provide more security, in addition also implementing SQL sanitation to prevent from SQL Injection Attack and generating log to inform administration about on going or attempted attacks.

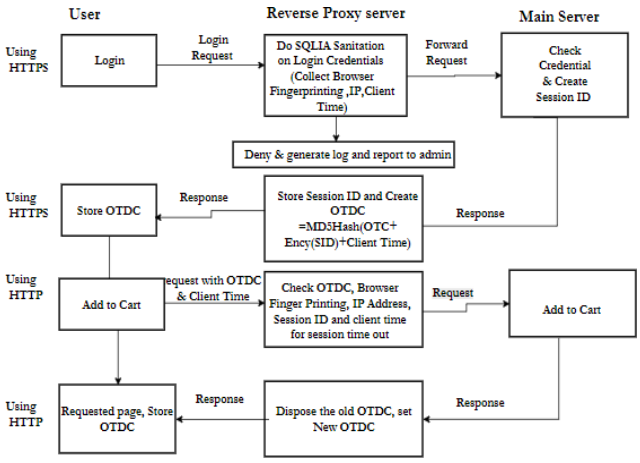


Fig. 4. Block Diagram of Proposed System

The proposed system will work as follows :

1) User request for authentication by doing login.

2) Request will go on Reverse proxy server(RPS)

3) RPS will check Sql Injection Attack Query process will be knowns as SQL Sanitation and collects browser information, time , IP Addess

4) If SQL Sanitation donot found any thing wrong then RPS will forward request to main server for authentication of the user.Other wise drop the activity and generate log and inform to administration.

5) Server will check clients credential, authenticate the user after that generate session id and sent it to the RPS.

6) RPS will save session id and generate new OTDC and sends this OTDC to client

7) Client will receives OTDC and start requesting resources or pages form server using current OTDC.

8) RPS receives request with OTDC and check with saved OTDC if it get matched then request is from actual user then RPS will forward this request to main server other wise it will deny the requested page and generate error log.

9) Forwarded request is received by Main Server then server will process request and send it to the RPS.

10) RPS will receive response and will generate new OTDC and sents response to the user along with new OTDC and disposes old OTDC.

11) This will be continued with evey request in web appliction for entire session.

VI. Expected Outcome

Following are some of the expected results can be given by proposed designed system

- To prevent server from session hijacking attack using concept One-Time Dynamic Cookie (OTDC)
- To prevent server from SQL Injection attack
- To create a system which will be independent to client and server or vice versa
- To Implement a system using Reverse proxy server which will act as intermediary between user and server to protect from malicious attacks.

- Generate logs on attack and inform administration about attack is going on or happened on server.

- To authenticate client, system will use Browser Fingerprinting Concept that include client Ip address, client system time and browser information.

ACKNOWLEDGMENT

I acknowledge this is the proposed system over existing system to provide more security for web applications and this proposed system is in implementation state will complete soon.

REFERENCES

[1] Annies Minu Sathiyaseelan, Vincy Joseph, Anuradha Srinivasaraghavan, "A proposed system for preventing session hijacking with modified one-time cookies", 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), IEEE, March 2017.

[2] D.Madhavi," A Survey on Detection Tools and Prevention Techniques for Session Hijacking Attack", International Journal for Scientific Research & Development (IJSRD), Vol.12, December 2015.

[3] Jerry Louis," Detection of session hijacking", University of Bedfordshire Repository, Department of Computer Science and Technology, Supervisor: Dr. Xiaohua Feng, 10547/211810, AY10/11, January 2011.

[4] Neil Patel, Neel Patel, Manan Doshi, Yash shah," A review on prevention for Session Hijacking using OneTime Cookies", International Journal of Advance research Science and Engineering, Vol. 6, December 2017.

[5] Anuj Kumar Baitha, Prof. Smitha Vinod," Session Hijacking and Prevention Technique", International Journal of Engineering & Technology, March 2018.

[6] Kuldeep Kumar, Dr. Debasish Jena, Ravi Kumar," A Novel Approach to detect SQL injection in web applications", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 2, Issue 6, June 2013.

[7] Italo Dacosta, Saurabh Chakradeo, Mustaque Ahamad and Patrick Traynor," One-Time Cookies: Preventing Session Hijacking Attacks with Disposable Credentials", Converging Infrastructure Security (CISEC) Laboratory Georgia Tech Information Security Center (GTISC), June 2012.

[8] Karis D'silva, J Vanajakshi , K N Manjunath, Srikanth Prabhu, "An effective method for preventing SQL injection attack and session hijacking", 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT),May 2017

[9] Asif Muhammad, NitinTripathi, "Evaluation of Open ID-Based Double-Factor Authentication for Preventing Session Hijacking in Web Applications", Journal of computers, Academy Publisher, Volume 7, No 11, Nov 2012. http://ojs.academypublisher.com/index.php/jcp/article/view/jcp071126232628/5787

[10] Kefei Cheng, Meng Gao and Ruijie Guo, "Analysis and Research on HTTPS Hijacking Attacks", 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing, 2010.

[11] Qijia Zeng, "Random Cookie protocol, a new solution to prevent against", Helsingin yliopisto -Helsingfors universitet- university of Helsinki, May 2016.

Embedded System for Remote Controlled Automobile using RFID

Keyur Pandya *Pareen Harsora* *Kruten Patel*

***Abstract*—Here a Remote Controlled Automobile (Robot) with RFID based authentication system and android phone as a Remote for navigation. The Robot will also support live streaming of camera, connected to robot, interfacing with the android application .This live streaming will help in the main purpose of the drone that is providing security services to the respective client ,the live streaming would be carried out with the help of a WAN that is an Internet connection .The camera will be able to picture capture the entire scene and further on the navigation will be helping the client in order to trace the device and reach to the mentioned location on the maps .All of this will be controlled by making the use of a mobile phone in which a mobile application that is an android application will be used to perform the tasks , which are basically controlling the device.**

I. INTRODUCTION

A Remote Controlled Automobile (Robot) with RFID based authentication system and android phone as a Remote for navigation. The Robot will also support live streaming of camera, connected to robot, interfacing with the android application.This live streaming will help in the main purpose of the drone that is providing security services to the respective client, the live streaming would be carried out with the help of a WAN that is an internet connection.The camera will be able to picture capture the entire scene and further on the navigation will be helping the client in order to trace the device and reach to the mentioned location on the maps.All of this will be controlled by making the use of a mobile phone in which a mobile application that is an android application will be used to perform the tasks, which are basically controlling the device.

II. MOTIVATION

The android application will be helping in tracing the exact location of the device in order to find the captured camera location by the device this justifies the need of navigation tools.

One of the main purpose of the main purpose of the device is to give video based mode to know the things or to see the things which is out of reach of human intervention may be because of a danger zone or less space for a human body to pass through this justifies the need of a camera.

III. PROBLEM DEFINITION

The criteria of problems that would be covered and solved could be defined as:
Android application to navigate the robot:
The android application will be helping in tracing the exact location of the device in order to find the captured camera location by the device this justifies the need of navigation tools.

Robots will have camera to provide live streaming:
One of the main purposes of the main purpose of the device is to give video based mode to know the things or to see the things which are out of reach of human intervention may be because of a danger zone or less space for a human body to pass through this justifies the need of a camera.

IV. OBJECTIVES OF PROJECT

RFID chip is used to maintain a strong authentication for any of its specified users as the RFID numbers of the registered users will be stored in the database and only the registered users will be allowed through the RFID scan to access the android application and control the device.

Android application will be helping in tracing the exact location of the device in order to find the captured camera location by the device which makes the device to also provide navigational service one of the main purpose of the device.

The navigational tools are a very helping factor to locate the device as in where it has reached and what is the current location as by knowing this one is able to know the correct path and follow its route to reach the specified area which is the information given by the device.

The compactness, the small size makes the robot very flexible in penetrating in the defense mechanism of any top class enemy which also gives it a advantage of going in any battleground where a human body could not even enter.

V. SCOPE

This project will consist of a very small budget, a small team, and a limited amount of time to complete. It will leverage as many commercial components as possible, get a baseline system working as quickly as possible and then focus on body design, system integration, and mission evaluation.

VI. ORGANIZATION OF PROJECT

We come to know about the actually proposed work and how things are going to work out. It also covers the literature review part which elaborates on the classification and what type of classification is to be used. It basically gives the whole overall outlying of the project.

Discuss the feasibility of the project and timeline over which the whole of the project will be made. Methodologies to be used to implement the project are also discussed. It also mentions the hardware and software requirements

We see the basic outline of the design of the project. We understand the interconnections between various modules and the hardware component of the project. The ways in which the project components will be linked are also discussed in this chapter.

Here we have explained how the final Project will be demonstrated and in which way it'll be helpful.

VII. THEORY

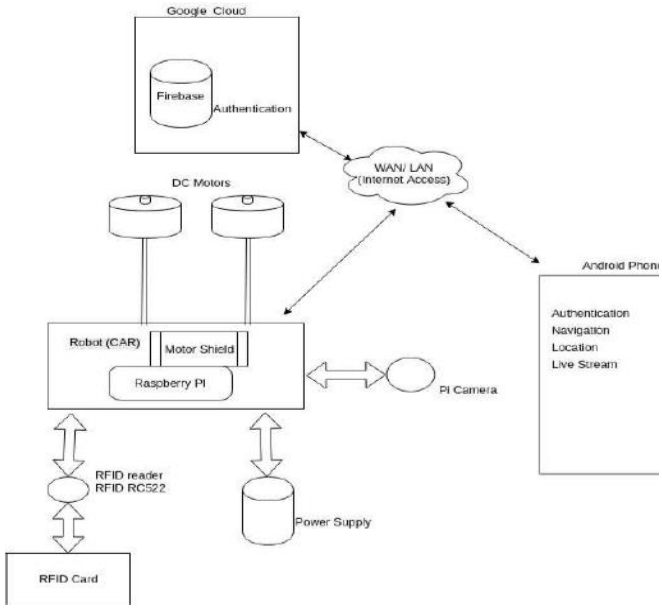
The android application will be helping in tracing the exact location of the device in order to find the captured camera location by the device this justifies the need of navigation tools.

One of the main purposes of the main purpose of the device is to give video based mode to know the things or to see the things which are out of reach of human intervention may be because of a danger zone or less space for a human body to pass through this justifies the need of a camera.

VIII. PROPOSED WORK

To develop a robot that will support live streaming of camera, connected to robot, interfacing with the android application .This live streaming will help in the main purpose of the drone that is providing security services to the respective client ,the live streaming would be carried out with the help of a WAN that is an Internet connection .The camera will be able to picture capture the entire scene and further on the navigation will be helping the client in order to trace the device and reach to the mentioned location on the maps .All of this will be controlled by making the use of a mobile phone in which a mobile application that is an android application will be used to perform the tasks , which are basically controlling the device.

A. System Architecture



B. Flowchart

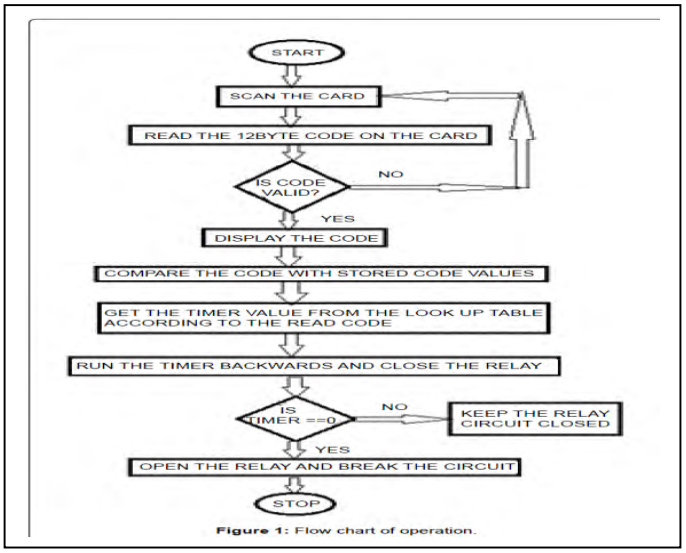


Figure 1: Flow chart of operation.

In a website named viral tech it was mentioned in an article headlined “The boom of automated systems’’ it mentions that “Further enhancement to this system can be done by integrating DTMF technology which makes it possible to control the robot via cell phone which includes an advantage of providing long communication range as compared to RF technology. Using Stealth mode to hide him in case of any casualties, the robot can be used like a helicopter so to get better visuals from all angles.”

X. ACKNOWLEDGEMENT

We sincerely thank our guide Ms. Ashwini Patil for his/her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

We thank the HOD, Dr. Sheetal Rathi, Dean Academic, Dr. R.R Sedamkar, the Principal, Dr. B. K. Mishra and the college management for their support.

XI. LITERATURE SURVEY

In a website named viral tech it was mentioned in an article headlined “The boom of automated systems’’ it mentions that “Further enhancement to this system can be done by integrating DTMF technology which makes it possible to control the robot via cell phone which includes an advantage of providing long communication range as compared to RF technology. Using Stealth mode to hide himself in case of any casualties, the robot can be used like a helicopter so to get better visuals from all angles.”

XII. REQUIREMENT SPECIFICATIONS

Functional

The functional requirement is describing the behavior of the system as it relates to the system's functionality. There are other hardware requirements as well such as Raspberry Pie 3 B+, RFID reader, Robot Body, 2X DC motors, L2933D motor driver and WAN server.

Non-Functional

he non-functional requirements here are more restrictive and most important to meet the end user of the product. Software and compatibility in implementing, programming languages used is python, so the code is easily understandable to facilitate incorporation of further improvements in the future. In this way it is easily portable and alright weight automobile.

XIII. FEASIBILITY ANALYSIS

Here the Feasibility analysis is given as:

The EFS is composed of two required forms:
Business case
The Business Case provides an analysis of the business environment including-
Expected customers : Government
The nature of the business : National
Process of payment : using Net Banking
The Business Case also presents the benefits of the proposed project.

Cost Benefit Analysis
The Cost Benefit Analysis summarizes the revenues and costs involved with the proposed project. As the proposed system will be used for the benefits of the nation or government and the products used in it are lower cost, a high end deliverable is manufactured.

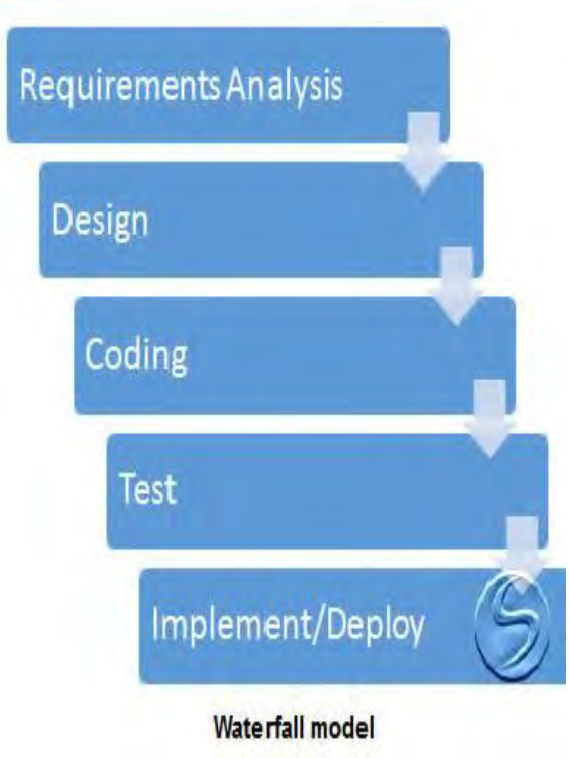
Technology used at front end: Raspberry Pi , Camera
Technology used at Back end: C, C++
Resources Required: Manpower, Programmers, testers, debuggers
Software required: Testing Tools (to perform black box and white box testing), GitHub
Editor: Arduino UNO
Hardware required: PC for development, server for deployment

Management support, employee involvement, and commitment are key elements required to gauge managerial feasibility in the proposed project. The success and the profitability of the project partly depend on managerial competence of the major ingredients of the proposed project which are the users i.e. companies and the benefiters i.e. vagrants. The capability of the infrastructure of a process is to achieve and sustain the properties of isolation, atomicity, durability, consistency etc. In the matter of the data stored of the users and benefiters. Also, constant attention is required in the case of amount of log in files being created.

The proposed system is concerned with the military use. It provides a strong authentication and location tracker on itself which can be used as a spy robot in military.

The capability of the infrastructure of a process is to achieve and sustain the properties of isolation, atomicity, durability, consistency etc.

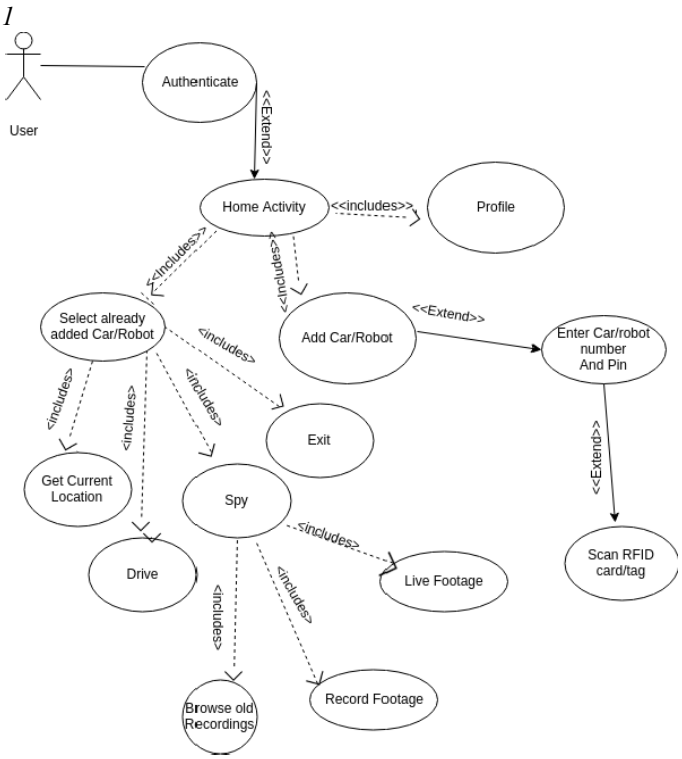
Since we are familiar with the Agenda and Requirements, this portal will be developed using waterfall software development model.
The development will be divided in five phases:
Analysis
Design
Coding
Testing
Deployment



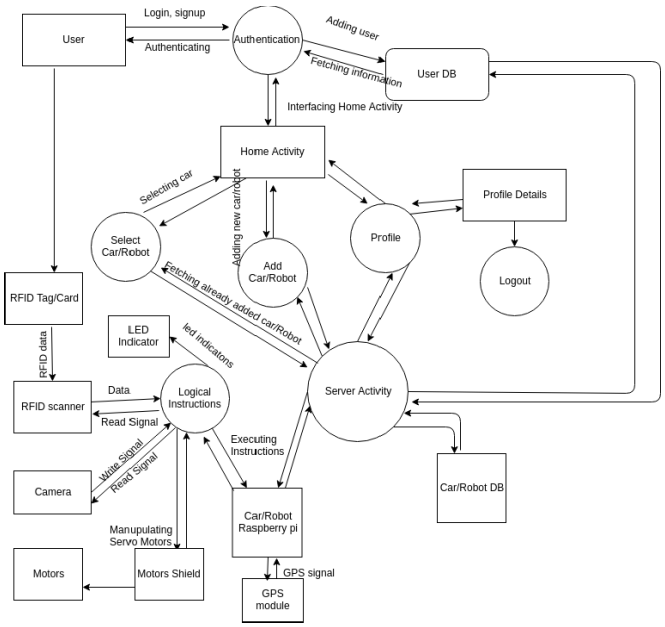
Waterfall model

XIV. USE CASE MODEL

I



XV. DFD MODEL



XVI. CONCLUSION

The main purpose of creating the robot is to provide advanced levels of security which will be achieved due to the fabulous characteristics of the product.
The Second part is that by using the product is to reduce human efforts and by doing this reducing the casualties that can be caused if a human enters a dangerous war prone region. Also the small size of the product concludes that it can enter small areas with less space easily.
Hence this remotecontrolled automobile is convenient and efficient to be used in all sectors where security is topmost priority.

XVII. FUTURE SCOPE

Further enhancement to this system can be done by integrating DTMF technology which makes it possible to control the robot via cell phone which includes an advantage of providing long communication range as compared to RF technology.
Using Stealth mode to hide himself in case of any casualties the robot can be used like a helicopter so to get better visuals from all angles.

XVIII. APPLICATION OF PROJECT

The remote controlled automobile is helping in tracing the exact location of the device in order to find the captured camera location by the device this justifies the need of navigation tools which also helps to improve tracing.

The main aim of the device is to give video based mode to know the things or to see the things which is out of reach of human intervention may be because of a danger zone or less space for a human body to pass through this justifies the need of a camera as it can capture certain picture which may help in the following mission.

XIX. OUTCOME

The expected outcome of creating the robot is to provide advanced levels of security which will be achieved due to the fabulous characteristics of the product.
The other aspect is that by using the product is to reduce human efforts and by doing this reducing the casualties that can be caused if a human enters a dangerous war prone region which will help the to be convenient and efficiently used in all sectors where security will be of utmost importance

XX. REFERENCES

1. *IJSER platform 2018, www.ijraset.com (visited on 2018-08-2).*

2. *Semantischolar platform 2015, pdfs.semantischolar.org (visited on 2018-10-10).*

3. *Oaji platform 2016, www.aoji.com (visited on 2018-09-5).*

4. *About Raspberry Pi: www.raspberrypi.org: This is the official website of the Raspberry Pi project (visited on 2018-10-2).*

“Cloud Based Automated Malware Analysis for Advanced Persistent Threats for Government organization”

Nikhil Pandey Amit Tiwari Ekta Tiwari

Abstract

Goal of this project is to make your web application more difficult to hack. Web Application or any software for that matter will never be completely secure and free for detect. It is only matter of time. Before a determined attacker will find some vulnerability or misconfiguration to exploit and compromise either your site or one of its user.

You should take moment to come to terms with this truth before processing. Many people wrongly assumed that hiring smart developer or deploying commercial security product will magically make their site hacker proof. So, we are making a virtual image of system implemented by top security features.

1. Introduction

The Project will be Self configurable Virtual Machine Image which can be used by an enterprise to protect its own application by putting VM image in front of Application. Enterprise need to develop application which can demonstrate correct functionality. Non Functional Requirement of the application like Scalability and Security will be handle by Self configurable Virtual Machine. Here VM will protect Enterprise wide application though various Hierarchical mechanism. VMs will allow Access level Protection along with Predictability of Attack on Application.

The outcome of the Project will be Self configurable Virtual Machine Image which can be used by an enterprise to protect its own application by putting VM image in front of Application. Enterprise need to develop application which can demonstrate correct functionality. Non- Functional Requirement of the application like Scalability and Security will be handle by Self configurable Virtual Machine. Here VM will protect Enterprise wide application though various Hierarchical mechanism.

2. Theory

Now a day’s malicious attack keep occurring. One of the main reasons is the poor quality of the software used in systems and application software. So, we are came up with the innovative idea for making network secure from malicious file entering network. We are making a product which will be act as intermediate between the client and users (legitimated users / hackers). This product will fully responsible for detecting malicious files that enter the organization network

Network will never be completely secure and free for detect. It is only matter of time before a determined attacker will find some vulnerability or misconfiguration to exploit and compromise either your site or one of its user. You should take moment to come to terms with this truth before processing. Many people wrongly assumed that hiring smart developer or deploying commercial security product will magically make their network hacker proof.

Making a product with highly configured security core rules. Initially we are developing a sandbox for static and dynamicanalysisofafileinanetwork.

The existing system include direct access of client and server, where in people are just relying on where they are hosting there website, or a platform where they are deploying there website which is completely dependent on the users concern.

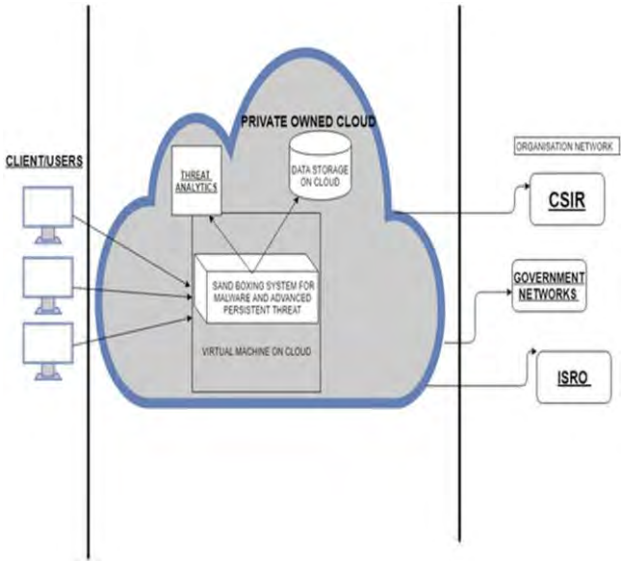
We need to analyze the need of the clients for our project also the demand in the market. Everything depends on demand and supply policy. If the demand is more and use of our product will be more. We need to analyze how our system is better compared to other systems this can be done by identifying gaps and also providing features that our others are not providing and make our product more useful.

This project will majorly attract the population of the conservative mindset with proper use of technology to lover the exhaustion of cloud based knowledge and startups.

3. Project Work

Our Project idea is to provide Cloud based Security against signature/anomaly based attacks i.e. sophisticated malware and zero-day threats to government network and Websites without sharing any data to any publiccloud(AWS,Azure)vendors.

Many government organizations use third party cloud which is not a full proof safe idea considering the importance of the data. Our idea is to have a government owned private cloud that will protect the government network against the malicious attack by implementingasandboxsystem insideavirtualmachine thereby analyzing the malicious file and giving a comprehensivereportregardingthefile.



SystemArchitecture

4. Conclusions

This cloud based system can be used to protect the government network by blocking any malicious activity by its sandbox (efficiency of sandbox depends on its implementation). The detailed reporting of the malicious files provided can also be used for the security auditing of the organization. The proposed use of government owned private cloud adds to the security and also to the overall performance of the system. Deployment of the system on cloud will provide much flexibility to its use. The cloud will act as an intermediary between the end user and the government network thereby securing the government network. It will also provide the protection from zero day attacks which cannot be achieved by the products of current leading vendors.

5. FutureScopes

This project will majorly attract the people those running his /herbusiness online and want to grow his business with full boost. In initial state it may not stand with the today's security service provider like GoDaddy and CloudFlare and many more. But it will be best product in future. In market there are many security service providerascompetitor.

6. Acknowledgement

We sincerely thank our guide Mr. Vikas Singh for her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry outtheprojectwork.

We thank the HOD, Dr. Sheetal Rathi, Dean Academic, Dr. R.R Sedamkar, the Principal, Dr. B. K.Mishraandthecollege managementfortheirsupport.

7. References

[1] Ryan C. Barnett, Web Application Defender's Cookbook, Battling Hackers and Protecting Users, December 2012.

[2] Ivan Ristic, Modsecurity Handbook,2010.

[3] OWASP top ten securitycorerulesetfor latest security in web application_ https://www.owasp.org/index.php/Category:OWASP_Top_Ten_Project (Accessed on 12 October,2016)

[4] OWASP Appsensor Project https://www.owasp.org/index.php/OWASP_AppSensor_Project (Accessed on 15 September,2017)

Design of Smart Ambulance System with an added feature of Traffic Overriding

Lydia Suganya

CMPN, ThakurCollege of Engineering and Technology, Kandivali-East, Mumbai, India
lydia.jacob@gmail.com

***Abstract*—Patients in Intensive Care Unit stay in critical condition and needs special care and treatment. At this state the team of doctors need to have a real time update of patient's health related parameters like their blood pressure, heart pulse and temperature. It is very frustrating task to calculate it manually, and if patients are in large numbers then it becomes close to impossible. To avoid this type of circumstance a system based on Internet of Things (IOT) can be used because of its automation nature that can give updates to the Doctors all time using internet. This system is an Arduino based system that takes all patients' information using few sensors. It uses Wi-Fi module to communicate this information to the internet. By using sensors for finding blood pressure range and heart beat rate, we can monitor patient's health electrically, provided they have to be physically connected by the patient. When we press the button, the sensor detects the blood pressure and also the heart beat and these details will reach doctors system using internet. A temperature sensor present in the system detects the temperature of environment, when this sensor is close to the patient it measures the patient's body temperature. Also an IOT based automated traffic signal monitoring system mechanizes entire traffic signalling and allows us for overriding the traffic signal over internet.**

Keywords-IOT, Patient Monitoring System, Heartbeat Sensor, Automated traffic signal monitoring.

I. INTRODUCTION

Taking care for anyone after their own family members and other relative in today's fast moving days has become difficult. Keeping track of the health status of the patient at home is a difficult task. Especially for the patients who are very old in their age for them this is the most important work to check their health on time to time based at work. Thus we are trying to make an innovative system that automated this task with ease of use. Our system uses different type of sensors such as temperature as well as for a heartbeat sensing to keep track of patient health. The sensorswhichhaveconnectivitywith a microcontroller are used tomonitor the status of patientby using the interface of an LCD display as well as Wi-Fi connection fortransmitting alert message. If there is any changes in patient heartbeat or body temperature, the system will automatically send alert message to the user about the patient's status over IOT and also shows information of heartbeat and temperature of patient live on the internet. Thus IOT based patient health monitoring and tracking system effectively uses internet to monitor the patient health status and save precious lives on time.

And forthe module of traffic overriding it aims to provide an effective solution for the traffic signal system to change the normal signals during emergency automatically. There is always an emergency situation for ambulance or fire brigade where they can stuck in traffic; they require priority to go first. Also if there is hightrafficat some area, then system uses an android application for controlling traffic and overrides the signals immediately giving green signal in the vehicle direction and red signal for all others. The density is measured in three ways low, medium and high according to which the timings are allotted for signals.[1] The timing overridden is done using Bluetooth technology in android device.

II. BACKGROUND

We are trying to monitor the patient's health from ambulance itself by using the values of various sensors like heartbeat, and body temperature. Hence all these data will be monitored and will be directly send to the doctor's web portal via Wi-Fi module form ambulance itself. When the ambulance arrives at patients door in the emergency it take time to reach hospital and it is possible that patients health can be in danger before reaching hospital then checkup takes place and then according to health conditions precarious are taken and there is also problem for ambulance to get stuck in traffic while reaching hospital. Above described problem can easily get solved, if we use smart ambulance system with traffic overriding. So we are proposing to make a Smart ambulance system with additional facility for ambulance to reach hospital as quick as possible by providing module of traffic overriding in which our idea is to create a system in ambulance itself that can measure the patient's heartbeat, and body temperature and sends this data to the doctor's web portal form ambulance itself, upon which doctor can make a list of prescribed medicine or make sure that all the required things are arranged before ambulance reaches hospital.

III. PROBLEM DEFINITION

There will be a system which includes two modules for making the ambulance smart and automated. One will makes sure that all the patient's health details like heart-beat and body temperature will reaches the hospital via Wi-Fi module the hospital. And the second will make sure that the ambulance does not stuck in the traffic and reaches the hospital as fast as it can by providing the application on the mobile of the driver by which the traffic signal can be overridden and ambulance can make its way to hospital.[2] This both modules are secure by authentication way which includes user-name and password facility to both of them. Which means no other then authorized party can use this system.

IV. SCOPE OF THE SYSTEM

The system collects vital information like body temperature, and heartbeat of the patient and regularly needs to be updatedtothe doctor. The doctor can monitor thedetails of patients' health on web portal now and then take all the required action which are necessary before the ambulance reaches to the hospital as quick as possible. And android application will make sure that at any point there is no problem for the ambulance to get stuck in traffi.[3] The system can be extended by adding more features like checking blood pressure of the patient, interaction of doctor form hospital into ambulance and etc., Doctors can create awareness about diseases and their symptoms through information that they got on their web portal. From the evaluation and the result obtained from analysis the system is better for patients to reach hospital easily and for doctor to take decision earlier before ambulance reaches hospital.

V. APPLICATION OF THE SYSTEM

The paper has a variety of application such as system for fire bridged driver, police vans, sending health details using internet to hospital directly in serious conditions, overriding traffic by the authorized person in serious conditions and etc.[4]

The System has various impact such as doctors in hospital stays very alert and takes decisions by taking time, Patient health parameter data is reached to doctor before patient physically reaches in hospital. [5] So it is more beneficial as there can be precarious taken before time, In critical condition an emergency call is given to ICU and ambulance, Doctors can take all necessary action without visiting to patient, Highly sensitive.

VI. LITERRATURE SURVEY

We have conducted a literature survey based upon which we have found the gaps in the existing system which can be seen in the table below (Table-I).

TableI: Literature Survey

Existing System Authors	Title of the paper	Gap Identified
Vikas Vippalapalli and Snigdha Ananthula[1]	Internet of things (IoT) based smart health care system	Harmful for patient body as sensors are placed on body, Sensors can get defect because of outside world conditions, And the cost of sensors used are very high.
Xiaofeng Ye, Bin Sun,Weiwei Lin [2]	A remote control system on traffic lights based on IEC61131-3 standard language	Setting of the traffic signal needs monitoring on fixed time delay, Battery need to change at regular time interval, Resolution of the screen used for monitoring purpose is very small.
Mukremin Ozkul,Elton Domnori[3]	A traffic signal control system through anonymous messages	Not Compact, Message traffic due to unnecessary needed message, communication is based on short message, Real time monitoring of traffic can get incorrect, Wireless communication with traffic signal can get disturb due to connection issue.
Punit Gupta , Deepika Agrawal, Jasmeet Chhabra[4]	IoT based smart healthcare kit.	Not Portable, Overhead due to carrying, Emergency message can get disturbed and unnecessary messages can be send.
B.Janani Saradha, G. Vijayshri, G. Vijayshri[5]	Intelligent traffic signal control system for ambulance using RFID and cloud	Speed is slow, Time required is more, Overhead due to cloud implementation.
Salman Ahmed, Sabrin Millat, Md. Aymanur Rahman[6]	Wireless health monitoring system for patients	Overhead due to Telecommunication technology implementation
Y M Jagadeesh, G. Merlin Suba, S Karthik[7]	Smart autonomous traffic light switching by traffic density measurement through sensors	Complex implementation because of Artificial Intelligence, There can be problem generated due to dynamic control of traffic signal.
Yujun Ma, Yulei Wang, Jun Yang[8]	Big Health Application System based on Health Internet of Things (IoT) and Big Data	Required huge amount of data storage for big data.

The Existing System uses Raspberry pi kit which increases the cost of the entire system and overhead as the existing system build a network, speed of execution is also slow and time taken is more, in our system we have proposed to use a Arduino kit which reduces cost as well complexity and No Overhead, our System contains One small sized system in which all components will be fixed and one more advantage is after prescription of doctor prescribed medicine will be taken before it reaches hospital and traffic problem is solved with the following components such as Bluetooth, Wi-Fi Module, Switch, Heartbeat, and Temperature Sensors, Power Supply Modules and Software like IOT Gecko libraries, and embedded language.

VII. PROPOSED WORK

Smart ambulance system with traffic overriding is very user friendly and in terms of handling also. [6] Smart ambulance system with traffic overriding can be needed to monitor and overriding it in emergency. Our system consist of two modules i.e.; patient health monitoring system (Fig 1.1) and trafficoverriding system (Fig 1.2)

It can detect and monitor the temperature and heartbeat of the patient and sends it to the hospital from the ambulance itself so that the crucial time which is needed to save the patient's life can

be saved because it gives details to the hospital's management and doctor for taking the precautionary measures.

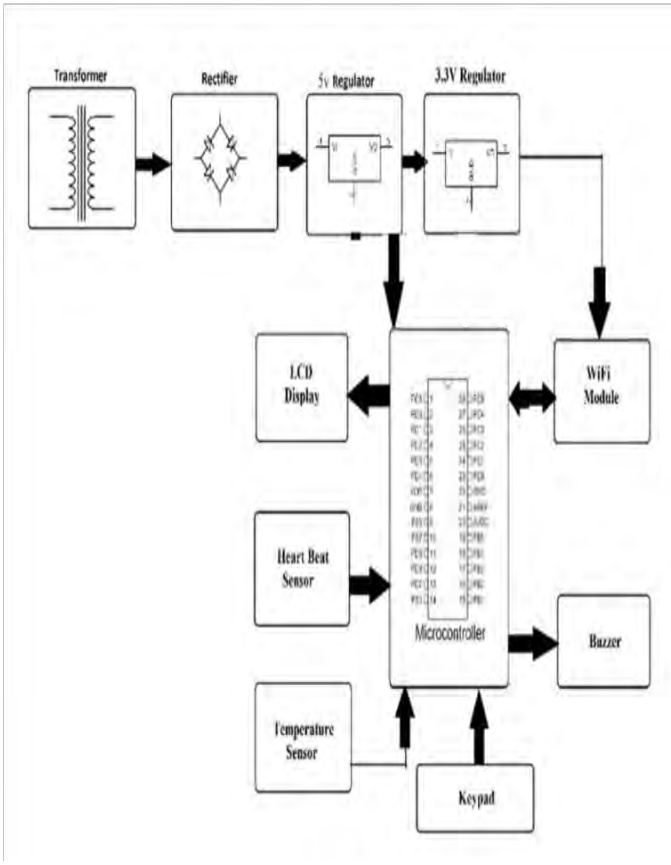


Fig 1.1: Block Diagram (Patient Health Monitoring Module)

When there is lot of traffic the density of the traffic will be measured by the sensors present on the board which will help in toggling the traffic light from red to green and vice versa. [7]

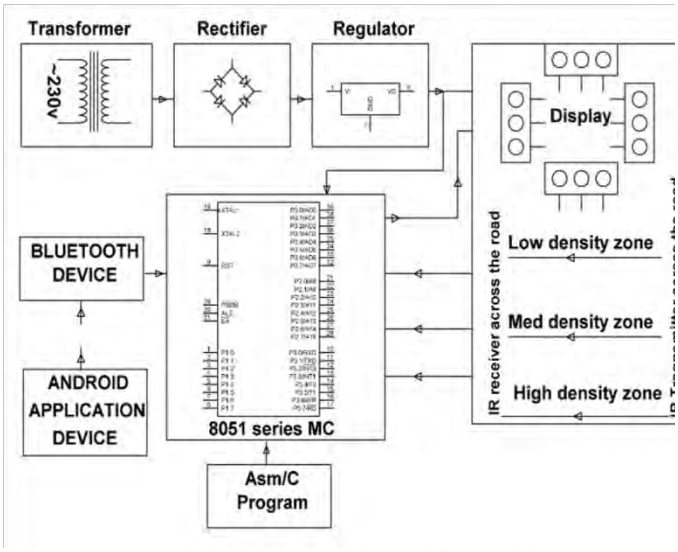


Fig 1.2: Block Diagram (Traffic Overriding Module)

Overriding of traffic can be used to make the ambulance to reach hospital as quick as possible. Most important thing is that the measures which have been given by the doctor from hospital can be given to the patient so that his life can be saved**.** Data transfer speed given from the ambulance is quite high so that the doctor's precautionary measures given can be sent quickly

VIII. EXPECTED OUTCOME

The system will provide facility of monitoring patient's health using various types of sensors. Using the android application the driver of ambulance can avoid the traffic and makes its way to the hospital. [8] It will improve the efficiency of the health monitoring and traffic signal overriding with this system. Fig-2 Shows the system Architecture of Smart Ambulance System.



Fig 2: System Architecture for Smart Ambulance System

IX. SYSTEM WORKFLOW

A) Patient side (Arduino module):

Temperature sensor will sense the temperature, heartbeat sensor will give the pulse count of patient continuously through Arduino using Wi-Fi module. Monitoring will be done continuously. This kit will be near to the patient. The details of the patient will be transferred to the doctor side web portal.

B) Doctors Side Web Portal (PC):

Server will collect the data from patient side kit and will update it in data log. It will check if the condition is normal or critical, according to that action will be taken. If the condition is not normal the doctor will make prior arrangements before the patient arrives. If doctor gives any prescription to patient, it will a send signal to secondary doctor present in the ambulance. And the secondary doctor will provide the medication to the patient in the ambulance.

C) Traffic Signal overriding (android app):

When there is lot of traffic the density of the traffic will be measured by the sensors present on the board which will help in toggling the traffic light from red to green and vice versa. Suppose there is a ambulance travelling from a lane which has traffic congestion so we can override the traffic signal with the help of this android application. This application will be handed over to an authorized person so that one cannot misuse or play around with traffic signal.

X. EXPERIMENTAL SETUP

The System is proposed with the aim of providing efficient healthcare to the patient for continuous monitoring and providing help in case of critical situation for which we have used Arduino, temperature sensor, heartbeat sensor, Wi-Fi module and LCD display as shown in figures listed from Fig 3.1 – Fig 3.4



Fig 3.1:- Heartbeat Sensor (Measures the heartbeat of the patient)

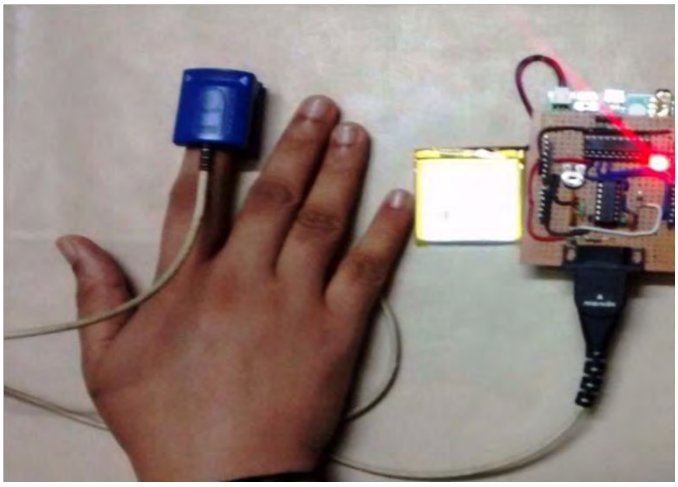


Fig 3.2:- Temperature Sensor (Measures the body temperatureof the patient)

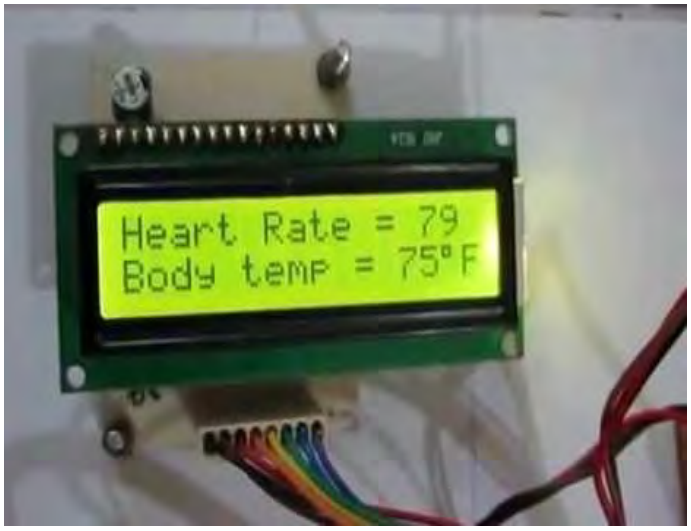


Fig 3.3:- LCD Display (Displays Result)



Fig 3.4:- Wi-Fi Module (Used for transferring of data)

XI. RESULTS

The User interface of the android application used in traffic overriding module proposed System is as shown in the fig 4: User Interface (Traffic Signal Overriding) and we have shown the comparison between the existing system and the proposed system with respect to certain parameters as shown in Table 2.

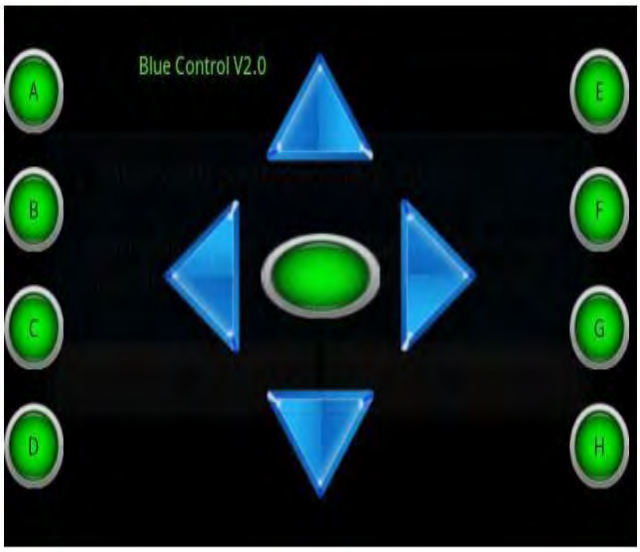


Fig 4: User Interface (Traffic Signal Overriding)

The table below shows the comparison between the existing system and the proposed system based on certain parameters as described below:

Table II: Existing vs. Proposed System

Parameter	**Existing System**	**Proposed System**
No. of Components used	Many	Medium(As compared to others)
Cost	High	Medium
Portable and compact	Portable and compact(due to wearable tags)	Portable and compact (due to size)
Type of Kit(Arduino /Raspberry pi)	Raspberry Pi	Arduino
Overhead and Speed/Time	Yes	No
User Interface	Complicated	User Friendly

XII. CONCLUSIONS

An idea is proposed for saving a precious life of patient's in a better and efficient way possible. It is beneficial for users in case of emergencies as it saves time. The smart ambulance provides necessary equipment required to track the patient's health status. The live feed data sent through the ambulance to the hospital helps in keeping track of patient's health details. Sending patient's health information to the hospitals helps the hospital staff to get the necessary pre-requisites regarding the patient's treatment. Hence it reduces the time complexity and helps to provide faster medical services.

The traffic signal overriding module only gives the authority to the ambulance driver in order to have control over the traffic signal. The signal can be changed depending on the traffic condition. The signal is changed to green in case of heavy traffic. Thus the ambulance takes the patient to the hospital in a short period of time without any delay.

REFERENCES

[1] VikasVippalapalli ETM Dept., GNITS, Hyderabad, India http: and SnigdhaAnanthula ETM Dept., GNITS, Hyderabad, India "Internet of things (IoT) based smart health care system" Published in: Signal Processing, Communication, Power and Embedded System (SCOPES), 2016 International Conference //ieeexplore.ieee.org/document/7955637/.

[2] Xiaofeng Ye College of Metrology and Measurement Engineering, China Jiliang University, Hangzhou 310018, China; Bin Sun College of Metrology and Measurement Engineering, China Jiliang University, Hangzhou 310018, China; and Weiwei Lin College of Metrology and Measurement Engineering, China Jiliang University, Hangzhou 310018, China "A remote control system on traffic lights based on IEC61131-3 standard language" Published in: Intelligent Control and Automation (WCICA), 2010 8th World Congress on //ieeexplore.ieee.org/document/5553830/

[3] Mukremin Ozkul Department of Computer Engineering, Epoka University, Tirana, 1039, Albania, Elton Domnori Department of Computer Engineering, Epoka University, Tirana, 1039,

[4] Albaniahttp:"A traffic signal control system through anonymous messages Sign In or Purch" Published in: Intelligent Vehicles Symposium (IV), 2017 IEEE //ieeexplore.ieee.org/document/7995893/

[5] Punit Gupta Jaypee University of Information Technology, Himachal Pradesh, India, Deepika Agrawal Jaypee University of Information Technology, Himachal Pradesh, India,

[6] Jasmeet Chhabra Jaypee University of Information Technology, Himachal Pradesh, India "IoTbased smart healthcare kit" Published in: Computational Techniques in Information and Communication Technologies (ICCTICT), 2016 International Conference //ieeexplore.ieee.org/document/7514585/

[7] B. Janani Saradha Department of IT, Sri Sairam Engineering college, India, G. Vijayshri Department of IT, Sri Sairam Engineering college, India, T. Subha Department of IT, Sri Sairam Engineering college, India "Intelligent traffic signal control system for ambulance using RFID and cloud" Published in: Computing and Communications Technologies (ICCCT), 2017 2nd International Conference http://ieeexplore.ieee.org/document/7972255/

[8] Salman Ahmed Department of Electrical and Electronic Engineering (EEE) American International University Bangladesh, Dhaka, Bangladesh, Sabrin Millat Department of Electrical and Electronic Engineering (EEE) American International University-Bangladesh, Dhaka, Bangladesh, Md. Aymanur Rahman Department of Electrical and Electronic Engineering (EEE) American International University Bangladesh, Dhaka, Bangladesh "Wireless health monitoring system for patients" Electrical and Computer Engineering (WIECON-ECE), 2015 IEEE International WIE Conference http://ieeexplore.ieee.org/document/7443888/

[9] Y M Jagadeesh Department of Mechanical, Panimalar Engineering College, Chennai, India, G. Merlin Suba Department of Mechanical, Panimalar Engineering College, Chennai, India,

[10] S Karthik Department of Mechanical, Panimalar Engineering College, Chennai, India "Smart autonomous traffic light switching by traffic density measurement through sensors" Published in: Computers,

Communications, and Systems (ICCCS), InternationalConferencehttp://ieeexplore.ieee.org/abstract/document/7562885/

[11] Yujun Ma Center of Information Development and Management, Nanyang Institute of Technology, Nanyang, China, Yulei Wang Center of Information Development andManagement, Nanyang Institute of Technology, Nanyang, China,

[12] Jun Yang School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China; "Big Health Application System based on Health Internet of Things and Big Data" Published in: IEEE Access (Volume: 5) http://ieeexplore.ieee.org/document/7782325/

IOT Based Fire Fire Fighting Robot using Arduino

Chinmay Verlekar, Pratik Patel, Manoj Dixit, Aakash Kanojiya, Vaishali Nirgude

Computer Engineering, Thakur College of Engineering, Mumbai, India

Abstract— Fires are the accidents which occur anytime, anywhere and result is loss of human life and property. Fire fighters cannot access the source of fire due to some constraints such as damage of building, very high temperature, or the presence of explosive materials. Through this project proposal, we can design a robot which can be placed where human lives are at risk. The robot can operate in the environment which is out of human reach in very short time. In such environments, fire fighting robots can be useful for extinguishing fire.It uses flame sensor for fire detection and Arduino board for processing. The robotic vehicle is loaded with water tanker and a pump which is controlled over wireless communication to throw water. An Atmega 8 microcontroller is used for controlling the desired operation of Robot .The robot is also equipped with cameras and ultrasonic sensors. The camera plays role in giving feedback to user and in finding the source of fire. Ultrasonic sensors are used to avoid collisions during movement. Feedback provided by camera on the robot will display on a screen of Smartphone. Further the robot works on IOT mechanism which will help to identify the exact location of the fire and providing an alert message to user.

Keywords—Arduino, Ultrasonic Sensor, Temperature sensor, Bluetooth module, Auto sprinkler, IoT.

I. INTRODUCTION

The increase in the human population and technological development has led to an increase in accidents and fire risks. As per the survey done by National Crime Records Bureau (NCRB), it is estimated that more than 2.1 lakh deaths have been caused because of fire accidents in India from 2010 to 2017. In the event of a fire breakout, to rescue people and to put out the fire we are forced to use human resources which is not safe. An IOT based Fire Fighting Robot (using Arduino), is an initial step towards robotics in firefighting. A robot detects and extinguishes fire that can be controlled from a specified distance. It uses flame sensor for fire detection and Arduino board for processing. Fire extinguisher is used to extinguish the detected fire. The vehicle consists of a water tank along with a pump which can throw water when needed. The android device is used as a transmitter to send over controlling commands to the vehicle. The android device provides a good touch based GUI for controlling the robotic vehicle.

The robotic vehicle is loaded with water tanker and a pump which is controlled over wireless communication to throw water. An Atmega 8 microcontroller is used for controlling the desired operation of Robot At the transmitter side, User sends command to the microcontroller which is on the robot and then translated into robotic movement either to move forward, backward and left or right etc. At the receiving end five motors are interfaced to the microcontroller where four of them are used for the movement of the vehicle and the remaining one to position the arm of the robot. Remote operation is achieved by any smart-phone with Android OS, upon a GUI based touch screen operation. The android application device transmitter acts as a remote control that has the advantage of adequate range, while the receiver have Bluetooth device fed to the microcontroller to drive DC motors via motor driver IC for necessary work. A water tank along with water pump is mounted on the robot body and its operation is carried out from the microcontroller output through appropriate signal from the transmitting end. A motor driver IC is interfaced to the microcontroller through which the controller drives the motors. The robot is also equipped with cameras and ultrasonic sensors. The camera plays role in giving feedback to user and in finding the source of fire. Ultrasonic sensors are used to avoid collisions during movement. Feedback provided by camera on the robot will display on a screen of Smartphone. Further the robot works on IOT mechanism which will help to identify the exact location of the fire and providing an alert message to user.

II. LITERATURE SURVEY

TABLE 1.1 COMPARATIVE STUDY OF IOT BASED FIRE FIGHTING TECHNIQUES

Paper	Description
The Intelligent Fire Fighting Tank Robot (2010),Kristi Kosasih, E. Merry Sartika, M. Jimmy Hasugian, danMuliady	Robots can be used for a high risk task such as fire fighting. The tank robot is equipped with various sensors and detector for navigation and extinguishing fire.
Fire Fighting Robot : An Approach (2012), Ratnesh Malik	Fire Fighters have to face dangerous situations while extinguishing the fire. This results in need of firefighting machines to assist the fire fighters to avoid deaths by handling the dangerous situations.
Cell Phone Controlled Robot with Fire Detection Sensors (2015), Lakshay Arora , Prof.Amol Joglekar	The mobile phone controls a robot by making a call to the mobile phone that would be attached to the robot. The Robot makes use of dual-tone multiple-frequency (DTMF) tone. The received tone is decoded and accordingly an action is performed by the microcontroller

III. METHODOLOGY

A. Proposed System

The robot will enter the room where there is fire and will detect the fire in the room. It will move automatically in the room and look for fire. The robot will automatically avoid the obstacles in the way. When fire will be detected the sprinkler will turn on and fire will be put off. The water tank will be present with the robot. The pipe can turn through 90 degree so that it can be used efficiently for fire extinguishing. The camera on the robot will help fireman to guide the robot through the room. The robot can also be controlled by fireman through android phone externally. In the application Obstacle detection Module for the Check whether there is obstacle in front of the robot. If obstacle is within the certain range then stop the robot. When the Fire detection Module is for the robot will move in the room and detect the fire.If the temperature at certain place is more than the threshold set then start the water sprinkler. .Else detect the fire inside the room.the Fireman Module When the robot will be moving inside the room it can be controlled by the fireman using android phone. When obstacle detected the robot will be guided by the fireman where to move. The camera on the robot will allow fireman to analyze the situation and turn the pipe of water sprinkler.

microcontroller which is on the robot and then translated into robotic movement either to move forward, backward and left or right etc. At the receiving end five motors are interfaced to the microcontroller where four of them are



Fig 1.1 :Proposed system

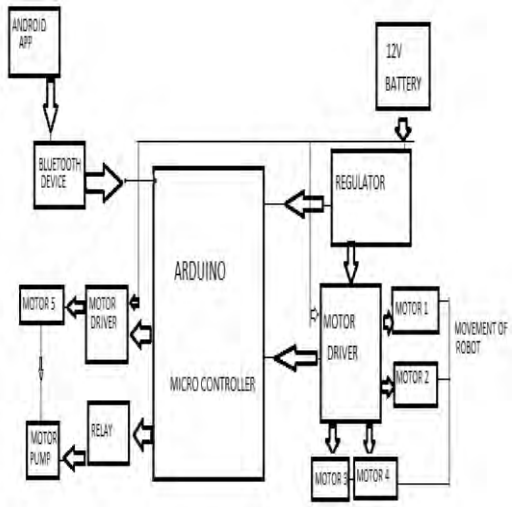


Figure 1.2: Block Diagram

IV. REQUIREMENT SPECIFICATION -

A. Software

1) Eclipse (Keil Compiler)

Keil an ARM Company makes C compilers, macro assemblers, real-time kernels,debuggers, simulators, integrated environments, evaluation boards, and emulators for

ARM7/ARM9/Cortex-M3, XC16x/C16x/ST10, 251, and 8051 MCU families.

Compilers are programs used to convert a High Level Language to object code. Desktop compilers produce an output object code for the underlying microprocessor, but not for other microprocessors. the programs written in one of the HLL like 'C' will compile the code to run on the system for a particular processor like x86 (underlying microprocessor in the computer). For example compilers for Dos platform is different from the Compilers for Unix platform So if one wants to define a compiler then compiler is a program that translates source code into object code.

2) Android Code:

Android Code: The first program makes use of the continuous data stream to detect, and stop the car when it has lost connection to the phone. Because of this continuous stream of characters, it also needs to filter repeated characters. The program is called "RC_Motor_Shield_Continuous.ino". It uses the "millis()" function which returns the number of milliseconds since the Arduino's execution started. This function lets the Arduino know when 500 milliseconds have elapsed since the last command was received. Once 500 milliseconds have gone by without receiving a command

B. Hardware

1) Microcontroller:

a. Features:

8K Bytes of In-System Programmable (ISP) Flash Memory

4.0V to 5.5V Operating Range

Fully Static Operation: 0 Hz to 33 MHz

256 x 8-bit Internal RAM

32 Programmable I/O Lines

Three 16-bit Timer/Counters

Eight Interrupt Sources

Full Duplex UART Serial Channel

2) Motors and Drivers

The L293 and L293D are quadruple high-current. Half-H drivers. The L293 is designed to provide bidirectional drive currents of up to 1 A at voltages from 4.5 V to 36 V. The L293D is designed to provide bidirectional drive currents of up to 600-mA at voltages from 4.5 V to 36 V. Both devices are designed to drive inductive loads such as relays, solenoids, dc and bipolar stepping motors, as well as other high-current/high-voltage loads in positive-supply applications. All inputs are TTL compatible. Each output is a complete totem-pole drive circuit, with a Darlington transistor sink and a pseudo- Darlington source. Drivers are enabled in pairs, with drivers 1 and 2 enabled by 1,2EN and drivers 3 and 4 enabled by 3,4EN. When an enable input is high, the associated drivers are enabled, and their outputs are active and in phase with their inputs. When the enable input is low, those drivers are disabled, and their outputs are off and in the high-impedance state. With the proper data inputs, each pair of drivers forms a full-H (or bridge) reversible drive suitable for solenoid or motor applications.

3) Bluetooth Module

Bluetooth is a wireless technology standard for exchanging data over short distances (using short-wavelength UHF radio waves in the ISM band from 2.4 to 2.485 GHz[3]) from fixed and mobile devices, and building personal area networks (PANs). Invented by telecom vendor Ericsson in 1994,[4] it was originally conceived as a wireless alternative to RS-232 data cables. It can

connect several devices, overcoming problems of synchronization.

Bluetooth is managed by the Bluetooth Special Interest Group (SIG), which has more than 20,000 member companies in the areas of telecommunication, computing, networking, and consumer electronics.[5] Bluetooth was standardized as IEEE 802.15.1, but the standard is no longer maintained. The SIG oversees the development of the specification, manages the qualification program, and protects the trademarks.[6] To be marketed as a Bluetooth device, it must be qualified to standards defined by the SIG.[7] A network of patents is required to implement the technology, which is licensed only for that qualifying device.

V. CONCLUSION

This project entitled as "Android Bluetooth Base Control Fire Fighting Robot" was created and developed effectively. For that demonstration purpose a prototype module is built for live demonstration, answers are discovered to be acceptable. As it is a demo module a toy type machine is built with less cost. While creating and developing this proto type module, we've consulted couple of expert's individuals who're understand in a variety of fields, these professionals working at different organizations goes to Hyderabad assisted us while fabricating the robot. As it is a prototype module, much amount isn't invested, the entire machine is built with in your area available components, particularly the mechanical components utilized in this project work are acquired from mechanical fabricators, and they're less than the necessity, large amount of modifications should be transported in design and it is essential to really make it just as real working system. Hence, the module will be enhanced further for acquiring better results. This project says creating a relatively inexpensive, high precision fire fighting robot that is targeted to manage through android mobile (remote). The thought of controlling through remote would be to boost the operator safety. The end result from the thesis is a straightforward robot that is controlled with a wise android phone & also has got the voice instructions. This thesis aims to supply simple recommendations for individuals thinking about building robots. As pointed out earlier, the work continues to be transported out several occasions and also the goal of the thesis would be to familiarize the scholars with basic principles of Adriano and Android to construct anything possible.

VI. FUTURE SCOPE

There is always an upgradation and updation in technology. Enhancement basically means future growth in any field. Here, Future Enhancement means the growth or updation in any our projects. This device has a bit of drone technology included in it. It is accessible under numerous situations and can provide immediate assistance to people in inaccessible areas. This device being a prototype contains basic features. But far more advanced features can be included.

REFERENCES

[1] W. Budiharto, Membuat Robot Cerdas, Jakarta: Gramedia, 2006.

[2] Ratnesh Malik, "Fire Fighting Robot : An Approach" , Indian Streams Research Journal Vol.2,Issue.II/March; 12pp.1-4

[3] Kristi Kosasih, E. Merry Sartika, M. Jimmy Hasugian, danMuliady, "The Intelligent Fire Fighting Tank Robot" , Electrical Engineering Journal, Vol. 1, No. 1, October 2010

[4] H. P. Singh, AkanshuMahajan, N. Sukavanam, VeenaBudhraja ,"Control Of An Autonomous Industrial Fire Fighting Mobile Robot", DU Journal of Undergraduate Research and Innovation

[5] Swati A. Deshmukh, Karishma A. Matte and Rashmi A. Pandhare, "Wireless Fire Fighting Robot", International Journal For Research In Emerging Science and Technology

[6] LakshayArora, Prof.AmolJoglekar, "Cell Phone Controlled Robot with Fire Detection Sensors", (IJCSIT) International Journal of Computer

[7] W. Budiharto, Membuat Robot Cerdas, Jakarta: Gramedia, 2006.

[8] Ratnesh Malik, "Fire Fighting Robot : An Approach" , Indian Streams Research Journal Vol.2,Issue.II/March; 12pp.1-4

[9] Kristi Kosasih, E. Merry Sartika, M. Jimmy Hasugian, danMuliady, "The Intelligent Fire Fighting Tank Robot" , Electrical Engineering Journal, Vol. 1, No. 1, October 2010

[10] H. P. Singh, Akanshu Mahajan, N. Sukavanam, VeenaBudhraja ,"Control Of An Autonomous Industrial Fire Fighting Mobile Robot", DU Journal of Undergraduate Research and Innovation

[11] Swati A. Deshmukh, Karishma A. Matte and Rashmi A. Pandhare, "Wireless Fire Fighting Robot", International Journal For Research In Emerging Science and Technology

[12] Lakshay Arora, Prof.Amol Joglekar, "Cell Phone Controlled Robot with Fire Detection Sensors", (IJCSIT) International Journal of Computer

Efficient Recommendation of IT Problems using Image Recognition

Divij Shah, Paridhi Somani, Siddhita Tawde

Department of Computer Engineering Thakur College of Engineering and Technology, Mumbai, India
divijshah11@gmail.com, psomani27@gmail.com, tsiddhita05@gmail.com

***Abstract*—Troubleshooting is an important factor of adaptation of technology in a wider way, however currently used methods do not provide a user friendly way of troubleshooting. As per our analyses of end user requirements, a simpler way of providing recommendation to problems would be by the means of image capturing and recognition. We will train a model which will analyze the data and detect indicators such as icons and error codes. We will train a model which will analyze the data and detect indicators such as icons and error codes. We will also use different classification algorithms for more accurate results. Collect query images from different open source repositories and train the following classification for different problems and its recommendations.**

I. INTRODUCTION

Troubleshooting is a systematic approach to problem solving that is often used to find and correct issues with complex machines, electronics, computers and software systems. The first step in troubleshooting is gathering information on the issue, such as an undesired behaviour or a lack of expected functionality. Other important information includes related symptoms and special circumstances that may be required to reproduce the issue. Once the issue and how to reproduce it are understood, the next step might be to eliminate unnecessary components in the system and verify that the issue persists, to rule out incompatibility and third-party causes. After common causes are ruled out, the trouble-shooter may resort to the more systematic and logical process of verifying the expected function of parts of a system. One common method is the split-half troubleshooting approach: With a problem resulting from a number of possible parts in series, one tests halfway down the line of components. If the middle component works, one goes to the middle of the remaining parts, approaching the end. If the test finds a problem at the midpoint, one does a split towards the start of the line until the problem part is found. The split-half process can save time in systems that depend on many components. Once the problem part is identified, it may be adjusted, repaired or replaced as needed. Evidence of effective troubleshooting is indicated when the issue is no longer reproducible and function is restored one. The success of troubleshooting often depends on the thoroughness and experience of the trouble-shooter. Template matching is a 'brute-force' algorithm for object recognition. Its working is simple: create a small template (sub-image) of object to be found, say a football. Now do a pixel by pixel matching of template with the image to be scanned for, placing (center of) the template at every possible pixel of the main image. Then, using a similarity metric, like normalized cross-correlation, find the pixel giving maximum match. That is the place which has a pattern most similar to your template (football). The challenge is to merge these two concepts, that is, to implement pattern matching for troubleshooting. The collected data, because of the vague and careless nature of taking screenshots needs to be filtered to remove unnecessary information. After this initial pre-processing, all that is left is to train a model and implement it for a proof of concept.

II. PROPOSED WORK

A. *System Architecture*

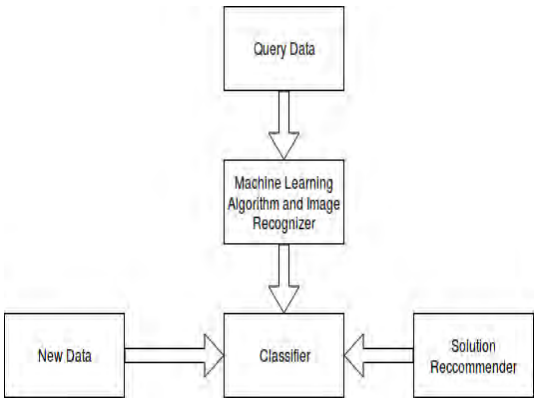


Fig 1: System Architecture

B. *Flowchart*

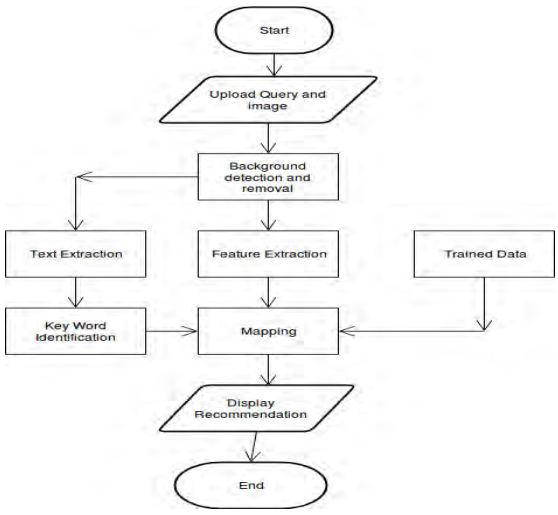


Fig 2: Flowchart

C. Class Diagram

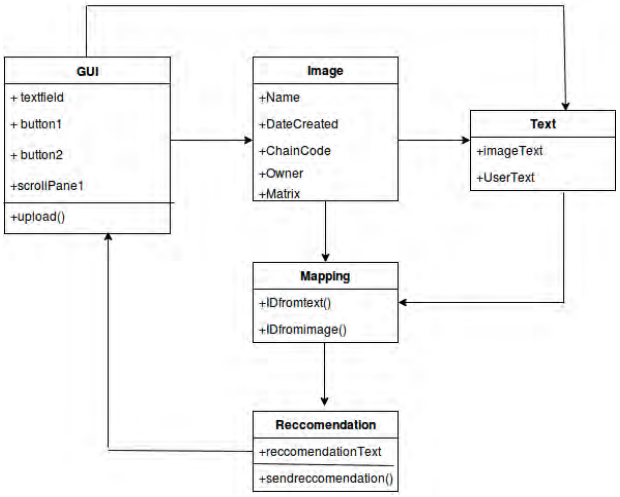


Fig 3: Class Diagram

D. Sequence Diagram

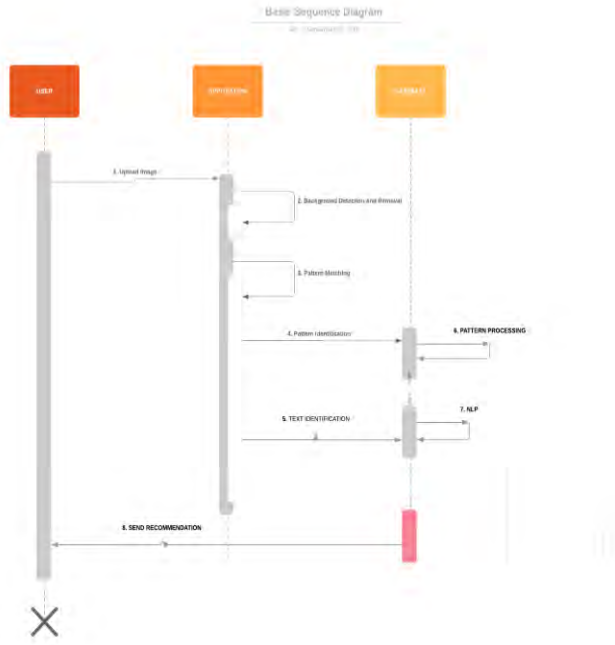


Fig 4: Sequence Diagram

III. LITERATURE SURVEY

Technology has become an indispensable part of human life and with Industry 3.0 fully evolved and Industry 4.0 on the advent of basic computer literacy has become an unavoidable life skill. Yet, the biggest problem faced by the IT industry is difficulty for the consumer to troubleshoot basic issues. Thus, we have decided to simplify the process using ML and Image Recognition techniques. As per our survey, traditional methods such as windows troubleshooting techniques fall short when it comes to user convenience and this leads to wastage of time by organization such as tech support teams. It also creates a hindrance as it leads to avoidance in usage of new technologies.

A. Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine

IV. FEASIBILITY STUDY

A feasibility study is a preliminary study which investigates the information of prospective users and determines the resources requirements, costs, benefits and feasibility of proposed system. A feasibility study takes into account various constraints within which the system should be implemented and operated.

Thus we evaluated the feasibility of the system in terms of the following categories:

A. Technical Feasibility

- Technology used in the front end: QtPy (Desktop App)
- Software Required: Anaconda3
- Hardware Required: Google Cloud (AI Platform)
- Resources Required: Data Sets from various repositories
- Editor: VisualStudio Code

B. Economic Feasibility

A simple economic analysis which gives the actual comparison of costs and benefits are much more meaningful. Our aim is to achieve maximum accuracy from our recommendation system with the help of an adaptive environment. This will reduce the need of checking the problems manually, thus reducing efforts, time invested and ultimately cost of the whole process whilst increasing efficiency.

C. Market Survey

The coming revolution is spearheaded by image recognition and AI technologies. A key challenge for users of image recognition technology is finding the proper solution that will address their specific consumer goods applications. On an observation of the local requirements, recommendation system is accessible, convenient and accurate addresses the daily technical issues hampering their routine work.

D. Reliability

The system adapts according to previously provided results and expectations following a spiral methodology for requirement gathering, short-listing and updating.

E. Scheduling Feasibility

A project will fail if it takes too long to be completed before it is useful. Typically, this means estimating how long the system will take to develop, and if it can be completed in a given period of time using some methods like payback period.

V. RESULTS AND DISCUSSIONS

The recommendation system allows users to upload image of issue terminal and the system will provide the solution steps to the customers, more efficiently and effectively, without facing many hassles or disadvantages. The system helps the users to deal with their daily I.T. muddles with a click of a button on their device. The system can be accessed from any device with an active internet service. Thus, users can then be uninterrupted from their work for a long time, which is the current case. This project will be helpful to the peripheral device companies, as well as to the common people who want to get acquainted to technology. It will be helpful to peripheral device companies as

it shall reduce the time and efforts invested by actual humans in detecting (and solving) an issue. It shall benefit the common man as the solution shall be customized as per their need and will require lesser investigation thus saving them time.

REFERENCES

[1] Mao, W. Xu, Y. Yang, J. Wang, A. Yuille, "Deep captioning with multimodal recurrent neural networks (m-RNN)", Proc. Int. Conf. Learn. Representations, 2015.

[2] https://www.marutitech.com/problems-solved-machine-learning (10/6/2018).

[3] K. S. Fu, Syntactic Pattern Recognition and Applications, Prentice-Hall, Englewood Cliffs, N.J., 1981.

[4] K. S. Fu and A. Rosenfeld, "Pattern Recognition and Image Processing," IEEE Trans. Computers, Vol. C-25, No. 12, Dec. 1976, pp. 1336-1346.

[5] Marilyn B. Troyer (1988) Issues and Problems in Teacher Computer Literacy Education, Journal of Research on Computing in Education, 21:2, 141-154, DOI: 10.1080/08886504.1988.10781867.

Online shopping Engine to search commodities

Saurabh Rai,Varun Shah, Drishti Sisodiya, Riddhi Vaidya, Shiwani Gupta

Department of Computer Engineering, Thakur College of Engineering and Technology saurabhrai251@gmail.com, varunshah1218@gmail.com, drishti255@gmail.com, riddhivaidya16@gmail.com, shiwanigupta3005@gmail.com

***Abstract*— With the increasing demand of Internet and e-commerce, the number of shopping websites has rapidly increased. Consumers spend a lot of time looking for commodity as per their likes and preferences. In recent years, there are some websites helping consumers to buy cheaper commodity. Although these websites can help consumers get the comparative price of commodities, the search results are not so ideal. Therefore, this study attempts to use web mining technique as a basic approach. This study proposes a novel commodity search system to track consumer demand. Whenever the commodity price of any website is lower than the consumer price conditions, the system will proactively notify the consumers. The results indicate that the novel commodity search system could assist consumers to search commodity and provide historical price information of commodity for consumers to decide in an efficient manner.**

Keywords—Chat-Bot, Commodities, Consumer, Web Datasets.

I. INTRODUCTION

Nowadays, the advance of Internet and Web technologies has continuously boosted the prosperity of e-commerce. Through the Internet, it has become daily life for people to online shopping, and the number of people buying, selling and performing transactions on the Web is increasing at a phenomenal pace. With the further development of e-commerce, it will not be easy for customers to single out the best commodity when faced with the massive commodity information in the Internet. Usually, customers utilize various E-commerce search engines to search and compare commodities when they do online shopping in the Internet. Therefore, E-commerce search engines have largely become the main methods for customer to acquire commodity information and relevant services in the course of e-commerce activities. However, common search engines (such as Google, baidu, etc.) and keyword-based search are not only low-efficient, but also sometimes the retrieved document contents of web pages are non-relevant with customer's query. In order to solve the existent problems that traditional information search methods have, many scholars both at home and abroad propose many new web search approaches, which are based on ontology and semantic web. Popov defines a general framework for document search that is supported by ontology and integrate full-text search with ontology-based methods.

consumers to search commodity, long term track for consumer demand, and proactively notify track results. In this

Pablo presents a semantic search model based on an adaptation of the classic vector-space model, including an annotation weighting algorithm, and a ranking algorithm. Wang also presents a semantic-similarity based information search method, which utilizes ontology to describe semantics of the customer queries and Web documents and computes the semantic similarity between the concept and property of domain knowledge to realize the semantic information search. Zhang presents a semantic annotation method based on ontology, which can be used in intelligent search systems for semantic reasoning. All of these methods can solve the semantic inaccuracy of information to some extent and can realize the semantic accordance between the customer's query and document information. However, in order to solve the information overload of commodities and provide accurate information search and shopping service for customers, one kind of effective commodity evaluation and comparison mechanism should be constructed, which is based on the realization of semantic retrieval of general information. In order to realize it, the paper presents an intelligent commodity information search model of E-commerce search engine, which integrates semantic retrieval and multi-attribute decision method. First, semantic similarity can be computed by constructing query semantic vector and document semantic vector respectively based on ontology, in order to realize the semantic consistency between retrieved results and customer's query.

II. LITERATURE REVIEW

The current, for example Haptik where the user logs-in chat with the system and at the backend there is a person sitting at another end who replies to all the queries whereas here the reply will be given as per the availability and datasets and in case of haptik there should be a person available 24x7 to reply queries but in our case there is no such requirement.[1]work on opinion mining of mobile phone reviews on E-Commerce site: Flip kart. The paper presenters have also done a feature-based classification of reviews. The objective is to benefit the customers and assist them in choosing the right product. This in case [2] attempts to use Semantic Analysis, Ontology, and A.I as a basic approach. Which will analyze commodity prices, provide the relevant information to consumer's reference, will aid

the [3] presents an original commodity information search model, which integrates the semantic similarity computation

based on semantic vector-space model.[4] The search model not only realizes semantic retrieval of commodity information, but also has the commodity selection and comparison mechanism, so as to provide the intelligent information search services for customers' online shopping.[5]This study mainly focuses on Online Product Recommendation affect consumer shopping efficiency, how consumer shopping efficiency influence consumers' loyalty and how product type moderates the relationship between quality of Online Product Recommendation and consumer shopping efficiency.[6]This paper presents our work on opinion mining of mobile phone reviews on E-Commerce site: Flipkart .A feature based classification of reviews has been done by us. The objective is to benefit the customers and assist them in choosing the right product. As future work we propose to offer a summary of reviews for more than two products and also automatically rank products based on the features that the user is interested in. [7]. The recommendation system is designed to recommends a clothing combination suitable to a specified situation and the user preference under the condition that every clothing item should be recommended with equal frequency by using Bayesian networks and the feedback of recommendation output evaluation by the user.

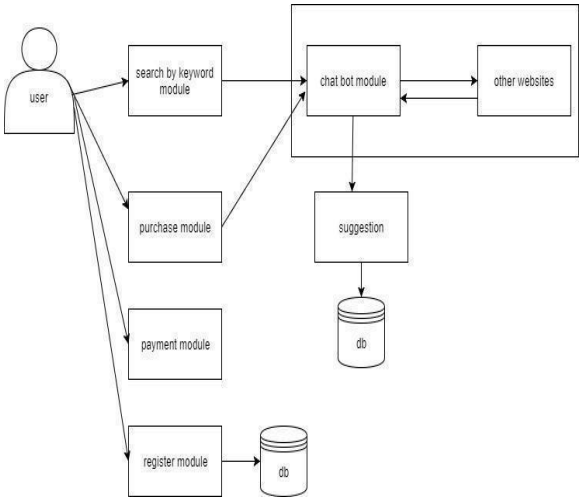
III. PROPOSED METHODOLOGY

At first, we will train our system by uploading dataset onto our system. The dataset will be containing data like Type of product, category, size, Age, Price. Based on these Factors the Systems will be trained. The System will be trained using Classification algorithm like NAIVE BAYES or ID3, etc. The Classification algorithm will generate a rule set onto the System. The User will provide input to the System like his product of his choice, category, size, age. Based on the product the System will predict the product and notify the User. Along with the Predicted Result i.e., Predicted the product, the System Will also suggest the related products or product of same variety etc. As our system also has Chabot which will interact with the user, it we'll have its own datasets which will be setting up at the file level.

This project mainly focuses on searching a commodity and finding best prices possible and enabling users to make quality decisions. The primary objectives of the project are as stated above, and the project aims at achieving all of them. With personalized search engines like Google and AltaVista which focused on searches based on keywords this project follows the same principle of searching based on keywords. It is basically a single software with multimillion use, this can be used to compare prices, get better deals or to know more about the product on a single platform. Through the study it is seen that ecommerce business is growing rapidly than a conventional business, which helps to optimize costs and a hassle-free shopping experience. It can be kept updated with all the new products or commodities advancing in the market. The more data collected on the user, the more personalized result will be. As in today's time where everyone has less time to search over other websites, this will lead to less consumption of time for searching of product.

The primary objective of the project is to create a one stop solution to find across all the ecommerce platforms and enhance the shopping experience of all the current users. Nowadays all the commodities are being searched before being purchased, therefore it is necessary that the search engine is well informed of all the products that the consumers want which is done using web mining. One of many objectives of the project is to reduce the time and effort of the consumer that he invests for searching a commodity on different platforms and provide a single platform to compare all the prices and quality across different ecommerce websites. Consumer gets an experience of hassle free shopping using a search engine. It makes use of web mining to search most relevant commodities according to a keyword. One of the most critical benefits is to improvise the quality of decision consumer makes. Using personalized search allows faster product and service discoveries for consumers and reduces the amount of necessary advertisement money spent to reach those consumers.

In our project as explained in incremental model, we are going to break to whole work into small modules. After breaking into modules, we are going to work on these modules which will increase the efficiency and faster execution of our project. Each module will pass through three phases of development (design, development, testing, and implementation). Planning, requirement gathering, and development will be done, after development of each module, testing of modules will be done to ensure that there are no bugs in the module. At the end implementation of each module will be done independently and after successful implementation all modules will be merged and the whole project will be released for customer feedback. Customer will give feedback and accordingly changes will be done if needed or else it will be released maintenance update from time to time. As we are implementing incremental model it will be easy to add any module in future. But before adding it will have designed, developed, tested and implemented. Since it is an incremental model, iteration of each phase will take place based on customer's requirement. Every time a new version of the product will be released with new upgrades.



Algorithm for User side:

User enters the statement.

→ if statement is a casual talk like hello, hi; then Chabot will reply with a casual answer like hi, hello, etc.

→ else if the statement is related to the products; the chatbot will tell the statement, process it and find the keywords related to the product.

→ depending upon the keywords received after processing of statement, the Chatbot will search for the products related to those keywords.

→if products related to the keywords are found then it will be displayed on the front page of website.

→else the chatbot will show "No such products available".

User can add the products to cart by clicking on add to cart button on any product whichever he wants to add to the cart button on any product whichever he wants to add to the cart.

When the user adds all the products desired in the cart then from there he can further checkout with those products and place that order.

When order is placed the mail will be triggered to the admin as well as the user(confirmation mail).

Fig 2 Algorithm for User Side

Algorithm for Admin side:

Admin will enter the email id and password on login page.

→the email id and password will then be checked with the entries in the database.

→if the credentials are valid, admin is allowed access in the admin part. else admin will be asked to reenter the credentials

Admin can now add the products which are to be displayed to the user for selling.

→for adding the products, the admin will be provided with a form where he has to enter all the products details and submit the form.

→on successful submission of the form the product will get added to the database.

Admin will be provided with the page (view products)

→on the view products, all the products from the database will be fetched using the MySQL query.

→the admin has the rights to edit or delete the products before checking out.

When admin clicks on the edit button in view products,

→admin will be taken to the page which will have the form which was used for adding the products. Using this page admin can edit whichever parameter that needs to be changed.

If admin clicks on delete on view products,

→delete query will be executed on the product which will delete the product entry from database on successful deletion.

Admin can view all the orders placed

→all the orders placed by the user will be fetched from the order table database and will be displayed here.

When the admin clicks on logout all the session will be destroyed and the admin will be out directed to the login page

Fig 3 Algorithm for Admin Side

IV. RESULTS

The System interacts with user efficiently in a user- friendly manner which not only makes the shopping more convenient and Effective but also helps the system to keep the record of the shopping requirement for the user. The chatbot can be more effectively developed by using vast base of Artificial Intelligence domain and thus making it

more competitive against its rival in the market like Haptic.

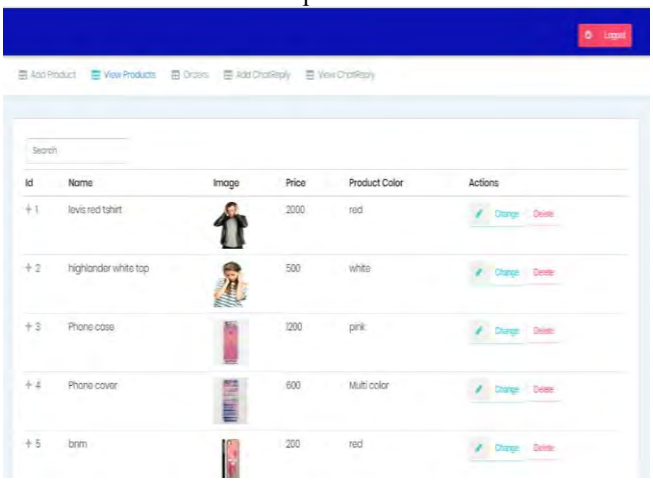


Fig 4 View Product Page

It can also be developed further and enhanced much efficiently not only available for shopping but also making it affordable for another task as well. The system at small scale can handle limited crowd but if it is implemented on large scale then number of servers need to be added thus to handle large crowd. The commodity information search algorithm also presents comparison and evaluation mechanism of the commodity contents and quality, not only considering semantic matching between searched document and customers' query keywords. The retrieval results of the commodity information search algorithm are more accurate, and more satisfy customers' shopping needs. The experimental results show that the algorithm presented here is superior to the traditional algorithm of semantic information retrieval and is more suitable for customers' shopping needs and also the custom algorithm used here is uneasy to detect and hack as per the security is concern because it will be changed as per the requirement of the system from time-to-time.

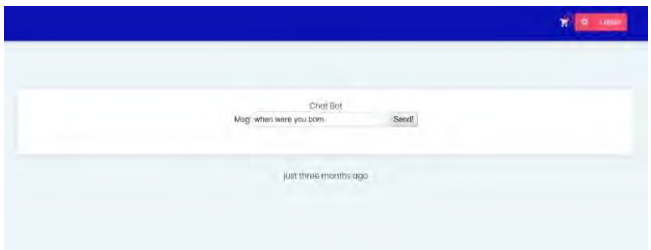


Fig 5 Chatting with Chatbot

V. CONCLUSION

The following project will give us the overview of the different commodities searched online with the help of AI chat-bot. It also gives us the related key searches of the commodities online based on the web datasets. The search

model not only realizes semantic retrieval of commodity information, but also has the commodity selection and comparison mechanism, to provide the intelligent information search services for customers' online shopping. The experimental results show that the algorithm presented in the paper is superior to the traditional algorithm of semantic information retrieval and is more suitable for customers' shopping needs. Based on present research work, we will improve the search model and make it to provide more accurate information search services for customers' shopping in the internet.

VI. REFERENCES

[1] Venkata Rajeev P,Smrithi Rekha V: Recommending Products to Customers using Opinion Mining of Online Product Reviews and Features ;2015.

[2] Ming-Hsiung Ying*, Yeh-Yen Hsu:A Commodity Search System for

Online Shopping Based on Ontology and Web Mining; 2014,International conference on software Intelligence Technologies and Apllication.

[3] Lin Yu-Chu Yuusuke, Kawakita Etsuko,Suzuki Haruhisa Ichikaw

Personalized Clothing-Recommendation System based on a Modified Bayesian Network; 2012 IEEE/IPSJ 12th International Symposium on Applications and the Internet.

[4] Lin, Q. X., Chang, C. H., & Chen, C. L. (2010) "A Simple and Effective Closed Test for Chinese Word Segmentation Based on Sequence Labeling" Computational Linguistics and Chinese Language Processing,Vol. 15, No. 3-4, pp.161.-180

[5] J. Liu, G. Z. Hu*, Y. Yu, W. J. Yi*, L. L. Zuo*: Researchon the Effect of the Recommendation System on Customer Online Shopping Experience,;2010, International conference on software Intelligence Technologies and Application.

[6] Ziming Zeng: Commodity Information Search Model of E-Commerce Search Engine Based on Semantic Similarity and Multi-Attribute Decision Method; Vol. 5, No. 7; July 2010 136 ISSN.

[7] Zhang gong-jie & Huang sui. (2008). Research and Implementation of Semantic Indexing based on Ontology. Computer and Engineering and Design, 29(8): 2078-2080.

Wireless Sensor Network For Environmental Monitoring And Control

Pankaj Malviya, Deep Modi, Kautilya Joshi, Zahir Aalam

Department of Information Technology Thakur College of Engineering & Technology, Mumbai, India

pankajmalviya81@gmail.com, modideep96@gmail.com, Kautilyajoshi97@gmail.com, zahiraalam@gmail.com

***Abstract*— This paper focuses on Wireless Sensor Network (WSN) technology for environmental monitoring and control. We illustrate here the technological difficulties and challenges that have been encountered in meeting end-user's requirements for information gathering. A WSN can monitor several environmental parameters such as underground water level, barometric pressure, ambient temperature, temperature measurement, atmospheric humidity, wind direction, wind speed and light intensity measurement. This in turn provides provide various convenient services for end users who can manage the data via a website from long-distance or applications in console terminal. This paper introduces the architecture of WSN system, the hardware of reduced function nodes and full function nodes, data acquisition, data processing and data visualization. In addition, the challenges of this technology in a field trial and the recorded data in terms of the sensor and communication reliability are addressed.**

Keywords- Wireless Sensor Network, ZigBee, Arduino, home automation, temperature measurement, humidity measurement, light intensity measurement.

I. INTRODUCTION

Wireless sensor network is a technology for wide range of wireless environments. Recently more research work has been done in direction to develop wireless network that works on low power, low data rate, low cost personal area network. Many organizations have developed WSNs for smart home, smart farm, smart hospital for patient monitoring, for traffic monitoring, fire monitoring in smart cities and home automation [1]. Sensor nodes fulfil tasks such as data gathering, data processing and communication and it makes it possible to follow the physical event. The sensor nodes collect data such as temperature, pressure, humidity, movement. Many sensor nodes, which distributed to the different areas and observed physical events, constitute to Wireless Sensor Networks. Basically, a wireless sensor network consists of a hundreds of sensor nodes that communicate wirelessly and spread over a certain area in order to monitor a physical event. Wireless sensor network has no infrastructure and topology and the sensor nodes are placed randomly. Therefore, Wireless Sensor Networks should be able to work different areas and different conditions and should be able to organize their own networks [2]. The importance and application has increased by the recent delivery of the IEEE 802.15.4 standard and the forthcoming ZigBee standard. The ZigBee Alliance has developed very low-cost, very low-power consumption, wireless communications standard for network and application layer to fulfil the demand of automation and remote control applications. IEEE 802.15.4 committee started working on a low data rate standard a short while later for physical and MAC sub layer. Then the ZigBee Alliance and the IEEE decided to join forces and ZigBee is the commercial name for this technology.

ZigBee is expected to provide low cost and low power connectivity for equipment that needs very long battery life as several months to several years but does not require data transfer rates as high as those enabled by Bluetooth. ZigBee can also be implemented larger networks than is possible with Bluetooth. ZigBee compliant wireless devices operate in the unlicensed RF worldwide (2.4GHz global, 915MHz Americas or 868 MHz Europe). The data rate is 250kbps at 2.4GHz, 40kbps at 915MHz and 20kbps at 868MHz.

Automating the control of various home appliances using this system facilitates human life, provides comfort and security to the user. In some cases, the automated devices also provide great energy consumption that helps user to decrease their expenses. Devices which are controlled can be anything that we use consistently in daily lives such as curtains, blinds, lighting (lamps, tube lights), air conditioning (heating, cooling), central heating boiler, small scale irrigation systems for personal gardens or even multimedia devices such as Televisions.

II. RELATED WORKS

A literature review is a text of a scholarly paper, which includes the current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic, issue, or area of research. In literature, many wireless sensor networks(WSNs) have been implemented for uses such as home automation, data mining and health care monitoring. Some of these are presented below:

A Smart home system was developed in [3] that monitored temperature, humidity, light and sound environmental parameters using Light Sensor. TSL2561 was used as SGSM Module for communication. We propose use of a more adequate light sensor for room temperature sensing with wider range of measurement and better accuracy. Our proposed system also notifies the user via SMS before switching ON/OFF the appliances connected to the proposed system.

It was found out in [4] that the hardware used in that paper consisted of electronic circuitry where a microcontroller is the principal processing unit. A graphical user interface is also implemented. A database was designed for comparison between past and present testing results. It was lacking multiple identical sensor nodes could be introduced. This would turn the master/slave (2 node star) topology into a mesh network. The communication range inside the mine can be increased by using smart mesh network.

In [5], findings as per the publications referred it states design of a real-time monitoring system (RTMS) for a patient by measuring the temperature using Temperature sensor (LM335Z) Real-time data is recorded which could be helpful for the corrective treatment by health care professionals. It can be more

reliable and extensible topology can be implemented within budget.

In [6], a wireless Sensor Network using TCN 75, PIC 16f877 and ZigBee was observed that responds to any emergency and inform appropriate individuals in a timely and cost effective manner. Better accuracy of sensor within adequate range of measurement with added feature of measuring humidity can be added to proposed system.

III. PROBLEM DEFINITION

There are several WSN hardware platforms available. In our prototype system, we chose our hardware platform to base on Arduino microcontroller board and an XBee S2 module. Our network consists of various motes with different sensors such as DHT11 for temperature and humidity sensing and BH1750 sensor for Light intensity sensing. Many organizations have developed WSNs for smart home, smart farm, smart hospital for patient monitoring, for traffic monitoring in VANET, fire monitoring in smart cities and home automation. The importance and application has increased by the recent delivery of the IEEE 802.15.4 standard and the forthcoming ZigBee standard.

1) After identifying the gap in Reference [1] the use of XBee series 1 module is substituted by XBee Series 2C module and the sensor TCN75 is substituted by the sensor DHT11 which will be added to the mote (sensor node). The DHT11 sensor provides a cheaper and adequate range of measurement as an alternative over the TCN75 sensor with added feature of measuring relative humidity.

2) After identifying the gap in Reference [2] the use of mesh topology is implemented over star topology as it provides higher degree of reliability as a failed node may result in automatic re-routing of the data packets. Use of DHT11 sensor is also used which provides digital output over the analog output provided by LM335A sensor with better accuracy. WAMP server is used to host web application over XAMPP server.

3) After identifying the gap in Reference [3] the use of XBee series 1 module is substituted by XBee Series 2C module which provides greater range, better power down current consumption and optional use of mesh topology.

4) After identifying the gap in Reference [4] the use of TSL2561 is substitutes by BH1750 which offers wider range of measuring capacity. The Web application is also used to send a prompt to the user to ask permission to switch ON/OFF the appliance instead of directly switching it ON/OFF and just notifying the user about it.

IV. SYSTEM OVERVIEW

We present the application of developing a wireless sensor network with temperature, humidity, and light intensity measuring capabilities. In this study, we have developed a small scale wireless network for automating home appliances by sensing environmental data and both, giving the user the control the devices according to his/her liking or putting the system to an auto mode which controls the connected devices in accordance with the data collected by the sensors by setting threshold values.

Firstly, we collect the environmental data by our sensor motes in areas we want to observe. The data collected by the sensor are then forwarded to the micro controller it's attached to (Arduino Uno in our case) which converts it into suitable form for wireless communication and forwards it to the wireless ZigBee modules. The reduced-function devices then send their respective data to the base station which uploads the received data to the website it hosts.

The website displays the real time data that the base station receives from various motes and it also display some information about the connected devices such as the last time the device was last switched ON/OFF and how much time has elapsed after that.

The website also contains an "AUTO" toggle button which enables the user to enable/disable the system to automate the connected devices according the environmental data received. When ON, the system has control over the devices too meaning that the system is able to turn ON/OFF the devices and notify the user when the observed environmental data values cross a threshold.

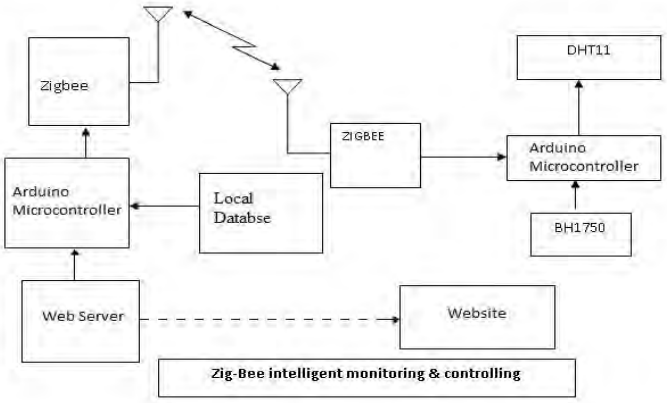


Fig 1: System block diagram

A block diagram is a diagram showing in schematic form the general arrangement of the parts or components of a complex system or process, such as an industrial apparatus or an electronic circuit.

The block diagram in figure 1 depicts the general layout of the proposed wireless sensor network. It comprises of two main parts, sensor nodes (motes) and the base station. The sensor nodes comprise of the sensor modules which allow real time environment monitoring. The collected data is forwarded by the sensor nodes to the base station that uploads it to the website.

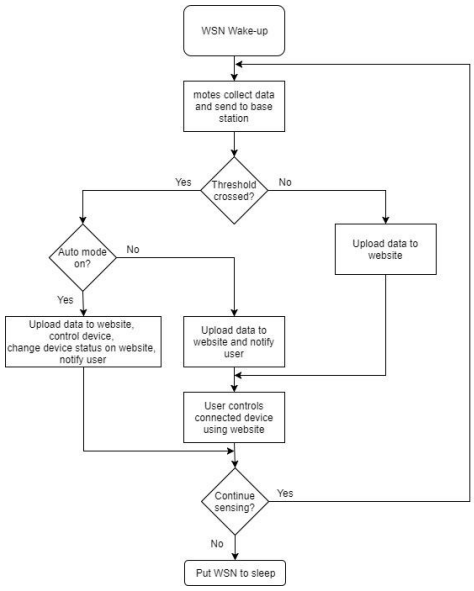


Fig 2: Proposed system data flow

A flowchart is a diagram that depicts a process, system or computer algorithm. They are widely used in multiple fields to study, document, plan, improve, and communicate often-complex processes in clear, easy-to-understand diagrams. The below figure is the flowchart that explains the basic working of our Wireless Sensor Network.

Figure 2 above depicts how our system keeps on monitoring the environment using various environmental sensors until it is switched off. One more thing about ZigBee modules is that they enter sleep mode whenever they are not required to sense or just forward any signal that leads to efficient energy consumption.

V. SYSTEM DESIGN

The system design of the proposed wireless sensor network comprises of the following core components:

- Temperature and Humidity sensor:
  The temperature and humidity sensor module used in the proposed system is the DHT11 digital sensor. DHT11 digital temperature and humidity sensor is a composite Sensor contains a calibrated digital signal output of the temperature and humidity. Application of a dedicated digital modules collection technology and the temperature and humidity sensing technology, to ensure that the product has high reliability and excellent long-term stability. The sensor includes a resistive sense of wet components and an NTC temperature measurement devices, and connected with a high-performance 8-bit microcontroller [6].



Fig 3: Temperature and Humidity sensor used in proposed system [8]

- Each DHT11 element is strictly calibrated in the laboratory that is extremely accurate on humidity calibration. The calibration coefficients are stored as programs in the OTP memory, which are used by the sensor's internal signal detecting process. The single-wire serial interface makes system integration quick and easy. Its small size, low power consumption and up-to twenty meters signal transmission making it the best choice for various applications, including those most demanding ones. The component is 4-pin single row pin package. It is convenient to connect and special packages can be provided according to users' request.

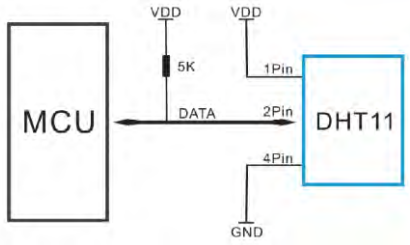


Figure 4: Block diagram of Temperature and Humidity sensors interface with microcontroller [9]

Note: Pin 3 – Null
MCU = Micro-computer Unit

- Light sensor board:
  This is a BH1750 light intensity sensor breakout board with a 16-bit AD converter built-in which can directly output a digital signal, there is no need for complicated calculations. This is a more accurate and easier to use version of the simple photo resistor which only outputs a voltage that needs to be calculated in order to obtain meaningful data. With the BH1750 Light sensor intensity can be directly measured by the lux meter, without needing to make calculations. The data which is output by this sensor is directly output in Lux (Lx). When objects which are lighted in homogeneous get the 1 lx luminous flux in one square meter, their light intensity is 1lx. Sometimes to take good advantage of the illuminant, you can add a reflector to the illuminant. So that there will be more luminous flux in some directions and it can increase the illumination of the target surface [7]. The BH1750 sensor board's major specifications are listed below:
  1. Light Range: 0 - 65535 lux
  2. Power Supply: 3.3V - 5V
  3. Sensor Built-in: 16-bit AD converter

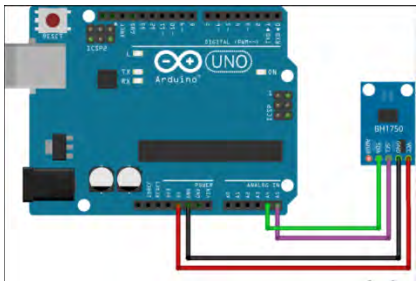


Figure 5: Light sensor board interface with microcontroller [10]

- ZigBee module:
  The Digi XBee S2 modules have been used in our system for wireless communication using ZigBee (IEEE 802.15.4/LR WPAN) standard. All the Reduced Function Devices (end nodes) send their respective data to the base station which in turn uploads it to the website. When not receiving or transmitting data, the RF module is in Idle Mode.

During Idle Mode, the RF module is also checking for valid RF data [11]. The module shifts into the other modes of operation under the following conditions:

1. Transmit Mode (Serial data in the serial receive buffer is ready to be packetized)
2. Receive Mode (Valid RF data is received through the antenna)
3. Sleep Mode (End Devices only)
4. Command Mode (Command Mode Sequence is issued)

- Base station:
  The base station comprises of a ZigBee module attached onto an Arduino microcontroller board. This node also hosts a website which the user can access to check the real time environmental data.
- Website:
  The website displays the real time environmental data collected by the sensory motes as well as current status of the connected devices. The website also has an auto button that enabled/disables the system to control the connected devices by itself without any human intervention.

VI. RESULT

The wireless system will help to monitor neighboring environment depending on the sensors used. The data collected by our system helps to control the devices that will result in us being able to control the environment as well.

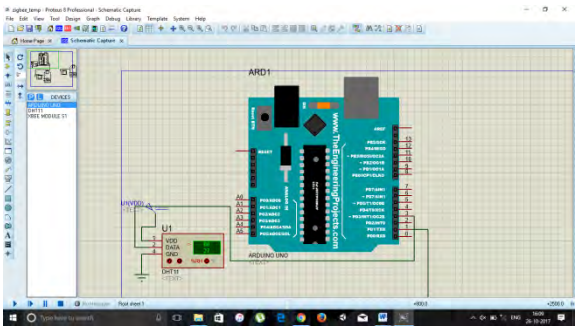


Figure 6: Simulation of temperature sensors working

The connection in a sensor node (mote) is shown in figure 6. The code in the above screenshot was written and compiled in Arduino environment and the resultant hex file's path was included inside the Arduino board's setting. BHT11 sensor is connected to the Arduino with the output from the DHT11 sensor being connected to Arduino's Digital PIN 3.

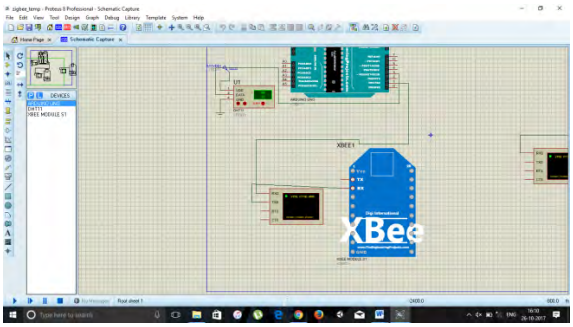


Figure 7: ZigBee module interfaced to microcontroller

The PD1/TXD PIN of the Arduino is connected to the virtual terminal's RXD PIN in this connection. Then the Virtual Terminals output pin, i.e., TXD is connected to the ZigBee module's input RX PIN.

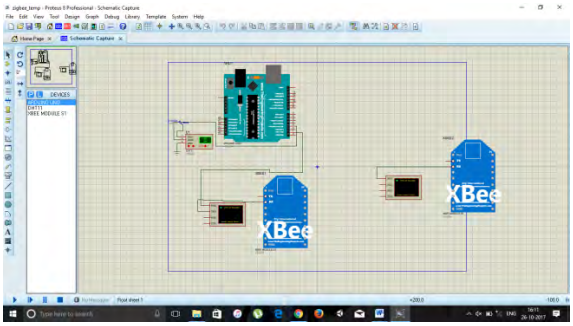


Figure 8: Proposed system layout depiction

The Sensing node in our simulation does the work of sensing the environmental data through the sensor, passing it on to the virtual terminal and then to the ZigBee module which transmits the data to the other ZigBee module on the receiving side. The purpose of receiving node in this simulation is just to display the real time data that achieved by connecting the virtual terminal directly to the ZigBee module of the second node.

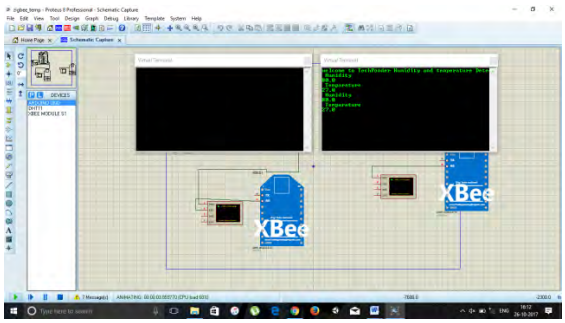


Figure 9: Base Station data being displayed on terminal

After running the above simulation, we can observe that the second terminal window in the diagram is displaying real time environmental data. The second widow is the second (receiver) node's terminal window. The first terminal window doesn't display the data because the data passed into it is forwarded to the ZigBee module via is output TXD PIN.

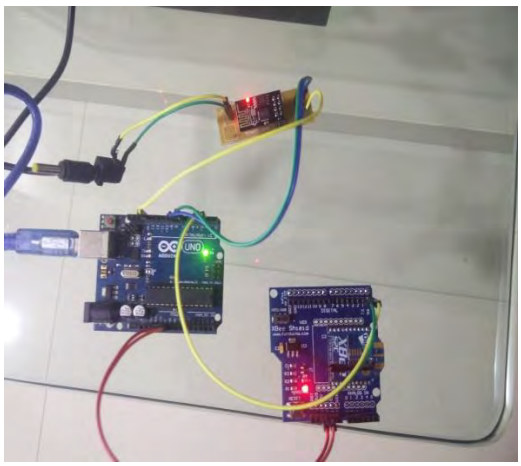


Figure 10: Base station of system

The above picture depicts the base station(master) node of the system that collects the data from all the sensor nodes and uploads it to the website. The ESP8266 IoT module interfaces with the website and upload it to the website.

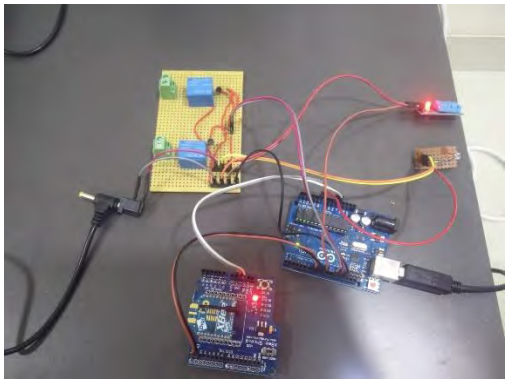


Figure 11: Sensor Node of the system

The above picture shows the connection of the sensor(slave) node with the DHT11 and the generic Light Dependent Resistor attached to it. It then forwards this data to the base station of the system.

A website was created in this research to view and access the real time sensor values. It was created ASP.NET and using Microsoft SQLServer 2017. The website's user interface is displayed below.

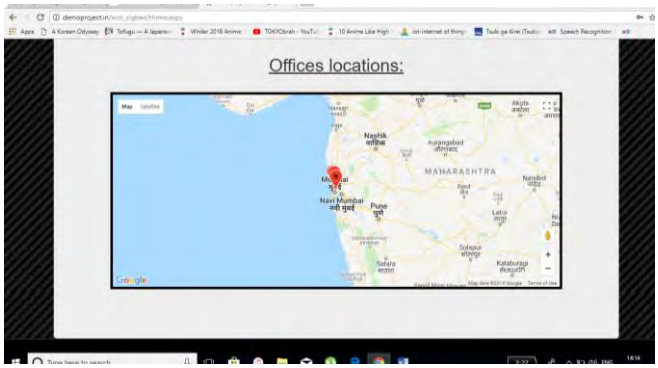


Figure 11: Homepage of website

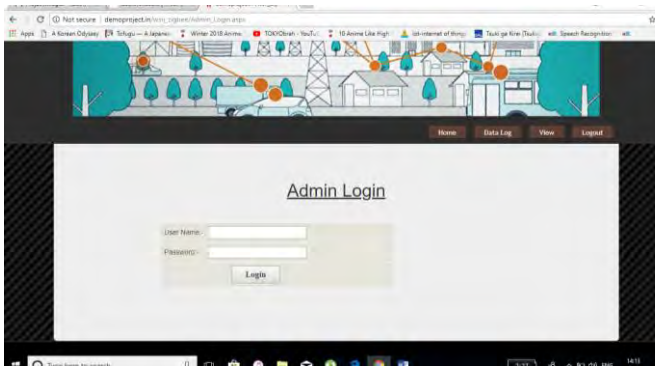


Figure 12: Admin login page of website

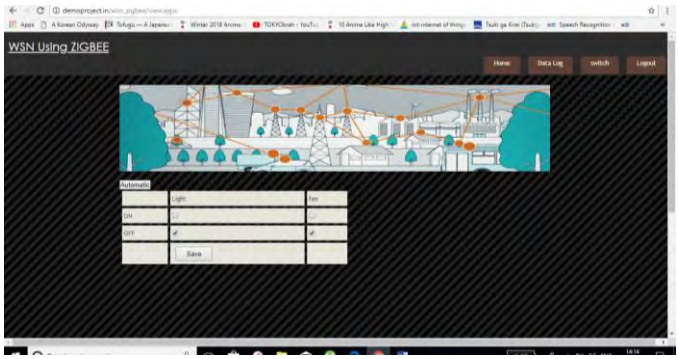


Figure 13: Control switches provided on the website

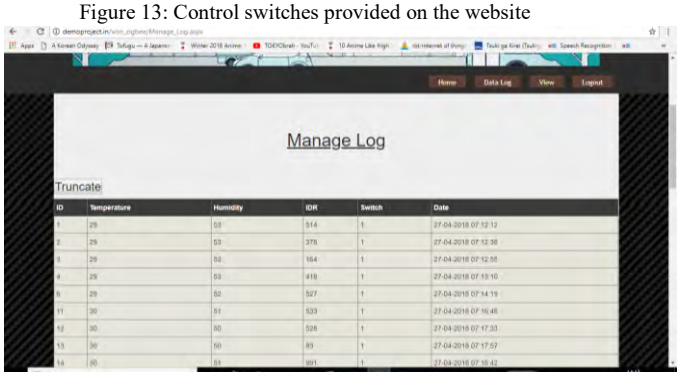


Figure 14: Sensor values displayed in tabular format in data log

VII. CONCLUSION

Through the various deployments, we have been able to thoroughly test our Wireless Sensor Network in real outdoor conditions, and it has proven to be robust and well fitted for environmental monitoring. Most of the problems we had to face were caused by hardware malfunctions and are all being addressed.

VIII. FUTURE SCOPE

The future work is trying to improve the topology structure to make all nodes communicate with each other, also to improve the stability of wireless sensors in communication by better software and hardware design. The proposed system is solely dependent on internet to allow the user to control the connected devices. This can create a bottleneck as not everyone has access to high speed internet in our country. This can be overcome by adding a GSM module for the user to control the connected devices directly via Short Message Service (SMS) without logging onto the website. This will further increase the usability of our system by allowing the user to interact with the devices in multiple ways in case on uncertainty

REFERENCES

[1] Nidhi Patel, Hiren Kathariya, Arjav Bavara, "Wireless Sensor Network using ZigBee", "International Journal of Research in Engineering and Technology", ISSN: 2319-1163 on Volume: 02 Issue: 06, June 2013

[2] Murat Dener, "Intelligent Home System Design Using Wireless Sensor Networks", "Research World International Conference, Istanbul, Turkey", ISBN: 978-93-86291-69-1 on Vol. 3, Issue. 2, 27th December 2016.

[3] Valdo Henriques and Reza Malekian (Member, IEEE), "Mine Safety System Using Wireless Sensor Network", "Institute of Electrical and Electronics Engineering Access", ISSN: "2169-3536" on Volume: 4, 16 June 2016, Pages: 3511 - 3521.

[4] Omar S. Alwan, K. Prahald Rao, "Dedicated real-time monitoring system for health care using ZigBee", "Institute of Electrical and Electronics Engineering Access", ISSN:" 2053-3713" on Volume: 4, Issue: 4, 8 2017, Pages: 142 – 144.

[5] S.S. Sarade, Prof. A.C. Joshi, Sachin S. Patil, A.N. Shinde, “Wireless Temperature Monitoring System Using Wireless Sensor Networks”, “International Journal of Advanced Research in Electronics and Communication Engineering”, ISSN: 2278 – 909X on Volume 1, Issue 4, October 2012.

[6] Temperature and humidity module DHT11 Product Manual
Available online on: https://akizukidenshi.com/download/ds/aosong/DHT11.pdf
Cited on: 15 January 2018

[7] Light Sensor board BH1750FVI
Available online on: https://www.dfrobot.com/product-531.html
Cited on: 22 January 2018

[8] DHT11 Humidity and Temperature Sensor
Available online on: http://www.micropik.com/PDF/dht11.pdf
Cited on: 15 January 2018

[9] DHT11 Humidity and Temperature sensor
Available online on: http://www.micropik.com/PDF/dht11.pdf
Cited on: 15 January 2018

[10] Light Intensity BH1750 Sensor for Arduino Raspberry Pi GY-302
Available online on: https://probots.co.in/index.php?main_page=product_info&products_id=753
Cited on: 22 January 2018

[11] XBee™ Series 2 OEM RF Modules
Available online on: http://www.farnell.com/datasheets/27606.pdf
Cited on: 1 February 2018

RFID Based Electronic Payment System

Shreya Mistry[1], Nishita Desai[1], Tanay Sanghavi[1], Ms. Shruti Mathur[2]
[1]IT Dept, Thakur College of Engineering and Technology, Mumbai, India
[2]Assistant Professor IT Dept, Thakur College of Engineering and Technology, Mumbai, India

Abstract

The project aims at creating a cashless payment system which will function properly in the college campus. It will handle all the transactions in all the parts of the college campus. This project will require proper working software as well as hardware. Connecting the software with the hardware is a challenging task. The RFID reader will have to accurately read the RFID card and update all the transactions and information in the system. However, it is essential to have accurate and proper transaction updates in the software.

Existing systems have used algorithms like AES, which are light-weighted. This project will be implemented using encryption algorithm present in SQL and also by writing a short algorithm of our own. This will ensure that the data stored inside the database will have proper encryption. It will also ensure that the data cannot be accessed or interpreted by any third person trying to get into the system.

The project requires the students to carry individual RFID cards. On the campus, at the time of any kind of transaction, the card will have to be placed on the RFID reader, which will be set up at various points in the college. The RFID card will have certain amount stored in it. Each time the card is placed over the reader, the reader will make changes in the data of the software. Once the changes are made inside the software, the transaction will be completed. Accordingly, the amount will wither be deducted or added into the individual account of the student.
The RFID reader and card provide efficiency and ensure that all the transactions are carried out in a hassle-free manner. The system aims at providing a complete cash-free environment all over the college campus. Thus, helping in, better management, keeping a record and conducting complete transactions. There by, helping in increasing theoverall efficiency of all transactions made inside the college campus.

I. Introduction

There have always been a lot of issues in transactions involving hard cash. In the college campus itself, there are numerous situations where the student needs to pay certain amount of money. The need to carry exact change is essential. Apart from this, there is always a danger of a student losing his wallet. However, these issues can be greatly managed by means of electronic payment. Electronic transactions are transactions which are made without the use of paper. Our project aims at providing an individual, unique card to each and every student. This card can be used to make payment in

various areas of the college campus, using RFID technology. This can greatly help in overcoming the problems faced by using hard cash.

II. PROPOSED WORK

This project presented a survey on RFID technology. RFID technology has a big potential to become ubiquitous in the near future. Today it

is already successfully used in supply chain management to track pallets of items. Tracking allows better coordination and control in the production cycle. Now the industry is pushing towards item-level tagging to increase the control even further. However, that also creates concerns, most common privacy concern, but also other security related issues. The paper presented possible scenarios how privacy can be compromised by RFID tags but also several solutions to protect against it. Since RFID technology becomes more and more common, attacks against the system itself start to appear. This paper listed the most common, starting from common sniffing and eavesdropping over denial of service to new RFID viruses.

The project concludes by looking at some social implications that RFID causes. Although not technically relevant, it provides a good outside perspective.

III. Literature Survey

According to the survey in the year 2014 it was reviewed that the security provided was 42.75% and expected security was 51.75% and RFID tag power was 10dB i.e. 100 microWatts. The research gap present in this is that the expected security can be increased by almost 10% to 61% and RFID tag power required could be less than 50 microWatts.[1]. In the year 2013 it was reviewed that high frequency RFID chips were used and for authentication lightweighted algorithm AES was used. The research gaps present in this are that we could use low frequency RFID chips instead of the high frequency ones. Also we could used heavy weighted algorithm in SQL[2]. In the year 2015 it was noticed that the recharge limit was Rs 200-2000. The research gap present in this is that the recharge limit is going to be Rs 1000[3]. In the year 2016 it was view that it made use of Web services and Ultra High Frequency and High Frequency cards were used and the frequency range for this cards was 125 or 134KHz. The research gap presented showed that web services is not used and low frequency cards were used which had a frequency range of 13.56MHz[4]. In the the year 2012 it was reviewed that active RFID cards was used with proximity of 100 metres and the frequency range was 455MHz, 2.45GHz or 5.8GHz. The cost of one card was Rs200. The research gaps presented in this is that pasive cards were used with the proximity of 10cm. The frequency range provided was 128KHz, 13.6MHz,915MHz or 2.45GHz. The cost for one card was Rs75[5].

IV. Problem Definition

Phase1:

Planning: The entire flow will be decided.
Requirement Gathering: The project will be planned and requirements will be enlisted
Designing: The front end will be ready,the design of all the phase will be done.

Phase2:

Coding: The code will be written and entire software code will be targeted to complete.
Encryption: DES algorithm will be used.

Phase 3:

Coding: The backend database will be created and connection will be established between front end and backend
Hardware Gathering: The hardware required will be procured

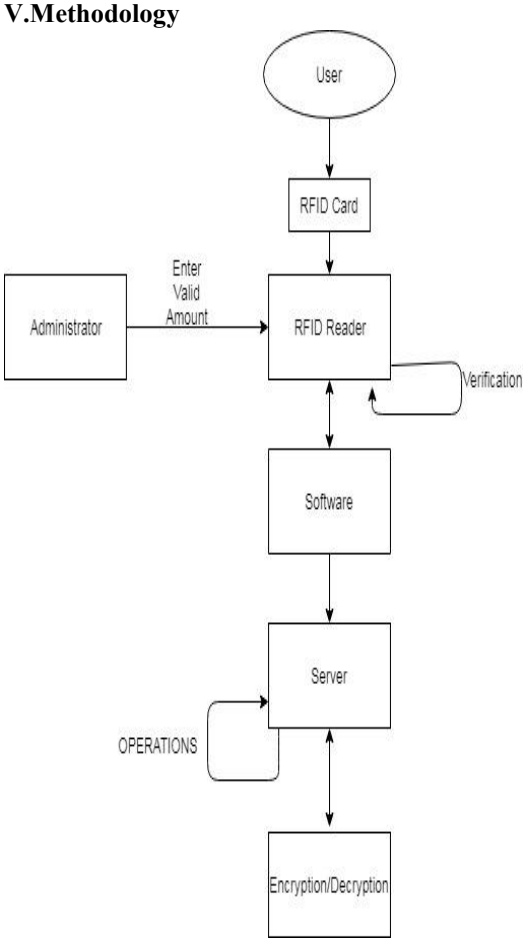
Phase 4:

Hardware-Software Compatibility: the hardware will be set up according to the software. Connection between them will be established
Testing: The connection will be tested and changes will be made accordingly

Phase 5:

The entire project will be tested and further required changes will be done .

V.Methodology



VI. Feasibility

RFID based Electronic Payment System:

Operational Feasibility:

There is no hardcoded input like mathematical formula and complex functions. The RFID reader will read the card which has been placed on it and will authenticate the student. If the student is authenticated, the amount entered will be deducted from his account. It will in all make transactions quickerand hassle free.

Economic Feasibility:

The program can be executed on computer of normal configuration of at least 4 GB ram The project can be executed independently that of other programs or modules and there is no need to install any extra dependent modules or programs to run this project without facing overhead or load on System.

Technical Feasibility:

All the complex algorithms and techniques are hidden from user and all input/outputs can be executed without much technical knowledge. All the algorithms and techniques are optimized to yield best possible results.

Legal Feasibility:

All the programs are developed using either open source software or freely available framework without much licensing issue. All the Framework used to develop allows us to use application developed to use and integrate anywhere free of cost.

Social Feasibility:

This Project can be helpful for implementing the 'cashless' system in the college campus. This Project will be useful the students to perform hassle-free transactions on the campus of the college.

VII. Outcomes

Phase1:

Planning: The entire flow will be decided.
Requirement Gathering: The project will be planned and requirements will be enlisted
Designing: The front end will be ready, the design of all the phase will be done.

Phase2:

Coding: The code will be written and entire software code will be targeted to complete.
Encryption: DES algorithm will be used .

Phase 3:

Coding: The backend database will be created and connection will be established between front end and backend
Hardware Gathering: The hardware required will be procured

Phase 4:

Hardware-Software Compatibility: the hardware will be set up according to the software. Connection between them will be established
Testing: The connection will be tested and changes will be made accordingly

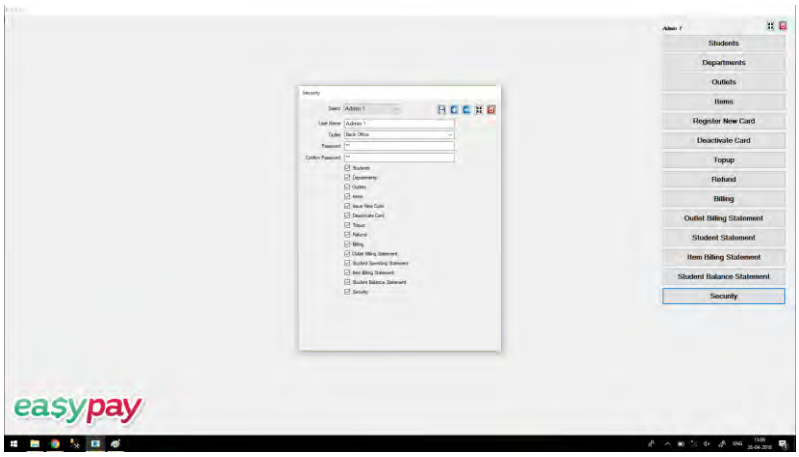
Phase 5:

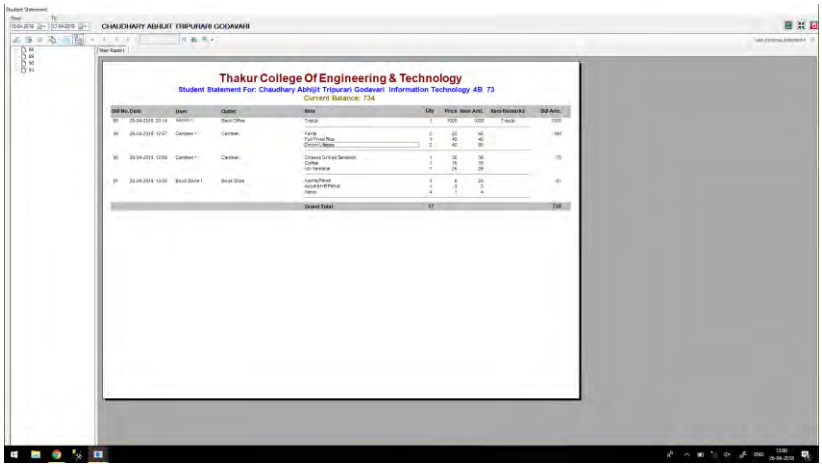
The entire project will be tested and further required changes will be done.

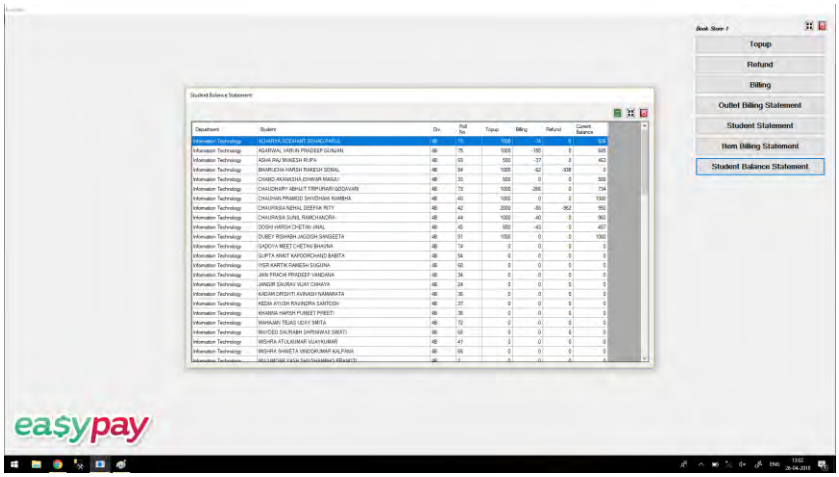
VIII. Improvements in Research gaps

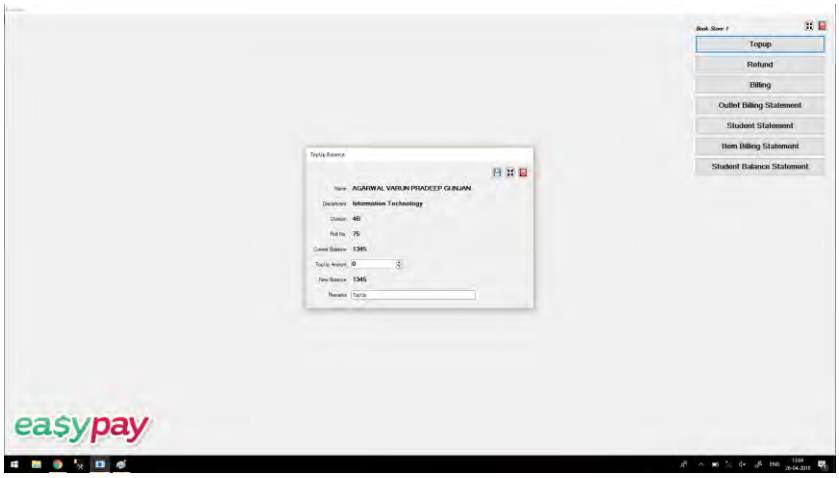
The real challenge in maintaining such payment systems is maintaining security in them. The research papers use AES algorithm, where as we use DES algorithm. Therefore, the security results will be different than the other projects. Along with the algorithm, we will also be implementing a small encryption algorithm of our own in order to increase the overall security of the system.

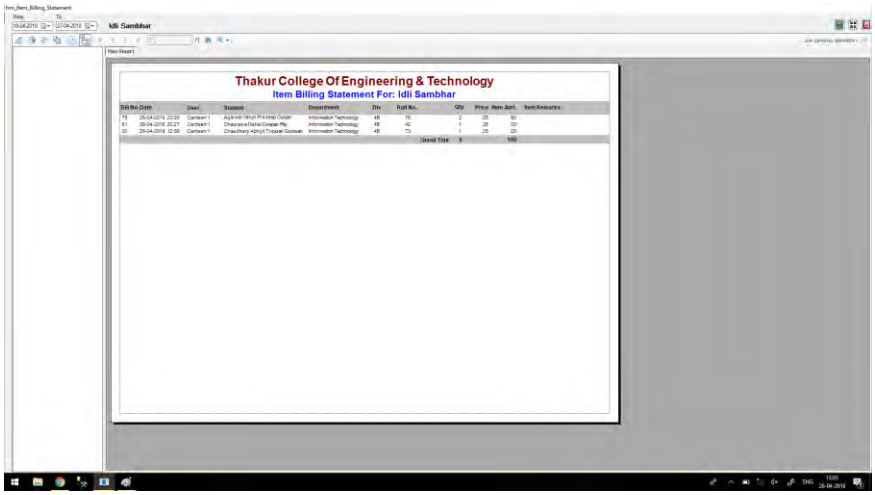
The previous system could achieve security up to 51.75%, whereas, our system aims at achieving 61% security.

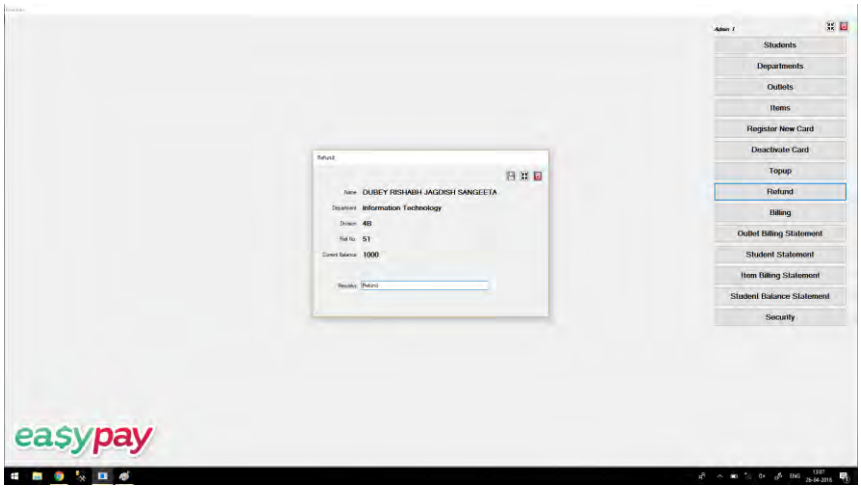

Students
Departments
Outlets
Items
Register New Card
Deactivate Card
Topup
Refund
Billing
Outlet Billing Statement
Student Statement
Item Billing Statement
Student Balance Statement
Security
easypay

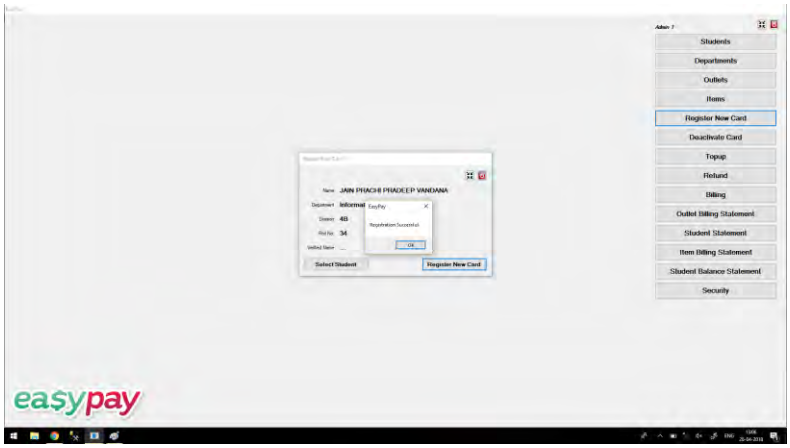

CHAUDHARY ABHIJIT TRIPURARI GODAVARI
Thakur College Of Engineering & Technology
Student Statement For: Chaudhary Abhijit Tripurari Godavari Information Technology 4B 73
Current Balance: 734

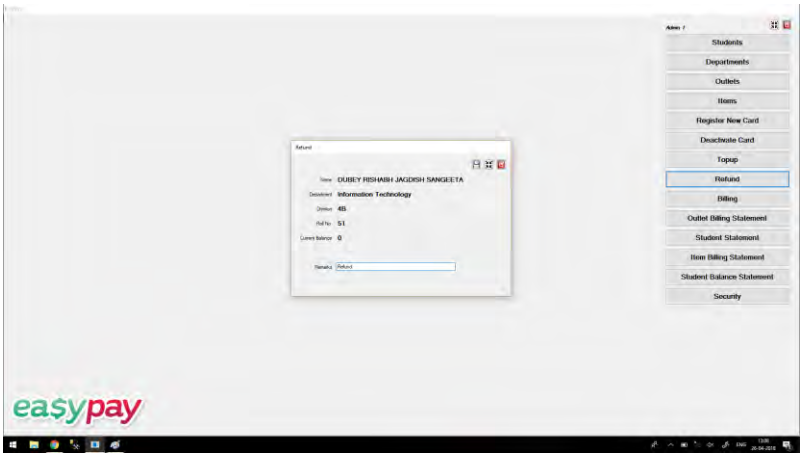

Topup
Refund
Billing
Outlet Billing Statement
Student Statement
Item Billing Statement
Student Balance Statement
easypay


Topup
Refund
Billing
Outlet Billing Statement
Student Statement
Item Billing Statement
Student Balance Statement
AGARWAL VARUN PRADEEP GUNJAN
Information Technology
easypay


Idli Sambhar
Thakur College Of Engineering & Technology
Item Billing Statement For: Idli Sambhar


Students
Departments
Outlets
Items
Register New Card
Deactivate Card
Topup
Refund
Billing
Outlet Billing Statement
Student Statement
Item Billing Statement
Student Balance Statement
Security
DUBEY RISHABH JAGDISH SANGEETA
Information Technology
1000
easypay


Students
Departments
Outlets
Items
Register New Card
Deactivate Card
Topup
Refund
Billing
Outlet Billing Statement
Student Statement
Item Billing Statement
Student Balance Statement
Security
JAIN PRACHI PRADEEP VANDANA
Select Student
Register New Card
easypay


Students
Departments
Outlets
Items
Register New Card
Deactivate Card
Topup
Refund
Billing
Outlet Billing Statement
Student Statement
Item Billing Statement
Student Balance Statement
Security
DUBEY RISHABH JAGDISH SANGEETA
Information Technology
easypay

IX. References

1. Life Science Journal 2014;11(1) Department of Computer Science, Faculty of Computing and Information Technology,
2. King Abdulaziz University, Jeddah 21589, Saudi Arabia.
3. International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 200- 203 (2015)
4. International Journal of Modern Trends in Engineering and Research
5. RFID: Technology and Applications by Sridhar Iyer
6. www.impinj.com
7. www.zebra.com
8. barcodesinc.com
9. Security analysis of RFID based devices in educative environments, Ahmed Alzahrani, 2014.
10. A Survey paper on Radio Frequency Identification(RFID) trends, ChristophJechlitschek, 2013.

Raspberry Pi Based Cost Effective Interactive Device for Blind, Visually Impaired, Deaf and Dumb People

Vijay Karotra
Department of Information Technology,
Thakur College of Engineering and Technology, Mumbai,
vj.gerrard@gmail.com

Shubham Maheshwari
Department of Information Technology,
Thakur College of Engineering and Technology, Mumbai,
shubh.maheshwari01@gmail.com

Tanishq Mehra
Department of Information Technology,
Thakur College of Engineering and Technology, Mumbai,
tanishqmehra97@gmail.com

Shridhar Kamble
Department of Information Technology,
Thakur College of Engineering and Technology, Mumbai,
shridhar.kamble@thakureducation.org

Abstract - Helping differently able people (visually impaired, physically challenged, deaf and dumb) is a difficult task as there is no solution offering a common platform or a solution that is cost effective. A solution is proposed where four modules viz - text to speech (dumb), gesture recognition (physically challenged), image to voice (visually impaired) and voice to text (deaf) are integrated onto a Raspberry Pi to provide a solution which is both, based on common platform and cost effective. The image to voice module is based on cutting edge neural nets (CNN and LSTM) to read an image and output the description. The gesture recognition module captures the image, identifies the gesture and relays the output by cross checking the gesture with predefined meanings. Both voice to text and text to speech are simple modules which work on natural language processing APIs. The Raspberry Pi is used as its small size and cheap cost suit the project requirements very well.

Keywords - Image-to-voice, gesture-recognition, speech-to-text, text-to-speech, CNN, LSTM.

I. INTRODUCTION

According to the World Health Organization, there are more than 39 million legally blind and 285 million visually impaired people living across the world, and this number is still growing at an alarming rate [1]. As currently most of the camera based devices are capable enough to capture photographs with high resolutions, we can make use of these images to generate captions based on graphics present on images which can be speak out for visually impaired people so that they also have look and feel where they are actually, what is their surrounding environment and finally get a better visualization of what is happening around them. The work focuses on finding a unique technique that aids the visually impaired by letting them hear what is represented as text and it is achieved by the technique that captures the image through a camera and converts the text available as voice signals. There are many existing solutions for assisting blind or visual impairments people but actually they are not affordable and not readily available to the community of visual impairments or blind peoples. It's a very tough job to address the issues of People with Visual, Hearing and Vocal Impairment through a single aiding system. The work focuses on gesture recognition and finding a unique technique that aids the visually impaired by letting them hear what is represented as text and it is achieved by the technique that captures the image through a camera and converts the text available as voice signals. Proposed system will describe the content of an image using well-formed English sentences for better understanding of their surroundings which will have great impact on them.

II. OVERVIEW

There are many existing solutions for assisting blind or visual impairments people but actually they are not affordable and not readily available to the community of visual impairments or blind peoples. It's a very tough job to address the issues of People with Visual, Hearing and Vocal Impairment through a single aiding system. The work focuses on gesture recognition and finding a unique technique that aids the visually impaired by letting them hear what is represented as text and it is achieved by the technique that captures the image through a camera and converts the text available as voice signals. Proposed system will describe the content of an image using well-formed English sentences for better understanding of their surroundings which will have great impact on them.

III. LITERATURE SURVEY

Text to speech module uses natural language processing to convert letters into phonetics which are combined to create speech [2]. Gesture recognition uses convex hull and adaptive boosting in OpenCV [3]. Image to voice is done in two steps first a CNN is used for feature extraction and the identified features are expressed in a sentence using a LSTM [4]. Speech to text uses an isolated word recognition with limited vocabulary [5].

IV. PROBLEM DEFINITION

A. Phase 1(Information Gathering):

Gather appropriate data set to train the neural networks. All available APIs for speech to text and text to speech need to find. Most important gestures and their meaning need to be gathered as well.

B. Phase 2(Analysis & Design):

The various algorithms to train the neural net are identified and compared and the best one is selected. The numerous APIs available for speech to text and text to speech need to be compared. Various possible methods of gesture recognition are compared to find the one best fitted for our use.

C. Phase 3(Implementation & Testing):

Separate implementation of each of the modules i.e. Caption generation, Text-to-speech, Speech-to-text and Gesture recognition and checking the efficiency of each of the modules. Finally, integration of the modules with the required hardware equipment and testing of the system as a whole.

D. Phase 4(Deployment):

Deployment of the integrated modules into Raspberry Pi after connection with the equipment's including WebCam, Zigbee, microphone, speaker, and screen. A full-fledged system will be provided which allows access to all the modules as per requirement.

V. METHODOLOGY USED

The figure below shows the connections to the Raspberry Pi. The webcam is to take the input for the Image Caption Generation and Gesture Recognition. The speaker acts as a output for the Image Caption Generation and the Text-to-Voice modules. Microphone takes the input for the Voice-to-Text which is finally displayed through the Monitor.

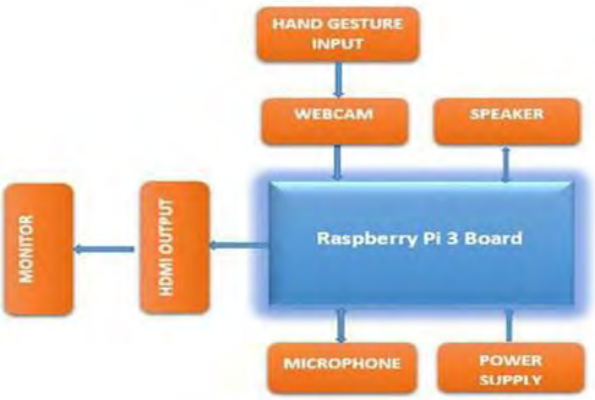


Fig. 1: Architecture of Raspberry-Pi based Automated Device

The speech to text module will be implemented using available open source APIs. A mic will be used for input and the result will be displayed on a screen once the input has been processed.

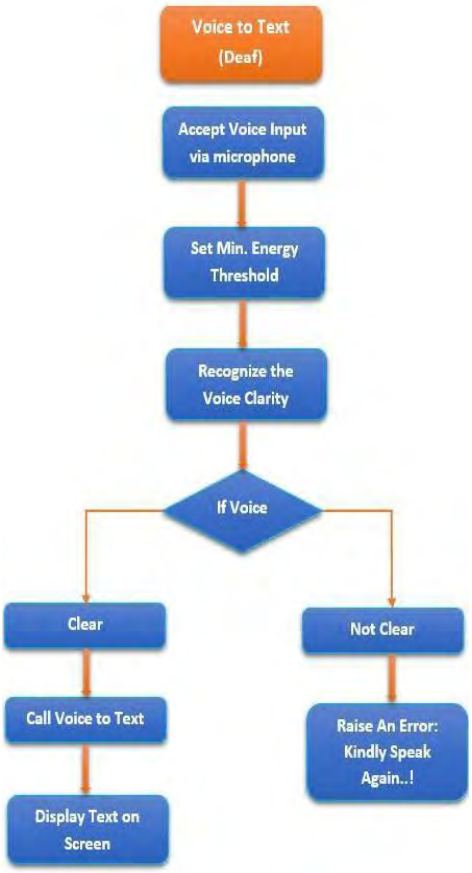


Fig. 2: Flowchart for Speech to Text

The text to speech module will also use available open source APIs. The input in this module will be a keyboard and the output will be on speakers.



Fig. 3: Flowchart for Text to Speech

The gesture recognition module captures an image from a video then the boundaries between the fingers is detected. Then we identify the region of interest and the gesture in that region is recognized and the corresponding predefined message is outputted on the screen.



Fig. 4: Flowchart for Gesture Recognition

The image to voice module works in two phases and each of these phases use a neural net. The first phase uses a Convolutional Neural Net (CNN) which is used for feature extraction from the input image. The output of this phase is are the objects which have been identified and they serve as the input of the next phase. The second phase uses a Long Short-Term Memory (LSTM) which converts the given words into a meaningful sentence. The dataset used in Flickr30k image set.

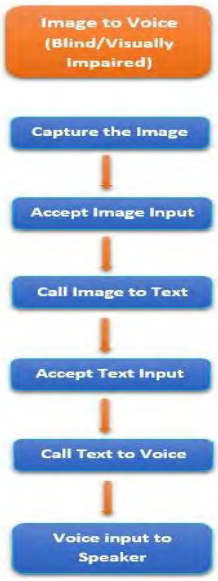


Fig. 5: Flowchart for Image Caption Generation

All of these modules are finally integrated on a Raspberry Pi via which the users will be able to interact with the modules.

VI. TECHNOLOGY USED

I. Python is used for scripting where the neural nets are programmed and trained and the APIs required for other modules are called.

II. All the modules are integrated onto a Raspberry Pi with Raspbian OS installed.

VII. RESULTS

A. Gesture Recognition:

Live streaming of camera was carried out through initialization. For effective finger recognition, the extracted images were compared with stored positive-negative integral image dataset fingertip tracking was performed by contour detection.

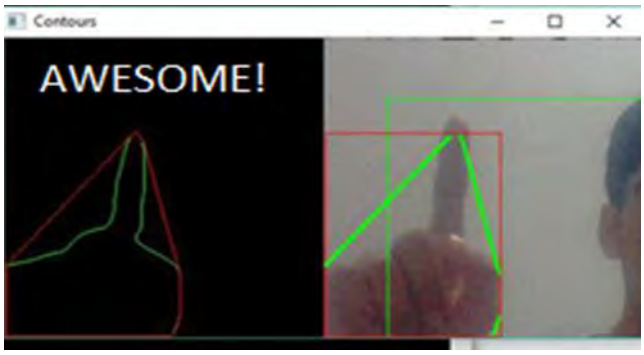


Fig. 6 Output for Finger 1

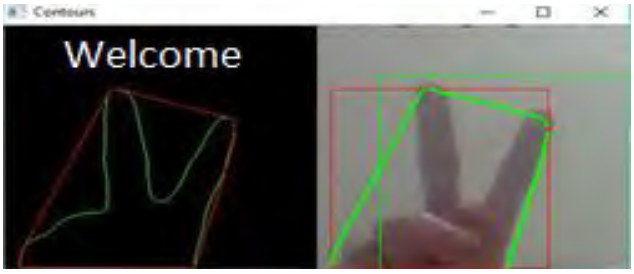


Fig. 7 Output for Finger 2

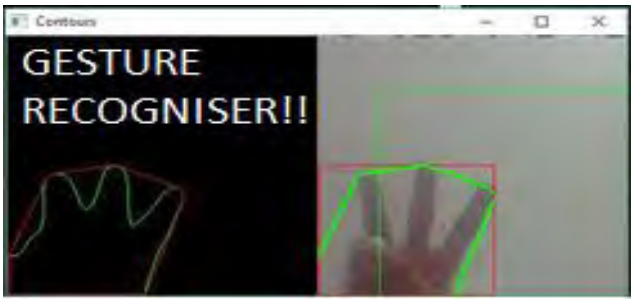


Fig. 8 Output for Finger 3

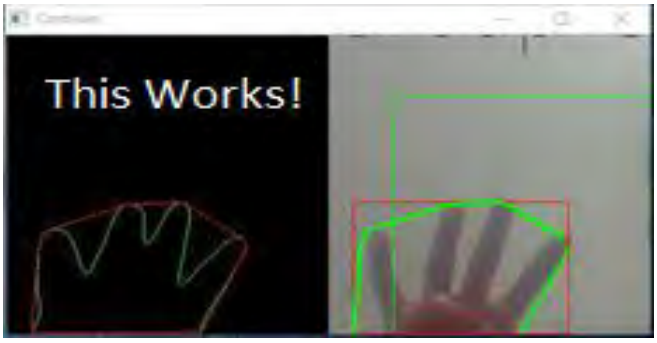


Fig. 9 Output for Finger 4

Table 1: Result for Gesture Recognition

	One Finger	Two Finger	Three Finger	Four Finger
Total No. of Trials	45	43	45	40
No. of correct finger gesture	42	40	44	38
% accuracy	93.33	93.02	97.78	95

B. Speech to Text:

Google API was used for effective speech to text conversion. The input was given via a microphone which was then processed by the Google's API and the output is then printed on to the screen.

The given input was as follows: Hey! This is a test for Google API Speech to text. In this module we are trying to convert the outspoken speech into readable text.

The output for the same input is shown in the figure below.

```
F:\Code\Personal\Major_Project\Code>python google_spee
ch.py
Say something!
Google Speech Recognition thinks you said hey this is
a test for Google API speech to text in this module we
 are trying to convert the outspoken speed into a read
able text

F:\Code\Personal\Major_Project\Code>
```

Fig. 10 Output for Speech to Text

C. Text to Speech:

The gTTS Package is used to convert a given input of text into a speech. The program accepts text as an input which is then converted into speech and saved in a mp3 format.

The following text was given as the input: "Look, I didn't want to be a half-blood. If you're reading this because you think you might be one, my advice is: close this book right now. Believe what-ever lie your mom or dad told you about your birth, and try to lead a normal life. Being a half-blood is dangerous. It's scary. Most of the time, it gets you killed in painful, nasty ways. If you're a normal kid, reading this because you think its fiction, great. Read on. I envy you for being able to believe that none of this ever happened. But if you recognize yourself in these pages-if you feel something stirring inside-stop reading immediately. You might be one of us. And once you know that, it's only a matter of time before they sense it too, and they'll come for you."

The resultant output was a 58 seconds long mp3 file which correctly recites the paragraph verbatim.

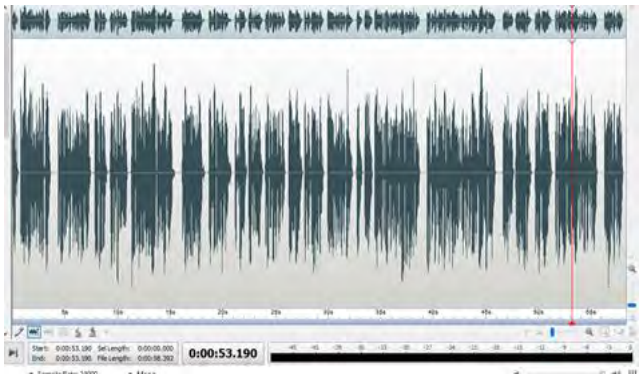


Fig. 11 Output for Text to Speech

D. Image to Speech:

The two neural networks are used to generate the caption, a convolution neural network and long short term memory.



Caption: A man and a dog are walking through snow

Fig. 13 Output for Image to speech

The above figure has the test image followed by the generated caption. It is clear that the model has been trained properly and the results are up to the mark.

Table 2: Result for Image Caption Generator

BLEU LEVEL	SCORE
BLEU-1	65.0945536216
BLEU-2	47.3736538727
BLEU-3	32.7219332641
BLEU-4	22.8155010438

VIII. CONCLUSION

It's a challenging task for physically impaired people (blind, deaf or dumb) to communicate with rest of the world. Blind or dumb people generally use sign language for communication. But in reality very few people understand sign language. Proposed system aims to lower this barrier in communication by developing a portable electronic device which will act as a smart assistant for physically impaired people. Primary goal of proposed system is to provide a standard lifestyle for physically impaired people. In a proposed system prototype model for blind, visually impaired, deaf and dumb people has been designed by combining four modules in a single compact device. The important key factor of this project to facilitate physically impaired people and provide them opportunity to interact with rest of the world with the help of technology. Device would be portable and less in weight.

IX. IMPROVEMENTS IN RESEARCH GAPS

The gesture recognition module allows for the accuracy of 70% which is at par with that achieved in the literature survey. Better algorithm is being worked upon which helps enhance the contour detection by the process of adaptive boosting which will make the camera emphasize more on the hand gestures as compared to the surrounding in the field of vision of the camera.
The image to voice module is currently working only on images. If we have better computing power and real time data we can apply the same concepts to video and thereby increasing the usability of the module.
The speech to text module which works on Google Speech Recognition does not take into account the usage of punctuations which can make the sentence difficult to understand in certain situations. Larger set of parameters and proper usage of time gaps between words can help improve structuring the sentences and make them grammatically more correct.

X. REFERENCES

[1] "Visual Impairment and Blindness." World Health Organization. (2014). Web. 10 Apr. 2016

[2] Dragos Burileanu, "Basic Research and Implementation Decisions for a Text-to-Speech Synthesis System in Romanian", International Journal of Speech Technology 5, pg. 211–225, 2002

[3] Premanand K. Kadbe, Ruchi Manish Gurav, "Real time finger tracking and contour detection for gesture recognition using OpenCV", International Conference on Industrial Instrumentation and Control (ICIC), Vol.3 Issue 5, March 2015

[4] Christopher Elamri, Teun de Planque, "Automated Neural Image Caption Generator for Visually Impaired People", Stanford University, 2017

[5] [Nuzhat Atiqua Nafis and Md. Safaet Hossain, "Speech to Text Conversion in Real-time", International Journal of Innovation and Scientific Research, Vol. 17 No. 2 Aug. 2015, pp. 271-277

[6] pypi.python
   a. https://pypi.python.org/pypi/SpeechRecognition/ accessed on 19/7/2017

[7] pythonprogramminglanguage
   a. https://pythonprogramminglanguage.com/text-to-speech/ accessed on 19/7/2017

[8] Google Drive, Flickr 30k dataset
   a. https://drive.google.com/file/d/0B_PL6p-5reUAZEM4MmRQQ2VVSlk/view accessed on 18/7/2017

[9] Yin, X., Zuo, Z., Tian, S., Liu, C.: Text detection, tracking and recognition in video: a comprehensive survey. IEEE Transaction, Image Processing, 2016

[10] Jaderberg, M., Vedaldi, A., Zisserman, A.: Deep features for text spotting. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 512–528. Springer, Heidelberg, 2014

[11] Chucai Yi, Student Member, IEEE, Yingli Tian, Senior Member, IEEE, and Aries Arditi "Portable Camera-Based Assistive Text and Product Label Reading From Hand-Held Objects for Blind Persons" 2013 IEEE.

[12] Bien, Z.; Park, K.-H.; Jung, J.-W.; and Do, J.-H. 2005. Intention reading is essential in human-friendly interfaces forthe elderly and the handicapped. IEEE Transactions on Industrial Electronics 52(6):1500–1505.

Comparison of time series forecasting model for predictive analysis

Munaf Patel, Zahir Aalam

Department of Information Technology Thakur College of Engineering & Technology, Mumbai
pmunaf123@gmail.com

***Abstract*—Forecasting is the method of making predictions of the future based on past and present data and most commonly by analysis of trend. There are several methods for making forecasts, but they all fall into two categories: causal methods and time series methods. In many cases, predictive algorithms implementing time series are good candidates for forecasting. In this paper we run a comparative study of three of these algorithms: Linear Regression, Multi regression and Moving Average in order to determine their performances in term of implementing times series for predictive systems. To assess the performance of these algorithms, we have conducted experiments over four representative datasets. The results exhibit that Multi regression produced the better forecasts. The other two algorithms show a good behavior as well**

General Terms—Data mining, Time series Analysis

Keywords—Decision Support, Comparative study, Time Series

I. INTRODUCTION

Day after day, decision support systems are becoming increasingly important in all companies, regardless of their size. The support systems for decisions cover a number of areas: Business intelligence, data mining and analytics... And so on. The possession of a software package that manages these areas is an important asset for the company in the era of information systems.

The present work is particularly interested in analytical processes, the predictive system, which can predict and forecast the activities of companies. Projection methods fall into two categories: Causal methods and methods of time series. In this paper, we want to make forecasts using the methods of the time series.

Time arrangement estimating is one of the foremost essential prescient analytics needs of numerous businesses. Numerous information components are watched as time arrangement. These may be item deals, product costs, temperature measures and so on. From a vital viewpoint, directors and decision makers will as often as possible ought to be able to foresee patterns and regular designs for these components [1].

As an example, a product manager may want to forecast sales for the next month so that he can schedule labor, raw material orders and equipment availability to meet the estimated demand. Similarly, a transportation manager may want to forecast shipping costs for the next quarter so that he can prepare and present a budget to his profit and loss team about estimated costs for his department. As a third example, a purchasing manager may want to predict the future costs of regularly ordered commodities so that he can quickly turn around and help the sales teams to prepare price quotes for products they manufacture using these commodities [2]. Time series forecasting is one of the entry-level predictive analytics processes that businesses need to develop confidence in implementing and deploying

Our motivation in this paper is to make a similar investigation of prescient calculations actualizing time arrangement to pick the reasonable one so as to assemble a prescient framework that can give expectations and gauges about the organizations exercises to support chiefs and leaders to settling on vital choice.

The rest of this paper is organized as follow. In section II we present a brief review on time series analysis and forecasting. Section III describes the three predictive algorithms used for our comparative study. In the section IV we present and discuss the achieved experimental results. Finally, section V concludes this work.

II. TIME SERIES FORECASTING

A time series is a chronological sequence of observations on a variable. Usually the observations are taken at regular intervals (days, months, years), but the sampling could be irregular. Common examples of time series are the Dow Jones Industrial Average, Gross Domestic Product, unemployment rate, and airline passenger loads. A time series analysis consists of two steps: (1) building a model that represents a time series, and (2) using the model to predict and forecast future values [3,4].

If a time series has a regular pattern, then a value of the series should be a function of previous values. If Y is the target value that we are trying to model and predict, and Yt is the value of Y at time t, then the goal is to create a model of the form:

$$Yt = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_{t-n}) + e_t \quad (1)$$

Where Yt_1 is the value of Y for the previous observation, Yt-2 is the value two observations ago, etc., and et represents noise that does not follow a predictable pattern (this is called a random shock). Values of variables occurring prior to the current observation are called lag values. If a time series follows a repeating pattern, then the value of Yt is usually highly correlated with Yt_cyck where cycle is the number of observations in the regular cycle. For example, monthly observations with an annual cycle often can be modeled by

$$Yt = f(Y_{t-12}) \quad (2)$$

The goal of building a time series model is the same as the goal for other types of predictive models which is to create a model such that the error between the predicted value of the target variable and the actual value is as small as possible [4,5]. The primary difference between time series models and other types of models is that lag values of the target variable are used as predictor variables, whereas traditional models use other variables as predictors, and the concept of a lag value doesn't apply because the observations don't represent a chronological sequence.

III. PREDICT WE ALGORITHMS

There are several methods of making forecasts, but they all fall into two categories: causal methods and time series methods.

In this paper, we are interested by making forecasts using the time series methods.

For this purpose, we find several predictive algorithms that can be implemented to make predictions and forecasts.

These algorithms include: Linear Regression, Support Vector Machines and Multilayer Perceptron.

A. Linear Regression

A regression model is a compact mathematical representation of the relationship between the response variable and the input parameters in a given design space [6]. Linear regression models are widely used to obtain estimates of parameter significance as well as predictions of the response variable at arbitrary points in the design space. One of the simpler forms of such models is

$$y = \beta_0 + \sum_{i=1}^{m} \beta_i x_i + \varepsilon \quad (3)$$

where y is the dependent or response variable, {x_i l< i < m} are the independent or regressor variables and is the residual -the error due to lack of fit. B0 is interpreted as the intercept of the response surface with the y-axis and {A l< i <m} are known as the partial regression coefficients. The coefficient values represent the expected change in the response y per unit change in x_i and indicate the relative significance of the corresponding terms. It is often the case that the regressor variables interact i.e. the effect of a change in x_i on y depends on the value of x_r. In such cases, the simple model in Eq. 1 is not sufficient. It is necessary to introduce terms that explicitly model two-factor interactions as shown below.

$$y = \beta_0 + \sum_{i=1}^{m} \beta_i x_i + \sum_{i=1}^{m} \sum_{j=i+1}^{m} \beta_{i,j} x_i x_j + \varepsilon \quad (4)$$

Eq. 5 represents a complete model that includes three-factor, four-factor and all higher order interactions. There are 2m terms in this model and an equal number of unknown regression coefficients.

$$y = \beta_0 + \sum_{i=1}^{m} \beta_i x_i + \sum_{i=1}^{m} \sum_{j=i+1}^{m} \beta_{i,j} x_i x_j + \sum_{i=1}^{m} \sum_{j=i+1}^{m} \sum_{k=j+1}^{m} \beta_{i,j,k} x_i x_j x_k + \cdots + \beta_{1,2,\ldots,m} x_1 x_2 x_m \quad (5)$$

The linear regression models we develop in this session can be represented as a sum of k terms from this complete linear model, expressed in a generic form as

$$y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \ldots + \beta_{k-1} x_{ik-1} + \varepsilon \quad (6)$$

where each $x_i 1$ is a distinct term from the complete model, and can be single factor, two factor, three factor or of any higher order. The collection of terms chosen for a given linear model will be referred to as the model terms.

In matrix terms, Eq. 6 can be written as

$$\boldsymbol{y} = \boldsymbol{X\beta} + \varepsilon \quad (8)$$

where 0 is the vector of regression coefficients and X is the model matrix. The model matrix has columns corresponding to the regressor variables x_l, x_2, ... , x_m, columns for interaction terms of any order, and a column of one's defining the intercept.

B. Multilevel Regression

This is quite similar to the simple linear regression model we have discussed previously, but with multiple independent variables contributing to the dependent variable and hence multiple coefficients to determine and complex computation due to the added variables. Jumping straight into the equation of multivariate linear regression [7].

$$Y_i = \alpha + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \ldots + \beta_n x_i^{(n)}$$

Y_i is the estimate of i^{th} component of dependent variable y, where we have n independent variables and x_i^j denotes i^{th} the component of the j^{th} independent variable/feature. Similarly cost function is as follows:

$$E(\alpha, \beta_1, \beta_2, \ldots, \beta_n) = \frac{1}{2m} \sum_{i=1}^{m} (y_i - Y_i)$$

where we have m data points in training data and y is the observed data of dependent variable.

As per the formulation of the equation or the cost function, it is pretty straight forward generalization of simple linear regression. But computing the parameters is the matter of interest here.

Normal Equation

Now let us talk in terms of matrices as it is easier that way. As discussed before, if we have n independent variables in our training data, our matrix X has n+1 rows, where the first row is the 0th term added to each vector of independent variables which has a value of 1 (this is the coefficient of the constant term a). So,X is as follows

$$X = \begin{bmatrix} X_1 \\ .. \\ X_m \end{bmatrix}$$

Y is the outcome, xi are the attribute values are

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ .. \\ Y_m \end{bmatrix}$$

Coefficient matrix C.

$$C = \begin{bmatrix} \alpha \\ \beta_1 \\ .. \\ \beta_n \end{bmatrix}$$

And our final equation for our hypothesis is

$$\boldsymbol{Y=XC}$$

To calculate the coefficients, we need ***n+1*** equations and we get them from the minimizing condition of the error function. Equating partial derivative of

$E(\alpha, \beta_1, \beta_2, \ldots, \beta_n)$

with each of the coefficients to 0 gives a system of equations. Solving these is a complicated step and gives the following nice result for matrix C.

$$C = (X^T X)^{-1} X^T y$$

where y is the matrix of the observed values of dependent variable.

This method seems to work well when the n value is considerably small (approximately for 3-digit values of n). As n grows big the above computation of matrix inverse and multiplication take large amount of time. In future tutorials lets discuss a different method that can be used for data with large no.of features.

C. Moving Average

In the statistical analysis of time series, autoregressive–moving-average (ARMA) models provide a parsimonious description of a (weakly) stationary stochastic process in terms of two polynomials, one for the autoregression (AR) and the second for the moving average (MA).

The notation AR(p) refers to the autoregressive model of order p. The AR(p) model is written

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t.$$

where $\varphi_1, \ldots, \varphi_p$ are parameters, c is a constant, and the random variable is ε_t white noise.

Some constraints are necessary on the values of the parameters so that the model remains stationary. For example, processes in the AR(1) model with $|\varphi 1| \geq 1$ are not stationary. (11)

The notation MA(q) refers to the moving average model of order q:

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

where the θ1, ..., θq are the parameters of the model, μ is the expectation of X_{t} (often assumed to equal 0), and the are again, $\varepsilon_t, \varepsilon_{t-1}, \ldots$ white noise error terms.

The notation ARMA(p, q) refers to the model with p autoregressive terms and q moving-average terms. This model contains the AR(p) and MA(q) models

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

The general ARMA model was described in the 1951 thesis of Peter Whittle, who used mathematical analysis (Laurent series and Fourier analysis) and statistical inference. [1][2] ARMA models were popularized by a 1970 book by George E. P. Box and Jenkins, who expounded an iterative (Box–Jenkins) method for choosing and estimating them. This method was useful for low-order polynomials (of degree three or less).

In some texts the models will be specified in terms of the lag operator L. In these terms then the AR(p) model is given by

$$\varepsilon_t = \left(1 - \sum_{i=1}^{p} \varphi_i L^i\right) X_t = \varphi(L) X_t$$

where φ represents the polynomial

$$\varphi(L) = 1 - \sum_{i=1}^{p} \varphi_i L^i.$$

The MA(q) model is given by

$$X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t = \theta(L) \varepsilon_t,$$

where θ represents the polynomial

$$\theta(L) = 1 + \sum_{i=1}^{q} \theta_i L^i.$$

Finally, the combined ARMA(p, q) model is given by

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t,$$

or more concisely,

$$\varphi(L) X_t = \theta(L) \varepsilon_t$$

IV. EXPERIMENTAL RESULTS

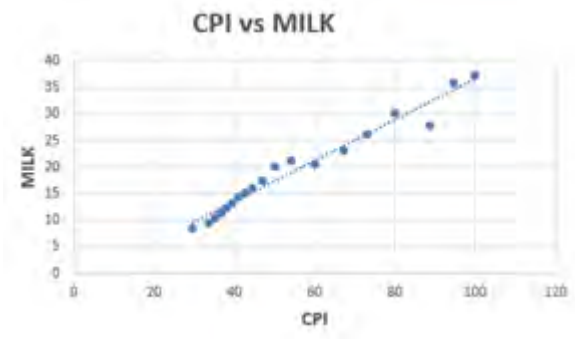
A. Algorithm

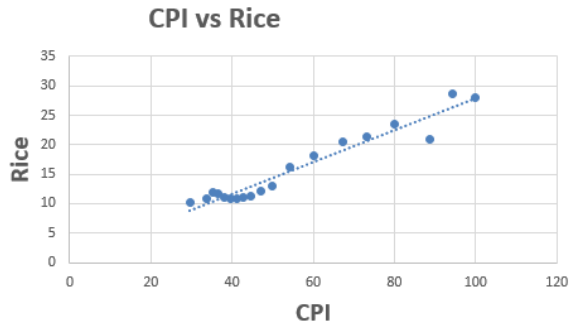
PHP-ML and PHP-AI were awarded for best quality. Machine Learning is a very complex and vast topic to implement in computer science. Machine Learning and Artificial Intelligence topics understanding are some of the most challenging problems to solve in computer science, requiring sophisticated deep learning algorithms to be trained on massive amounts of data and infrastructure. PHP-ML puts deep learning technologies within reach of all developers.

B. Datasets

Our experimentation is based on three time series algorithms: Food commodifies data represents the retail cost of food commodities from 2005 till 2015 for its actual value. Various other parameters and variables are also taken in consideration like prices of Petrol, Dollar, Value of CPI, temperature values from 2005 to 2015.

Almost all of the time series here selected are real world time series data. We have considered that application-oriented engineers in process control, manufacturing, industrial production and research centers would find more interesting this paper if data used here was from real world. Real world data can provide more credible results on forecasts of time series. They can also suffer by external issues of an occasional nature phenomenon such as storms, earthquakes, floods, strikes, war, technological advances, etc..., which make them an interesting dataset. Our time series are shown from Figures 2 to 5.





These time series present different seasonal and trended characteristics. As some time series have been measured monthly, they have a period of 12 (K = 12). On the other hand, some of them also show trend.

C. Results and discussion

Three different predictive algorithms were implemented to achieve our comparative study for time series forecasting. The available time series were divided into two subsets. The time series from January 2005 to December 2015 was employed as training set used for training the models of the algorithms.

The time series from January 2005 to October 2015 was used as the test set.

Our results are shown from Figures 5 to 9 and tables 1 to 4. Each graph presents the forecasting values of the three predictive algorithms, as well as the observed values obtained from the test data.

TABLE I. Milk Price Forecast

	Observed	MLR	MVA	LR
2006	17.58	17.4042	15.822	16.701
2007	19.9525	19.752975	17.95725	18.954875
2008	21.1275	20.916225	19.01475	20.071125
2009	20.4975	20.292525	18.44775	19.472625
2010	23.08	22.8492	20.772	21.926
2011	26.135	25.87365	23.5215	24.82825
2012	30.2175	29.915325	27.19575	28.706625
2013	27.7336	27.456264	24.96024	26.34692
2014	35.9525	35.592975	32.35725	34.154875
2015	37.165	36.79335	33.4485	35.30675

TABLE II. RICE PRICE FORECASTS

	Observed	MLR	MVA	LR
2006	12.0325	11.912175	11.0699	11.430875
2007	13.0375	12.907125	11.9945	12.385625
2008	16.26	16.0974	14.9592	15.447
2009	18.11	17.9289	16.6612	17.2045
2010	20.55	20.3445	18.906	19.5225
2011	21.4375	21.223125	19.7225	20.365625
2012	23.49	23.2551	21.6108	22.3155
2013	20.8303	20.621997	19.163876	19.788785
2014	28.64	28.3536	26.3488	27.208
2015	28.08	27.7992	25.8336	26.676

Table III. Dal Price forecasts

	Observed	MLR	MVA	LR
2006	32.0475	31.727025	29.4837	30.445125
2007	37.18	36.8082	34.2056	35.321
2008	42.755	42.32745	39.3346	40.61725
2009	66.465	65.80035	61.1478	63.14175
2010	67.3925	66.718575	62.0011	64.022875
2011	66.175	65.51325	60.881	62.86625
2012	71.7075	70.990425	65.9709	68.122125
2013	66.4396	65.775204	61.124432	63.11762
2014	76.6675	75.900825	70.5341	72.834125
2015	85.0875	84.236625	78.2805	80.833125

Table IV. Oil Price forecasts

	Observed	MLR	MVA	LR
2006	57.5125	56.937375	52.9115	54.636875
2007	62.9282	62.298918	57.893944	59.78179
2008	77.92	77.1408	71.6864	74.024
2009	67.8925	67.213575	62.4611	64.497875
2010	67.18	66.5082	61.8056	63.821
2011	87.88	87.0012	80.8496	83.486
2012	95.005	94.05495	87.4046	90.25475
2013	95.4228	94.468572	87.788976	90.65166
2014	90.0725	89.171775	82.8667	85.568875
2015	87.78	86.9022	80.7576	83.391

The graphs and tables indicated that the predicted values by all the selected algorithms reasonably matched the observed data. However, a closer look indicates clearly that the best forecasts have been produced by Multiple linear regression for all of the three-time series, following the shape of the observed values.

V. CONCLUSION

In this research work, we have made a comparative study of predictive algorithms to provide a time series forecasting. Three predictive algorithms have been conducted which are linear regression, multiple linear regression and moving average to evaluate their effectiveness of making forecasts. To assess the performance of the overall three algorithms, several experiments have been conducted using three time series which are Milk, Rice, Dal and Oil in a algorithm. According to our comparative study we can clearly conclude that the Multi linear regression is the best prediction algorithm that can be execute in Weka to provide the time series forecasting.

REFERENCES

[1] S. Makridakis, S. Wheelwright, and R. Hyndman, Forecasting Methods and Applications, 3rd Edition, John Wiley & Sons, USA, 2008.

[2] C. Chatfield, Time Series Forecasting, Chapman and Hall/CRC, USA, 2000

[3] G.E.P. Box, G.M. Enkins, G.C.Reinsel, Time Series Analysis: Forecasting and Control, 4th Edition, John Wiley, USA, 2008

[4] R.H. Shumway, D.S.Stoffer, Time Series Analysis and Its Applications

[5] G.P. Zhang, time series forecasting using a hybrid ARIMA and neural network model

[6] B.Heshmaty and A.kandel, Fuzzy Linear Regression and Its Applications to Forecasting In Uncertain Environment

[7] L. J. Cao and Francis E. H. Tay, Support Vector Machine With Adaptive Parameters in Financial Time Series Forecasting, IEEE Transactions on Neural Networks, Vol. 14, No. 6, November 2003

[8] S. Lahmiri, A Comparative Study of Backpropagation Algorithms in Financial Prediction, Internationam Journal of Computer Science, Engineering and Applications Vol.1, No.4, August 2011

[9] A.V. Devadoss, T.A. Alphonnse Ligori, Forecasting of Stock Prices Using Multilayer Perceptron, International Journal of Computing Algorithm, Volume: 02, December 2013, Pages: 440-449

[10] M. Tiwari, M. Bhai Jai and 0. Yadav, Performance analysis of Data Mining algorithms in Weka, IOSR Journal of Computer Engineering Volume 6, Issue 3 , PP 32-41,2012

[11] M. Forster, R. Passmore, Time series data sets 2013.

Solving Communication Networking Problems with Machine Learning Approach: A Case Study

Nidhi Kamothi, Himani Tiwari, Megharani Patil

Thakur College of Engineering & Technology, University Of Mumbai, Mumbai, India
nidhikamothi95@gmail.com, hemanipurohit23@gmail.com, megharanipatil@thakureducation.org

***Abstract*— The modern world relies on the availability of huge communication networks connecting computers and other electronic devices worldwide. There are an enormous number of problems that must be solved in order to maintain a good flow of data or information exchange. A Computer Network is the cornerstone of every organization used to share or exchange information which can be an image, text, video, sound clip or any other type of media or file. First problem is Broadcasting Information. One of the basic information tasks is to spread information from one node to all members of the communication network. Second Problem is Bandwidth Allocation. Bandwidth allocation in the Internet is managed by the Transport Control Protocol (TCP). If the acknowledgment does not appear, then a sending failure occurs possibly due to exceeded bandwidth. Third problem is Duplicate IP Address. If more than one PC is trying to use same IP address, it can cause network issue with the parent network of computer or also it can intermittent network communications. Fourth problem is improper Internet or Network Speed. It causes due to improper planning of network which causes the slower internet speed or data transmission over the network. Another problem is No network access issue. Sometimes we see "No Network Access" icon in our working computers or either we see a HTTP 504 web server error code. Network traffic classification is important for QoS, Network management and security monitoring. Intrusion is also one of the most serious problems with Network Security, as new types of intrusions are getting much more challenging to detect. Large amount of network traffic has been generated due to the use of internet. ML has been applied to various mundane and complex problems arising in network operation and management. It helps the network administrator to take quick decision by classifying the incoming traffic as either malicious or normal and hence efficient detection of intrusion. Therefore, this is a timely contribution of the implications of ML for networking that is pushing the barriers of autonomic network operation and management.**

Keywords— ***Machine learning, Network traffic, Network Security, Intrusion Detection, Transmission Control Protocol, Routing, Wireless Sensor Networks, Ad-hoc networks***

I. INTRODUCTION

Machine learning techniques have been applied in many areas of science due to their unique properties like adaptability, scalability, and potential to rapidly adjust to new and unknown challenges. Network security is a fast-growing field demanding a great deal of attention because of remarkable progresses in social networks, cloud and web technologies, online banking, mobile environment, smart grid, etc. Diverse machine learning methods have been successfully deployed to address such wide-ranging problems in network security. Network security is defined as a method to secure the network and any attempt to violate the CIA (Confidentiality, Integrity and Availability) of security is defined as Intrusion. The system used to detect the intrusion is called Intrusion Detection System (IDS). Intrusion detection has become a major problem in today's world. Broadcasting in processor networks means disseminating a single piece of information, which is originally known only at some nodes, to all members of the network. This is done in a sequence of rounds by pairwise message exchange over the communication lines of the network. Communication networks are often described as undirected graphs. The participants are nodes and information transmission is modelled by communication links between two nodes, abstracted by undirected edges. One of the basic information tasks is to spread information from one node to all members of the communication network. Bandwidth allocation in the Internet is managed by the *Transport Control Protocol (TCP)*. . The Transmission Control Protocol (TCP) used in the Internet is an end-to-end congestion control mechanism. One of its main features is the so-called Additively Increasing and Multiplicatively Decreasing algo. Duplicate IP Address can cause network issue with the parent network of computer or also it can intermittent network communications. It can be solved by assigning a unique IP address for your PC or every computer system. Slow Internet or network speed can happen due to improper planning of network which causes the slower internet speed or data transmission over the network. Intrusion detection has become a major problem in today's world with the increase in unlabelled traffic. Intruders are smarter than the detector as they always find a way to breach the recent IDS. IDS can be developed using unsupervised, semi supervised or supervised learning. Machine learning technique can be used for both misuse detection and anomaly detection

II. BACKGROUND

Discovery of malicious correlations in computer networks has been an emergent problem motivating extensive research in computer science to develop improved intrusion detecting systems (IDS). In this manuscript, we present a machine learning approach known as Decision Tree (C4.5) Algorithm and Genetic Algorithm, to classify such risky/attack type of connections. The algorithm obtains into consideration dissimilar features in network connections and to create a classification rule set. Every rule in rule set recognizes a particular attack type. For this research, we implement a GA, C.45 and educated it on the KDD Cup 99 data set to create a rule set that can be functional to the

IDS to recognize and categorize dissimilar varieties of assault links [1]

Network traffic classification is important for QoS, Network management and security monitoring. Current method for traffic classification such as port based or payload based suffered many problems. Newly emerged application uses encryption and dynamic port numbers to avoid detection. So we use unsupervised machine learning approach to classify the network traffic. In this paper unsupervised Kmeans and Expectation Maximization algorithm are used to cluster the network traffic application based on similarity between them. Performance of these two algorithms is compared in terms of classification accuracy between them. The experiment results show that K-Means and EM perform well but accuracy of K-Means is better than EM and it form better cluster [2]

Intrusion is one of the most serious problems with network Security, as new types of intrusions are getting much more challenging to detect. Large amount of network traffic has been generated due to the use of internet; most of the generated traffic is in the format which cannot be used directly to arrive at meaningful information. The cleansing and labeling of data each time needs a considerable amount of human effort, and is time consuming. In this paper we show how, Semi supervised machine learning technique can be used in intrusion detection, for both labeled and unlabeled data. In the proposed technique we take a small amount of labeled data to create model and using this model we show how to predict the unlabeled traffic. Machine Learning tool is used for this purpose which uses semi supervised classifier to build the model [3]

With the internet of objects, the number of devices with internet connection is increasing day by day. This leads to a very high amount of data circulating on the internet. It is one of the most common problems that can be distinguished from normal and abnormal traffic by analyzing in high data amount. In this study, an analysis was carried out by using machine learning approaches to determine whether the data received on the internet is normal or abnormal data. In order to achieve this goal, the KDD Cup 99 data set which is frequently used in literature studies is classified by Naive Bayes (NB), bayes NET (bN), Random Forest (RF), Multilayer Perception (MLP) and Sequential Minimal Optimization (SMO) algorithms. Classifiers are also compared with false rate, precision, recall, and F measure metrics along with accuracy rate values. Classification times of classifiers are also given by comparison [4]

Organization has come to realize that network security technology has become very important in protecting its information. With tremendous growth of internet, attack cases are increasing each day along with the modern attack method. One of the solutions to this problem is by using Intrusion Detection System (IDS). Machine Learning is one of the methods used in the IDS. In recent years, Machine Learning Intrusion Detection system has been giving high accuracy and good detection on novel attacks. In this paper the performance of a Machine Learning algorithm called Decision Tree (J48) is evaluated and compared with two other Machine Learning algorithms namely Neural Network and Support Vector Machines which has been conducted by A. Osareh for detecting intrusion. The algorithms were tested based on accuracy, detection rate, false alarm rate and accuracy of four categories of attacks. From the experiments conducted, it was found that the Decision tree (J48) algorithm outperformed the other two algorithms [5]

In the paper, we demonstrate novel approach for network Intrusion Detection System (IDS)for cyber security using unsupervised Deep Learning (DL) techniques. Very often, the supervised learning and rules based approach like SNORT fetch problem to identify new type of attacks. In this implementation, the input samples are numerical encoded and applied un-supervised deep learning techniques called Auto Encoder (AE) and Restricted Boltzmann Machine (RBM) for feature extraction and dimensionality reduction. Then iterative k-means clustering is applied for clustering on lower dimension space with only 3 features. In addition, Unsupervised Extreme Learning Machine (UELM) is used for network intrusion detection in this implementation. We have experimented on KDD-99 dataset, the experimental results show around 91.86% and 92.12% detection accuracy using unsupervised deep learning technique AE and RBM with K-means respectively. The experimental results also demonstrate, the proposed approach shows around 4.4% and 2.95% improvement of detection accuracy using RBM with Kmeans against only K-mean clustering and Unsupervised Extreme Learning Machine (USELM) respectively [6]

III. SOLVING ROUTING PROBLEMS USING MACHINE LEARNING

We use protocols and application in the networking domain that designed to work in large scale heterogeneous, hierarchically organized network to solve various routing problems in its basic form using Machine learning. Addresses are assigned to nodes and the nodes keep static routing tables in order to be able to find the path to some other node. If there is any change in routing information or any failure of any nodes then it must ne propagated through whole network. Latency and loss of data increase, until the new information is successfully transmitted. In a wireless ad-hoc scenario, new problems arise and traditional routing protocols cannot be successfully applied. Transmitting of data becomes much more expensive, and the quality of service suffers. Additionally, in energy-restricted environments like WSNs the overhead of keeping routing information fresh becomes unbearable. So the problem arisen many researcher turned their attention to the domain of Machine Learning (ML). Different properties of usual WSN applications have to be considered: memory and computational limitations, communication costs, restricted energy. A survey for existing works is willing to apply machine learning to problems in ad hoc networking. First, Section I gives short descriptions of suitable machine learning approaches for use in wireless ad-hoc networks, identifying their most relevant requirements and advantages and suggesting their applicability domain. Section II classifies and evaluates the most important existing and on-going research in the area. The presented works are evaluated in terms of their quality – meeting the application requirements.
The various routing problems and networking topologies solution using machine learning:

Finding the Right Algorithm

There are several properties of wireless ad-hoc networks, which have to be considered when selecting a good ML approach.

Wireless ad-hoc nature: The topology of the network is constantly changing either through mobility or failure of the nodes. The shared wireless medium sets additional restrictions on the possible communication between the nodes and new problems arise, like link quality and asymmetric links in WSNs.

Mobility and topology changes: MANETs are usually highly mobile scenarios, where new nodes constantly join the

network, and existing ones either move through the network or leave it. . On the other hand, energy is not restricted, since individual nodes are usually laptops or PDAs, which can be easily recharged.

Physical distribution: Data is also distributed throughout the nodes of the network and can be gathered on a central station only with high communication costs. Consequently, algorithms requiring global information from the whole network become very expensive. Thus, distributability of the algorithms is highly recommended.

In distributed ML algorithms, the solution is found in a collaborative manner and each node holds only the part of it, directly relevant to its actions (like the best next hop in a routing problem). The properties of distributed ML techniques, suitable for problem solving in ad-hoc networks are given in Table I. It has to be noted that the presented properties are based on both the ML technique itself and its requirements and on already implemented protocols and applications. The most used Machine Learning techniques are presented, together with further literature and their most relevant requirements/properties for the ad-hoc networking domain [7]

Reinforcement Learning – RL: Generally, there are three types of Machine Learning - supervised, unsupervised and reinforcement learning. Reinforcement learning is biologically inspired and differs from supervised learning in how it acquires its experience and knowledge. Consider as a simple example a child, learning to walk. The possible actions it can take are different movements of its legs, arms and body and by selecting them, it interacts with its environment. Reinforcement learning is well suited for distributed problems, like routing. It has medium requirements for memory and computation at the individual nodes, arising from the need of keeping many different possible actions and their values[3]. The approach also needs some time for learning, until the value function of the actions stabilizes (the algorithm converges). It is easy to implement, is highly flexible to topology changes and achieves optimal results [8]

Swarm Intelligence – SI: The term Swarm Intelligence refers to a class of machine learning techniques, biologically inspired by the behavior of social insects like ants, bees etc. Swarm intelligence is well suited for distributed network scenarios, where mobility and topology changes are of greatest importance, but energy is not restricted, like MANETs. Like Reinforcement Learning, Swarm Intelligence techniques need some reasonable amount of memory and computational resources on the network nodes and is very adaptable to topology changes.

Mobile agents: Mobile agents are often mistaken for a Machine learning or Swarm Intelligence approach. However, they refer to the usage of simple, small entities (packets), which traverse the system (in our case the network) and deliver fresh information to the system's nodes. In the case of routing for example, the agents update routing information (paths or next hops) on the nodes. Although very efficient in some applications (like routing in less mobile scenarios), they cannot be classified as a learning nor as a swarm intelligence algorithm.

Real Time Heuristic Search: heuristic search methods operate in two steps: planning and plan execution.This approach cannot be applied in real time scenarios, where agents traverse the search space and have to take their decisions based on locally available data only. Real time heuristic search methods, called also agent-centered search.

ML BASED PROTOCOLS AND APPLICATIONS

ML based protocols for routing in wireless ad-hoc networks (both MANETs and WSNs) will be presented**.** In order to simplify comparison, names are introduced for the protocols where necessary**:**

Routing in MANETs: The mostly cited routing protocols, using ant system optimization is AntHocNet. It is based on the routing paradigm for communication networks. AntHocNet was designed especially for the needs of a wireless ad-hoc network[4]. The protocol uses both reactive path setup and proactive path maintenance for single source . The routing algorithm presented in called ARS, is also based on AntNet and basically extends it to support quality of service guarantees in communication networks, like bandwidth and hop-count.

Other applications in wireless ad-hoc networks:

Topology control: The approach for finding an optimal topology in a MANET is described in the authors use ants to figure out the possible power levels for all nodes in a MANET in a distributed manner and to adapt them continuously also in highly mobile environment.

Localization: Machine-learning based approach to solve the localization problem in WSN. Given n + m nodes in the network, where the positions of n nodes are known and of m nodes unknown, and given the RSSI signal strength between any pair of nodes, the positions of the un-localized nodes have to be recovered.

Intrusion detection: ants can be used for intrusion detection in a sensor network. Ant colonies with extended properties (called emotions) traverse the network and are able to detect systematic scanning of the network. Upon detection, they change their emotion and thus alert the network administrator. The approach is fully distributed, but comes with the additional costs of the traversing ants.

IX. CONCLUSION AND FUTURE WORK

Machine learning has been applied to routing problems in wireless networks as early as 1994 and since then finds more and more applicability. Special distributed algorithms has been developed or adjusted to the wireless ad hoc scenario and a large spectrum of traditional ML approaches has been applied to different problems, beginning with reinforcement learning and swarm intelligence. In this work, a survey of the most important ML-based solutions to the routing problem in wireless ad hoc networks (WSNs and MANETs) is presented. Additionally, machine learning techniques and their applicability to the described domain is given. Future work will extend the list of presented ML techniques and networking applications. Additionally, some important work from the broader area of networking will be included. We also discussed two different unsupervised machine learning algorithm K-means and Expectation maximization for different types of network traffic application.

REFERENCES

[1] S.Selvakani k, Rengan S Rajesh, “Integrated Intrusion Detection System Using Soft Computing”, IJNS, Vol.10, No.2, pp.87-92, March 2010

[2] Nguyen T.T.T., Armitage G. (2008) A Survey of Techniques for Internet Traffic Classification using Machine Learning. In: IEEE Communications Surveys &Tutorials, Vol. 10, No. 4, p. 56-76

[3] Rohit Arora M.Tech. CSE Deptt. Hindu College of Engineering Sonepat, Haryana, India Suman Asstt. Prof. CSE Deptt. Hindu College of Engineering Sonepat, Haryana, India Comparative Analysis of Classification Algorithms on Different Datasets using WEKA International Journal of Computer Applications (0975 – 8887) Volume 54– No.13, September 2012

[4] H.-J. Liao, C.-H. Richard Lin, Y.-C. Lin, and K.-Y. Tung, “Intrusion detection system: A comprehensive review,” J. Netw. Comput. Appl., vol. 36, no. 1, pp. 16–24, 2013.R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[5] H. M. Shirazi, "Anomaly Intrusion Detection System Using Information Theory, K-NN and KMC Algorithms", Australian Journal of basic and Applied Sciences vol. 3(3), pp. 258 1- 2597, INSInet Publication, 2009

[6] Md Zahangir Alom, VenkataRamesh Bontupalli, and Tarek M. Taha. ”Intrusion detection using Deep Belief Networks.” Aerospace and Electronics Conference (NAECON), 2015 National. IEEE, 2015

[7] P. Beyens, M. Peeters, K. Steenhaut, and A. Nowe, “Routing with Compression in WSNs: A Q-Learning approach,” in Proc. of the 5th Eur. Wksp on Adaptive Agents and Multi-Agent Systems (AAMAS), 2005

[8] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction. The MIT Press, March 1998

Determining Level of Cognition Of Mild Cognitive Impaired Person Using Screening Test and EEG Signals

Amruta Koli, Prof. Shredevi Karande
MIT World Peace University, Pune, India
amruta7.koli@gmail.com, shridevi.karande@mitpune.edu.in

***Abstract*—Dementia describes the group of symptoms associated with decline in memory. Mild Cognitive Impairment(MCI) is the stage between expected cognitive decline of normal aging and more serious decline of dementia. As there is large increase in dementia disease day by day it become necessary to find techniques for diagnose such disease. In this paper techniques such as Screening test which includes Mini Mental State Examination(MMSE) and Clock Drawing Test(CDT), Electroencephalogram(EEG) for detecting dementia, its various types and stage applied individually and in combinations are reviewed. Detection Of MCI stage includes machine learning algorithm such as Support Vector Machine(SVM), Logistic Regression(LR), Neural Network(NN), Artificial Neural Network(ANN) Further, the limitations are studied and analysed to get better accuracy for detection of MCI.**

Index Terms*—Mini Mental State Examination, Clock Drawing Test, Screening, Dementia, Mild Cognitive Impairment, Electroencephalogram

I. INTRODUCTION

According to survey of World Health Organization(WHO), Dementia is recognized as global health challenge. Globally, India houses the second most number of individuals suffering from dementia with an estimated 4.1 million people suffering from it as per the 'Dementia India' report published by the Alzheimer's and Related disorders Society Of India. This is expected to double by 2035. According to the World Health Organisation, around 50 million people in the world have dementia, and there are nearly 10 million new cases every year. So it is necessary to detect this disease at early stage so that it helps in increasing life span of person.

Dementia is one kind of disease which is nothing but memory loss and mostly happens at the age of 60's. Person having dementia has to suffer from lots of problems. Person having such a kind of disease may not able to think properly like normal healthy person, may have fear of going outside individually, even they forgot to perform their daily routine tasks. There are so many types of dementia like Alzheimer's dementia, Lewy Body dementia, Frontotemporal dementia, Parkinson's dementia, Vascular dementia and Alzheimer's is the most common type of dementia among all. Mild Cognitive Impairment is one of the stage of dementia. There are many symptoms of dementia includes memory problem,

communication and language problem, have less ability in focus and attention, judgment problem, short term memory, problem in performing daily activities. There are series of tests carried out in hospital region for detection of dementia but, these several tests requires lots of amount of time and efforts. So there is need for such tools which ease in process of detection of dementia. One factor which is used for detection of such disease is EEG. Electroencephalography is one kind of method which helps in knowing the brain functionality which includes process of pre-processing, feature extraction, feature selection and finally classification which consists of algorithm such as Support Vector Machine(SVM), Logistic Regression(LR), Artificial Neural network(ANN) for detection of Mild Cognitive Impairment. Screening test includes two tools. One is Mini Mental State Examination(MMSE) and other is Clock Drawing Test(CDT). The Mini Mental State Examination (MMSE) is one of the most popular tool which is used to detect individual with Mild Cognitive Impairment(MCI). It actually consists of 11-questions which covers 5 different areas for checking cognitive function of participants: orientation, registration, attention and calculation, recall, and language. The clock-drawing test(CDT) is a one kind of screening tool which is used for checking cognitive function of individual. It simply includes drawing of clock by time given by physician and score is given based on how clock is drawn. It is often used in combination with some other tools, but individually it enough to give result in detecting cognitive ability of person.

MCI is one of the stage of dementia. Person having MCI may have higher risk of 31%-44% of converting into dementia as compared to healthy person. Screening test has lots of benefits. It helps in identifying patient with dementia at early stages so that it beneficial in further assessment and may help in improving person's cognitive function, behaviour. Identifying dementia at early stage also helps in better assessment of patients for their better health and also helps in increasing life span of patients.

II. LITERATURE OVERVIEW

This section describes the previous work that has be done for detection of dementia, its type and stage Mild Cognitive Impairment.

This paper[1] describes about how Screening Test such as Mini Mental State Examination(MMSE) is used for identifying the cognitive functionality of patients. Mini Mental State Examination(MMSE) is simplified scoring method for finding the cognitive mental state of patients which simply includes 11 questioner and takes only 5-10 minutes to complete the test. This paper include data of 206 patients of having dementia and 63 healthy control persons for applying the MMSE method. As a result it was found that MMSE is one of the best method for determining the cognitive state of patients. This method was easy as compare to other methods, takes only fewer amount of time, easy for taking tests for patients and simple in identifying the cognitive mental state of patients.

This Paper[2] explained how Mini Mental State Examination(MMSE) helps in identifying cognitive functions of patients who are highly educated. Overall 1141 data of patients were recorded which includes patients with dementia, patients having Mild Cognitive Impairment(MCI) and normal healthy

control persons. Different factors such as sensitivity, Specificity were calculated based on MMSE score got for detection of cognitive function. In detection of dementia considered MMSE score of 24 or below that got classification rate of 89%. Individuals with dementia and Mild cognitive Impairment having MMSE score of 27 or below that got classification rate of 90%. At the end as a result it was found that patients having MMSE score of 27 or below that are at greater risk of having dementia disease and they need to be diagnosed with further proper treatment.

In this paper[3] about Mini Mental State Examination was explained. MMSE is one of the screening tool used for detection of Mild Cognitive Impairment(MCI) and its subtypes. This tool basically used to know the cognitive function of patients. It includes questionaries' of different areas like orientation, registration, attention and calculation, recall and language. Based on different score obtained in each area total MMSE score of
each individual was calculated and classified in MCI and its different subtypes.

This paper[4] gives idea about how Clock Drawing Test is used for determining cognitive performance of Mild Cognitive Impaired person. In this literature review of regarding CDT test was explained. How different scoring systems are used previously for diagnosing MCI was given. Sensitivity and specificity for each CDT score was given. As a result it was found that CDT is not good tool for determining MCI.

In this paper[5] they explained how Clock Drawing Test(CDT) is used for determining cognitive function of participants at early stage. Mini Mental State Examination(MMSE) tool also used to diagnose cognitive function. CDT is performed individually and in combination with MMSE. Age, Sex, Education factors are considered. CDT divided into 2 groups i.e. CDT1 and CD2. MMSE and both CDT1, CD2 score was calculated considering age, sex, education parameter.

This paper[6] described the in detection of dementia in older people how sensitivity and specificity value of Clock Drawing Test(CDT) was plays important role in primary care. In this both screening test Mini Mental State Examination and Clock Drawing Test were applied on 41 elderly patients having dementia disease and 523 healthy individuals. In result it was found that in detection of dementia the sensitivity of CDT was found to be 76% where as in normal individual's it was 81%. MMSE got higher sensitivity and specificity range. The conclusion of this paper was CDT plays important role in detection of dementia in primary care. CDT was found to be good for detection of dementia with good sensitivity and specificity value but may not good at detecting Mild Cognitive Impairment.

This paper[7] briefly described about how Clock Drawing Test(CDT) is used for knowing the cognitive function of individual's. While performing CDT test different factors such as age, education, mood were consider as they can affect test result. So many different ways for scoring CDT are discussed. Depending upon drawing of clock and their respective scoring participants cognitive function determined so that he/she can categorize into having dementia or not. This method considered to be easy to implement, less threatening to patients, requires less amount of time, simple test for determining dementia but it fails in some cases such as determining early stage of Alzheimers Disease(AD).

In this paper[8] how Screening Test Clock Drawing Test(CDT) was used to detect individual with MCI. In this Mini Mental State Examination(MMSE) also used to detect dementia. The main aim here was to test whether combination of CDT and MMSE helps in distinguishing subjects with MCI or mild dementia or normal healthy controls. The conclusion shows that CDT good at distinguishing MCI patients from no impaired
individuals and MMSE good at distinguishing MCI patients from those having dementia.

In this paper[9] one of the screening test i.e. Clock Drawing Test(CDT) is discussed. CDT is used to determining individuals cognitive function. Seventy-five subjects were selected to perform this test. Different categorization of error occurred during performing CDT were discussed. In this some other screening tests also performed such as Mini Mental State Examination(MMSE), Short Mental Status Questionnaire(SMSQ), Geriatric Depression Rating Scale (GDR). Sensitivity and Specificity of each test was calculated and compared with CDT and from that error level of clock was determined.

In this paper[10] screening test named Clock Drawing Test(CDT) was used for detection of Mild Cognitive Impairment(MCI). In this scoring systems are used for CDT score. These are Cahn;s scoring, Rouleau scoring, Sunderland scoring method. So finding best among these protocols of CDT sensitivity and specificity was calculated. As a result it was found that for detecting MCI at early stage Cahn's protocol is the best one among all three protocol and give satisfactory result.

This paper[11] studied how the combination of screening test such as MMSE and CDT helps in improving the detection of Mild Cognitive Impairment and Mild Alzheimer's Disease. This combination performed to differentiate between individuals having Alzheimer Disease, Mild Cognitive Impairment(MCI) and normal patients. The total time required to complete these screening test was 8 minute 50 sec. In this ROC curve was used to differentiate between these individuals. The mean area under ROC curve was found to be higher when MMSE and CDT used in combination than when used individually. The combination of MMSE and CDT achieved higher sensitivity and specificity rate in detection of Mild Alzheimer's disease.

In this paper[12] comparison of two screening test MMSE and CDT was explained at clinical level for detection of dementia. CDT is simple tool in which individual was asked to draw a clock face with placing of numbers on it. Different methods for scoring clock drawing test was discussed. Classification of clock was done based on its drawing pattern. CDT is simple easy to implement and performed with minimal equipment. It has key advantage that it can be performed with individuals with poor education. CDT was found to be good at detecting Alzheimer's Disease but may not be sensitive at detecting Mild Cognitive Impairment. MMSE scoring method

was explained. MMSE was found to be good in evaluation of cognitive mental state of individuals. Cons of using MMSE is that to perform this test individuals need to be highly educated. So overall this paper explained the usage of CDT and MMSE in detection of dementia.

This paper[13] described about how combination of different screening tests helps in evaluation of Mild Cognitive Impairment(MCI) which mostly occurred in elder people. In this paper combination of Mini Mental State Examination(MMSE), Verbal Fluency Test(VFT) and Clock Drawing Test(CDT) were used for determining individuals with MCI and Alzheimer's disease(AD).Overall 247 of older individuals was examined. Receiver Operating Characteristic(ROC) curve was used for evaluation of performance. As a result it was found that Combination of these tests provide good accuracy in detection of AD but individual these test not so good at detecting patients with MCI from healthy subjects. Conclusion provided here says that combination of these tests does not warrant to give good result at detecting MCI patients. MMSE score single sufficient for detecting patients with AD.

In this paper[14] how EEG signals were used in detection of Mild Cognitive Impairment(MCI) was explained. In this EEG data was recorded from 17 patients to examine brain activity. Event Related brain Potentials(ERP) was recorded as response generated from EEG signals. In this spectral and temporal features of ERP were extracted. Screening test such as Montreal Cognitive Assessment(MOCA) was used for identifying cognitive function of patients. Process of MCI detection includes 1)feature extraction 2)feature scoring 3)classification. 4) evaluation. In classification different algorithms such as Support Vector Machine(SVM), Logistic Regression(LR) , Random Forest(RF) was used for performance evaluation. At the end as a result it was proved that time frequency features are best among all other features. Classification accuracy got here was 87.9%. In this single channel EEG signals was used, in future it can be possible to use multichannel signals.

In this paper[15] role of EEG was studied as it plays important role in detection and classification of dementia. In this EEG signal processing stages was explained. Description of different waves generated by EEG such as α-wave, β-wave, γ-wave, δ-wave were explained. How the placing of electrodes was performed for EEG signal generation was given. Independent Component Analysis(ICA), Wavelet Transform(WT) methods was used for EEG signal processing. EEG signal generation, feature extraction, feature selection, classification all these stages were performed. Support Vector Machine(SVM), Linear Discriminant Analysis(LDA) was used for classification. Result showed that SVM is best technique for classifying features. Conclusion found that EEG is great biomarker for detecting and classifying dementia individuals. Dataset used here was with fewer sample so it need to be increase for better accurate result.

In this paper[16] they have explained how EEG signals are used for detection of Alzheimer Disease(AD). They have explained how EEG signals are processed to get result. EEG methodology includes Pre-processing, Feature extraction, feature selection, classification and finally Diagnosis. Linear Discriminant analysis(LDA), Support Vector Machine(SVM) algorithms was used for classification. Different methods used for generation of EEG signals were studied and their corresponding accuracy were compared with each other. Output of classification helped to diagnose whether person is in early stage of dementia or he/she have AD. So finally it was conclude that EEG helps in detecting different stages of dementia.

This paper[17] described about usage of dominant frequency analysis of EEG signal in diagnosis of dementia. The main objective was to detect MCI individual from healthy control. For this purpose different frequencies such as Dominant frequency, dominant frequency variability and dominant frequency prevalence was tested and their corresponding performance were compared. As a result it was found that EEG can be a good helping tool for detecting dementia at early stage.

In this paper[18] EEG analysis regarding Mild Cognitive Impaired patients was explained. EEG signals help in detecting whether person is suffering from MCI stage or not. EEG complexity includes Lempel-Ziv complexity, LZC. P300 value was also determined. All these values are compared between two groups i.e. healthy control and MCI patients. As a conclusion it was found that in healthy control LZC factor was higher than person having MCI and also most of the brain area like frontal and temporal area had more LZC value.

This paper[19] discussed about one of the stage of dementia called Mild Cognitive Impairment(MCI). In this participants involved were healthy control and person having Alzheimers Disease(AD) and MCI patients and there corresponding performance were compared. The comparison done based on demographic factors and different tests like Mini-Mental State Examination(MMSE), Wechsler Adult Intelligence Scale-Revised, Wechsler Memory Scale-Revised, Dementia Rating Scale, Free and Cued selective Reminding Test, and Auditory Verbal Learning Test. As a conclusion it was found that those above parameters considered helped in clear distinguish of individual having AD, MCI and normal healthy controls.

This paper[20] described about how emotions of individuals generated from EEG signal using Emotive device was evaluated. In this EEG signal processing stages such as EEG signal generation, feature extraction, Classification and finally evaluation was discussed. For EEG signal generation Emotive epoc headset was used. Emotive epoc headset placed on head as per instruction given. After that feature extraction was performed so that any irrelevant activity done by individual during EEG signal generation can be removed. Finally classification algorithms such as Support vector Machine(SVM), Linear Discriminant Analysis was used on features selected for performance evaluation. Results showed that EEG signal generated from Emotive epoc device contains enough information to classify different emotions such as happiness, sad, anger, calm. In future different feature extraction methods can be used in order to improve accuracy.

In this paper[21] from EEG signals how emotions of human can be detected is discussed. For humans emotion brain plays important role and brain activity can be detected using EEG. Emotions can be happy, sad, anger, calm can be detected using EEG signals. Generation of EEG signals involves different

stages such as Pre-processing, feature extraction, feature selection, classification and finally evaluation. Classification involves different algorithms such as Support Vector Machine(SVM), Logistic Regression. EEG generate various kinds of waves such as α-wave, β-wave, γ-wave, δ-wave. Corresponding frequency values were also recorded for these waves which helps in understanding current running emotion in human brain.

TABLE 1: COMPARATIVE ANALYSIS

a. Comparative Study for Classifiers

Sr. No.	Classifiers	Advantages	Disadvantages
1	SVM	Use of kernel techniques makes it more powerful .	It takes more time for larger dataset.
2	LR	Very efficient as it does not require too many computational resources.	It can't solve non-linear problems.
3	DNN	Have ability to Model non linear and complex relationship.	Training of data is very expensive.
4	RF	Good performance on many problems including non linear.	Need to choose the number of trees.

III CONCLUSION

This paper discussed about how the different factors are used for detecting dementia, its various types and stage Mild Cognitive Impairment individually and/or in combination. Screening test Mini Mental State Examination(MMSE) and Clock Drawing Test(CDT) applied in combination gives better performance. The main goal is to develop a such system in combination with Screening Test and EEG so that it helps in detecting dementia at early stage. From the above study it is found that there are some flaws in it. Machine Learning algorithm which are used currently in classification phase of EEG signal generation for detection of dementia gives accuracy up to certain mark. By using different algorithms such as Artificial Neural Network, Deep Neural Network will be able to increase accuracy in future so that we get better result in detection of Mild Cognitive Impairment.

REFERENCES

[1] Marshal f. folstein, susan E. folstein and Paul R. Mchugh, "Mini-Mental State A Practical method for grading the cognitive state of patients for the clinician", J. Psychiat, 1975.

[2] Sid E O'Bryant, Joy D. Humphreys, MA; Glenn E. Smith, PhD;, "Detecting Dementia With the Mini-Mental State Examination in Highly Educated Individuals", Arch Neurol, 2008.

[3] Breno S.O. Diniz, Monica S. Yassuda, Paula V. nunes, Marcia Radanovic and Orestes V. Forlenza, "Mini-mental State Examination performance in mild cognitive impairment subtypes", International Psychogeriatric Association, 2007.

[4] Lena Ehreke, Melanie Luppa, Hans-Helmut Konig and Steffi G. Riedel-Heller, "Is the Clock Drawing Test a screening tool for diagnosis of mild cognitive impairment? A systematic review", International Psychogeriatric Association, 2009.

[5] C Mittal, Col SP goethi, Maj Gen S Rohatgi, "Early Cognitive Impairment: Role of Clock Drawing Test", MJAFI , Vol. 66, No. 1, 2010.

[6] Michael Kirby, Aisling Denihan, Irene Bruce, David Coakley and brian A. Lawlor, "The Clock Drawing Test in primary care: sensitivity in dementia detection and specificity against normal and depressed elderly", International Journal Of Geriatric Psychiatry, 2001.

[7] BERIT AGRELL, O V E DEHLJN, "The Clock-drawing Test", Geriatric Section, Department of Internal Medicine, Lund University Hospital, 22185 Lund, Sweden, 1998.

[8] S. Umidi, P.D. Trimarchi, M. Corsi, C. Luzzati, G. Annoni, "Clock Drawing Test In The Screening Of Mild Cognitive Impairment(MCI)", Elsevier, 2009.

[9] Kenneth I. Shulman, Md, Ralph Shedletsky, Phd A N D Ivan L. Silver, Md, "THE CHALLENGE OF TIME: CLOCK-DRAWING AND COGNITIVE FUNCTION IN THE ELDERLY", International Journal Of Geriatric Psychiatry, Vol. 1, 1986.

[10] Sayaka Yamamoto, Nanaka Mogi , Hiroyuki Umegaki ,Yusuke Suzuki, Fujiko Ando,Hiroshi Shimokata ,Akihisa Iguchi ,"The Clock Drawing Test as a Valid Screening Method for Mild Cognitive mpairment", Dementia and Geriatric Cognitive Disorder, 2004.

[11] Jesús Cacho, Julián Benito-León, Ricardo García-García , Bernardino Fernández-Calvo , José Luis Vicente-Villardón and Alex J. Mitchell , "Does the
Combination of the MMSE and Clock Drawing Test (Mini-Clock) Improve the Detection of Mild Alzheimer's Disease and Mild Cognitive Impairment?", Journal Of Alzheimer's Disease, 2010.

[12] Delnaz Palsetia, G. Prasad Rao, Sarvada C. Tiwari , Pragya Lodha , Avinash De Sousa,"The Clock Drawing Test versus Mini-mental Status Examination as a Screening Tool for Dementia: A Clinical Comparison", Indian Journal of Psychological Medicine, Vol. 40, Issue 1, 2018.

[13] Rodolfo B. Ladeira, Breno S. Diniz, Paula V. Nunes, Orestes V. Forlenza, "COMBINING COGNITIVE SCREENING TESTS FOR THE EVALUATION OF MILD COGNITIVE IMPAIRMENT IN THE ELDERLY",CLINICS, 2009.

[14] Saleha Khatun, Bashir I. Morshed, Gavin M. Bidlman, "Single Channel EEG Time-Freqency Features to Detect Mild Cognitive Impairment, IEEE, 2017.

[15] Noor Kamal Al-Qazzaz, Sawal Hamid Bin MD. Ali, Siti Anom Ahmad, Kalaivani Chellappan, Md. Shabiul Islam, and Javier Escudero, "Role of EEG as Biomarker in the Early Detection and Classification of Dementia", The Scientific World Journal, 2014.

[16] Nilesh Kulkarni, Prof. Dr. V. K. Bairagi, "Diagnosis of Alzheimer Disease using EEG Signals", International Journal of engineering Research & Technology(IJERT), Vol. 3Issue4, April-2014.

[17] G. Anuradha, Najumnissa Jamal and Syed Rafiammal, "Detection pf Dementia in EEG Signal using Dominant Freqency Analysis", IEEE Internationl Conference on Power, Control, Sigmals and Instrumentation Engineering, 2017.

[18] Benju Zhu, Changfeng Chai, Shane Gao, haiyan Ren, Limei Cao, Zhiqiang Dong, XiangGeng, Jia Zheng, Xuemei Qian, Xiaoying Bi and xu Chen, "Analysis of EEG Complexity in Patients with Mild Cognitive Impairment", Journal of Neurological Disorders, Vol. 5, Issue 4, 2017.

[19] Ronald C. Petersen, PhD, MD; Glenn E. Smith, PhD; Stephen C. Waring, DVM, PhD; Robert J. Ivnik, PhD; eric g. Tangalos, MD; Emre Kokmen, MD, " Mild Cognitive Impairment", ARCH NEUROL, Vol. 56, MAR 1999.

[20] Rafael Ramirez, Zacharias Vamvakousis, "Detecting Emotion from EEG Signals Using the Emotive Epoc Device", Springer-Verlag Berlin Heidelberg, 2012.

[21] Rahulraj Singh and M. Uma , "A Study Of Emotive Brain Mapping using EEG", International Science Press(IJCTA), 2016.

Evaluation Of Aurdino Based Soil Moisture Measurement Using IOT

Renuka Sahani,Manu Nair,Smriti Saxena, Ms Vaishali Nirgude
Computer Department, Thakur College of Engineering & Technology, Kandivali (East), Mumbai-400101, India.
renukasahani16@gmail.com,manm770@gmail.com,smririrenuka16@gmail.com

***Abstract*-An automated soil moisture measurement system will be developed to help the farmers in knowing the moisture of the soil. The system will have required sensors of our interest like temperature, moisture,Ph meter placed soil. Automation is done through IOT , IOT is a shared network that can interact with objects through internet connection. This system will be developed for monitoring and maintaining the situation at farm field with sensors. Etc. IOT is widely used in home automation, smart agricultrure. The objective is to characterize and classify the soil to provide more details about the subsurface morphological information. An additional objective of this project is to seek out an approximate value of soil moisture and Ph value which will help the farmers to predict the fertility of soil**

Keywords: IOT:- Internet of Things,PH.

I. INTRODUCTION:

Automatic control has various applications which includes home automation, factories- boilers, switching telephone devices, ships, air craft, etc which reduces human intervention, it reduces labour, energy, Automation can be done using two prototypes namely

• Wi- Fi, Bluetooth • Ethernet

Wi-Fi: Wireless Fidelity (Wi-Fi) is standard for (WLAN) Wireless Local Area Network. It provides us with surfing internet at broadband speed by connecting to access point . Access point has many components that provide LAN.

Ethernet: Ethernet is physical later of Local Area Network (LAN). it has become the most commonly used LAN network technology in last 30 years as it provides low cost, high speed and easy installation.

Components used for Implementation of the system Arduino Board (UNO) Arduino is micro-controller based on Atmega328, having 14 digital In/Out pins of which 6 are for PWM output, 6 are for analog input. Operates at 16 MHz, with a USB, Power jack, reset button .

Moisture Sensor Soil Moisture Sensor detects the moisture content of soil, it consists of a plurality of soil moisture sensors. Technology used is Frequency domain Sensor i.e., capacitive sensor, moisture meter characteristic of the use of water in the neutron moderator

II. THEORY

In the current situation, there are no such recommendations in the available system online and also farmer has to physically go and get the recommendations based on their soil report provided by quality executives. Each and every sector in the digital world is going through a dramatic change due to the influence of IT field. But, till date not a whole lot work has been done in the agricultural sector. The use of various data mining techniques used in agricultural section will be a continuing area of research. It will offer an internet based answer for the soil testing facilities and as a free messages for the farmer that contains information on soil testing code and the chemical that is important for the crop and additionally the essential knowledge. Additionally the farmers can specify their next crop whereas they furnish their sample to scantiest therefore in keeping with next crop the chemical can recommend. The results which are obtained supported the classification of contains that should be gif tin soil and in keeping with result report are generated. The ultimate goal is to increase the growth of the agricultural sector

For the future developments, it can be enhanced by developing the system for large acres of land. Also, the system can be integrated with additional hardware to check the quality of the soil and the growth of crop in the soil. The sensors and the microcontrollers are successfully interfaced and wireless communication is achieved between various nodes. Also, the system can further improve its function by adding machine learning algorithms, which are able to learn and understand the requirements of the crop. This would help in making the field an automatic system. The observations and the obtained results tell us that this solution can be implemented for reduction of water loss and reduce the amount of man power required for a field. Agriculture sector is a field that is technically lagging in the implementation of new world innovation and applications based on modern techniques. Our proposed system will make optimized use of the sources and solve the problems like water shortage. The data is stored in the server. Based on the available conditions, a data would be retrieved.

DHT 11 Sensor:

The DHT11 is the basic, low-cost digital temperature and the humidity sensor. It gives out a digital value and hence we can give its output directly to the data pin instead of an ADC. It has a capacitive sensor for the measuring of humidity. The only real shortcoming of this sensor is that the user can only get new data from it only after every 2 seconds.



Figure: DHT 11 Sensor

Soil Moisture sensor(YL-69):

The soil moisture sensor measures the water content in the soil. It uses the property of electrical resistance of the soil. The relationship between the measured property and soil moisture is calibrated and it varies depending on the environmental factors such as temperature, soil type, or the electric conductivity. Here, It is used to sense the moisture in the field and transfer it to the raspberry pi in order to take controlling action of switching water pump ON/OFF.



Figure:: Soil Moisture sensor(YL-69)

ARDUINO Mega 2560:

The ARDUINO Mega 2560 have 54 digital I/O pins (of which 15 can be used as the PWM outputs), 4 UART's (hardware serial ports), 16 analog inputs, a 16 MHz crystal oscillator, a port to the USB connect, a power jack, an ICSP header, and a reset button. It contains everything that is needed to support the microcontrollers; just connect it to a computer with the USB or power it with an AC-to-DC adapter or a battery to get it started. This board can be powered by the USB an connection or external supply. The power source is then chosen automatically. Now rather then pressing a physical button to reset the MCU, it can be done from the software.



Figure: ARDUINO Mega 2560

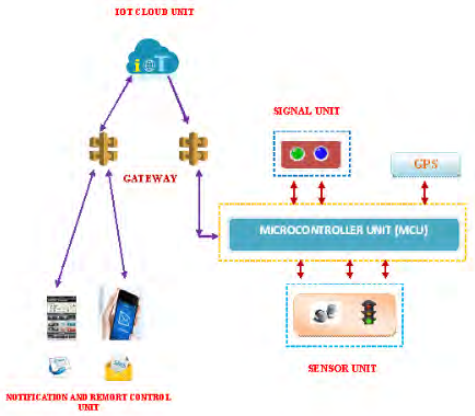
Relay:

A relay is a electrically operated operating principle are also used, such as the solid-state relays. Relays are used where it is necessary to control a circuit switch. Many relays use a electromagnet to mechanically operate a switch, but other by the separate low-power signal, or where several circuits must be controlled by just one signal.

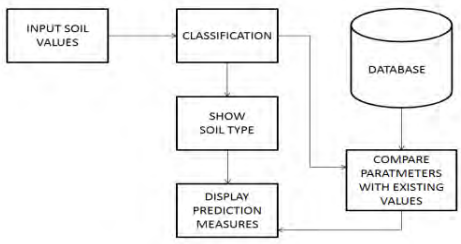


III. PROPOSED WORK:

System Architechture:



Proposed work have many advantages over the existing : Here, we will use IOT to upload data on cloud using wifi module.We will use Ph meter, moisture sensor and temprature sensor which can be accessed by the farmers.





IV. RESULTS AND DISCUSSIONS

The researchers in the agriculture field and farmers deploy sensors at their agricultural-lands to obtain the data on temperature, humidity, soil moisture and so on. The automatic collection of these data greatly helps the analytical works. Currently, they rely on network providers, for example, cellular phone network to achieve the wanted automatic collection from their remote sites, which is not feasible or affordable for most of the farmers due to its high operational cost.

The review and meta-analysis of the yield data comparing organic and conventional agriculture showed that the current organic yields of an individual crop are on average 80% of conventional yields . The analysis of 362 datasets also showed a high variation in the yield gap of organic agriculture (standard deviation 21%). Some of this variation seems to be systematic.

E.g. soybeans and some other pulses, rice and corn score higher than 80% and wheat, barley and potato scoring lower than 80%. Most regions have relative production fairly close to the overall average.

In a specific research carried out by Leisa J. Armstrong, comparative study of current data mining techniques used for research such as cluster analysis and statistical methods was carried out to establish the most effective and efficient technique .They used a large set of data extracted from the Western Australia Department of Agriculture and Food (AGRIC) soils database to conduct the research. The experiment analyzed a small number of traits contained within the dataset to determine their effectiveness when compared with standard statistical techniques.

In the current situation, there are no such recommendations in the available system online and also farmer has to physically go and get the recommendations based on their soil report provided by quality executives. Each and every sector in the digital world is going through a dramatic change due to the influence of IT field. But, till date not a whole lot work has been done in the agricultural sector. The use of various data mining techniques used in agricultural section will be a continuing area of research. It will offer an internet based answer for the soil testing facilities and as a free messages for the farmer that contains information on soil testing code and the chemical that is important for the crop and additionally the essential knowledge. Additionally the farmers can specify their next crop whereas they furnish their sample to scantiest therefore in keeping with next crop the chemical can recommend. The results which are obtained supported the classification of contains that should be gif tin soil and in keeping with result report are generated. The ultimate goal is to increase the growth of the agricultural sector.

V. CONCLUSION

From this we can know the moisture, Ph level and temperature of the soil which will help in predicting that which crops should be grown in that soil which will help the farmers. This system will characterize and classify the soil to provide more details about the subsurface morphological information. An additional objective of this project is to seek

out an approximate value of soil moisture and Ph value which will help the farmers to predict the fertility of soil.

FUTURE SCOPE:
For the future developments, it can be enhanced by developing the system for large acres of land. Also, the system can be integrated with additional hardware to check the quality of the soil and the growth of crop in the soil. The sensors and the microcontrollers are successfully interfaced and wireless communication is achieved between various nodes. Also, the system can further improve its function by adding machine learning algorithms, which are able to learn and understand the requirements of the crop. This would help in making the field an automatic system. The observations and the obtained results tell us that this solution can be implemented for reduction of water loss and reduce the amount of man power required for a field. Agriculture sector is a field that is technically lagging in the implementation of new world innovation and applications based on modern techniques. Our proposed system will make optimized use of the sources and solve the problems like water shortage. The data is stored in the server. Based on the available conditions, a data would be retrieved.

VII. LITERATURE REVIEW

We have studied various papers in order to know the previously made soil moisture measurement system using various techniques. The technologies and the methods of some of the papers are listed below.
In paper titled "The Gravimetric Method of Soil Moisture Measurement"

by S.G Renoylds[1], a sampling device such that a screw or coring auger is done for moisture determination. After 7 days storage of soil sample the moisture losses were observed to be almost negligible. Since the rate of soil drying varies from soil to soil hence preliminary were undertaken. The materials used were sampling instruments, sample containers, balances, oven and drying time. The data was entered manually and no sensors were there and hence it required a lot of labor and Time Consuming as the samples has to be stored for days. The Gravimetric method is the only way of soil measurement and hence it is indispensable for calibrating way of soil moisture measurement method. It has an advantage that it requires only simple, inexpensive instruments and just as soon as the experienced has been gained large areas can be sampled.

In paper titled "Watermark Soil Moisture Sensors" by Julie Chard[2]: Watermark sensors are used for soil moisture measurement. Watermark sensors are not that expensive and are used for measuring soil water potential over a wide range(0-2 bars or 0 to 200 kpa).Watermark sensors can be easily installed, low maintenance and compact. Watermark sensors consist of two concentric electrodes embedded in reference material matrix. It is usable over several seasons with proper care continuous measurements at same location but watermark method lack in accuracy.

In paper titled "Design and Test of a Soil Profile Moisture Sensor Based on Sensitive Soil Layers" by Zehran Gao and Yan Zhu[3] :Soil moisture sensor was designed based on high frequency capacitance. There are five sensors groups of a data processor, sensing probes and some accessory components. Copper rings of low resistivity were used as components of sensing probe. For realization of intelligent farmland irrigation accurate measurement of soil moisture is necessary. The main role of data acquisition is to know the output frequencies of sensing probes. Data processing was designed to call the digital filtering module to remove the invalid data among multiple frequency signals. It has ability to read soil volumetric water content directly and no special maintenance is necessary but it is expensive.

Table 2.1: Comparative Study

Sr No.	Title	Author	Advantages
1	The Gravimetric Method of Soil Moisture Measurement	S.G. REYNOLDS	1.Very accurate
2	Watermark Soil Moisture Sensors	JULIE CHARD	1. Good accuracy in the medium to fine soils due to their fine sized particle similar to its inner granular matrix. 2. Affordable(about 20$ per sensor, 250$ for the meter). 3. Easy handling(light weight, pocket size,easy installation and direct reading) 4. Larger moisture reading range(0 to 200cb, or kPa 5. Usable over several seasons with proper care. Continuous measurements at same location.
3	Design and Test of a Soil Profile Moisture Sensor Based on Sensitiv	YAN ZHU,ZHENRAN GAO	1.Ability to read soil volumetric water content directly 2.No special maintenance necessary 3.High accurate when sensors are installed properly in good contact with soil

	e Soil Layers		4.Large range of operating environment(90 to 50 degree Celsius) and range of measurement(0% to saturated water content) Continuous measurements at same location.

ACKNOWLEDGMENT

We sincerely thank our guide Ms. Vaishali Nirgude for her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work. We thank the HOD, Dr. Sheetal Rathi, Dean Academic Dr. R.R. Sedamkar, the Principal, Dr. B. K. Mishra and the college management for their support.

REFERERENCES

[1] Ahmed, A.U., R.V. Hill, L.C. Smith, and D.M. Wiesmann. 2007. The World's Most Deprived: Characteristics and Causes of Extreme Poverty and Hunger.
[2] Allen, J.C., and N. Mar. 1985. Soil Response to Forest Clearing in the United States and the Tropics : Geological and Biological Factors Soil Response to Forest Clearing in the United States and the Tropics : Geological and Biological Factors1. 17(1): 15–27.
[3] Andersson, J. a., and S. D'Souza. 2014. From adoption claims to understanding farmers and contexts: A literature review of Conservation Agriculture (CA) adoption among smallholder farmers in southern Africa. Agric. Ecosyst. Environ. 187: 116–132Available at http://linkinghub.elsevier.com/retrieve/pii/S0167880913002685 (verified 28 May 2014).
[4] Anthofer, J. 2004. The Potential of the SRI for Poverty Reduction in Cambodia. Consultancy Report. FSNPSP
[5] Asfaw, S., B. Shiferaw, F. Simtowe, and L. Lipper. 2012. Impact of modern agricultural technologies on smallholder welfare: Evidence from Tanzania and Ethiopia. Food Policy 37(3): 283–295Available at http://linkinghub.elsevier.com/retrieve/pii/S0306919212000176 (verified 4 June 2014).
[6] Baker, C.J., K.E. Saxton, W.R. Ritchie, W.C.T. Chamen, D.C. Reicosky, and F.R.S.E. Justice. 2006. No-tillage seeding in conservation agriculture (CJ Baker, KE Saxton, WR Ritchie, WCT Chamen, DC Reicosky, MFS Ribeiro, SE Justice, and PR Hobbs, Eds.). Available at http://www.cabi.org/CABeBooks/default.aspx?site=107&page=45&LoadModule=PDFHier&BookID=486.
[7]Topp, G. C. (2003), State of the art of measuring soil water content. Hydrological Processes, 17, 2993-2996.
[8] Šanda, M., Haase, T., Wild, J. and Jankovec, J. (2014), TMS3: Temperature and moisture system for spatially distributed measurements. 15th Biennial Conference of the Euromediterreanean

Disease Prediction System Using Data Mining

Surabhi Pandey, Gauraja Parab, Srushti Sharma, Rashmi Thakur

Computer Engineering, TCET, University of Mumbai, India
surabhi.pandey48@gmail.com, gauraja25@gmail.com, srushtish1997@gmail.com, thakurrashmik@gmail.com

***Abstract*— Healthcare is a sector where decisions usually have very high-risk and high-cost associated with them. One bad choice can cost a person's life. With diseases like Swine Flu on the rise, which have symptoms quite similar to common cold, it's very difficult for people to differentiate between medical conditions. We propose a novel method for recognition of diseases based on symptoms and display do's and don'ts for that particular disease. We do this by assigning different coefficients to each symptom of a disease, and filtering the dataset with the severity score assigned to each symptom by the user. The diseases are identified based on a numerical value calculated in the fashion mentioned above.**

Keywords— Data Mining, Decision Tree, Common platform, Automation, Mainstream.

I. INTRODUCTION

The successful application of data prediction in highly visible fields like e-business and marketing has led to its application in other industries and sectors. Among these sectors just discovering is healthcare. The healthcare environment is still information rich but knowledge poor. There is a wealth of data available within the healthcare systems. This project intends to provide current techniques of knowledge discovery using data prediction techniques that are in use in today's medical research particularly in Disease Prediction. Number of experiments has been conducted to compare the performance of predictive data prediction. Medical data prediction has huge potential for exploring the hidden patterns in the data sets of the medical domain. The occurrence of every disease shows a pattern based on its symptoms. The main focus of this project is to propose a system to exploit these symptoms for predicting the associated diseases. The core idea behind this was that every symptom of a disease has a unique impact on severity and recovery time.

By prediction we mean to forecast an occurrence of a condition based on some mathematical calculation. For implementing this prediction, we need a recommender system. A recommender system is a system which reads an input, finds a pattern in it which is based on the dataset given to train the system. Based on the pattern it figures out a solution for the problem. A naive solution can be to create a database of every possible disease and its symptoms and predicting diseases based on that. The biggest drawback about this solution is that the efficiency and speed of this solution are very less and the size of this dataset would be very large.

II. PROPOSED WORK

Developing front end using HTML, CSS and JSP (Java Server Pages). Python is used in the development of Backend. Decision Tree algorithm is implemented for the disease prediction system. SQL is used for the database.

III. METHODOLOGY

With big data growth in biomedical and healthcare communities, accurate analysis of medical data benefits early disease detection, patient care and community services is necessary. The symptoms are common in human, but for handling errors sometimes occur diagnosis that can lead to death. Such errors can occur due to lack of expertise or reluctance of patients to check themselves since the symptom is common.

By considering the issues, we conduct a study to design an application that can help patients with the disease symptoms. Hence, agile methodology is used in the iteration of the development process followed by testing in the duration software development lifecycle.

IV. TECHNOLOGY

Data mining technology is used for unearthing of patterns in the database (i.e. electronic health record). The goal of data mining is an intelligent extraction of meaning from the given dataset. It is the automatic or semi-automatic analysis of data using database techniques and analytics.

The following steps are major phases int the data mining process- Extraction, Transformation and Loading of Data in data warehouse - managing of data in multi dimension databases; analysis using algorithms and presentation of the result. It involves management and processing of data followed by post processing and visualization of it. Decision Tree algorithm is implemented for the prediction of diseases in the general population.

Languages used –

- Front end: HTML, CSS, JSP
- Back end: Java, Python
- Database: SQL

V. LITERATURE SURVEY

Wasan et. al. have proposed application of various data mining techniques as diagnostic tools to identify patterns in medical data. They have identified knowledge discovery in hospital management systems as a potential field which can benefit from application of such

techniques. Data mining is very useful in discovering patterns in large amounts of data. Medical diagnostics is one such field where pattern discovery can be of immense use.

The applications of such techniques have been presented by Scales et. al. Durairaj and Ranjani have performed a comparative study of data mining algorithms and tools on various diseases. They have also analyzed the success rate of medical techniques over existing datasets. They have concluded that combination of multiple data mining methods may yield better results in medical domain.

A method for predicting disease risk through feature selection has been proposed by Yang et. al. They have applied random forest and SVM techniques for this purpose on multiple UCI datasets.

Meisam Shabanpoor and Mehregan Mahdavi proposed a methodology for prediction of diseases and their cure time based on symptoms. Their method classifies diseases into groups based on age and Body Mass Index. They have developed a collaborative filtering method which considers neighborhood selection for prediction. Their method does not take into account different coefficients/weights for different symptoms of a disease.

S Sudha and S Vijayarani suggested to use data mining techniques to predict diseases mainly of three kinds. They focused more on heart diseases, diabetes and breast cancer. They used different algorithms for predicting different diseases.

VI. Feasibility Study

The project feasibility study is a document containing a detailed description of the project, followed by a set of different feasibility areas. These are aspects of the project that will drive the success or failure of the project. This study will provide the necessary information so that you can decide whether or not your project will begin or whether it has a shot at success.

A. *Market Feasibility*

- Our project is a needful product and service.
- Most of the persons are need this product for security purpose.
- Any normal user can use this product.
- Disease prediction is the system is similar product/service currently available but that cannot provide information about all disease.
- For our product use customer don’t want to pay any money.

B. *Technical Feasibility*

- Our project is possible for implementation.
- It is possible to create this product/service.
- We have the necessary expertise or technical resources (hardware/software).
- Our project do not rely on 3rd party products that do not already exist

C. *Financial/Economic Feasibility*

- We have the money to complete the project.
- Our project is technically feasible, and there is a market for our product or service.
- Our project is not as much as expensive. That is, we have sufficient budget for your project.
- You will need to conduct a primary investigation into how expensive the project may be.
- We just require laptop for implementing our project.
- It will take 4-5 months for research and development.
- No any additional resources project needed.

VII. Process Model Identification

In a country like India, where population is high, it is important for a person to have access to basic health-care facilities. Further with the basic facilities it is necessary to have the infrastructure and resources for the smooth service to the people. Therefore, we come up with a common platform for a general person to access doctors and medical campaigns, proposing a novel method for recognition of diseases based on symptoms and display do’s and don'ts for that particular disease.
This portal will be developed using waterfall software development model. The development will be divided in five phases:

- Analysis
- Design
- Coding
- Testing
- Deployment

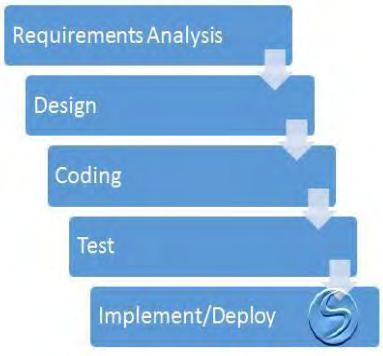


Fig. 1. Waterfall Model

VIII. OBJECTIVES OF PROJECT

In this modern society where large number of humans follow a sedentary lifestyle following an 8-hour job cycle, which is becoming one of the leading causes of mortality worldwide.

The computers at the hospitals of the healthcare industries are used to collect huge amounts of data of the patients and their ailments. This huge repository of data is a treasure of knowledge. The hidden patterns and relationships in the data is mostly overlooked.

With the data expansion in biomedical and healthcare sector, accurate analysis of medical data aids early disease detection and patient care. However, the accuracy of analysis is decreased when the quality of medical data is insufficient. Also, different regions exhibit unique characteristics of certain regional diseases, which may weaken the prognosis of any disease outbreaks. Symptoms of a disease is one of the criteria that can interfere with human health. The symptoms maybe common in many patients, but during the diagnosis there is a chance of some occurrence of an error which may lead to a death. Such errors can occur due to lack of expertise or reluctance of patients to check themselves since the symptom is common.

By considering these issues, we conducted a study to design an application that can help patients with such disease symptoms.

Main objectives of the project are listed as follows-

- General public can analyze their afflicted health symptoms and take a decision to further contact a doctor.
- They're provided with general do's and don'ts and can further opt for a medical campaign.
- Associations involved in healthcare can avail the necessary information for the conduct of campaign.
- Doctors get a chance to join Digital India Environment by providing their expertise to the public.

IX. OBJECT DIAGRAM

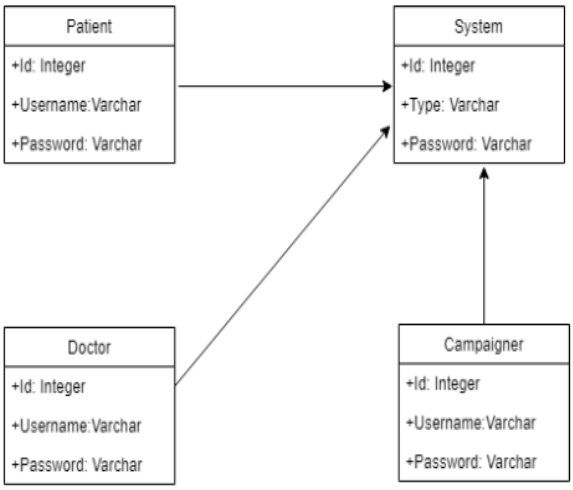


Fig. 2. Object Diagram

X. ARCHITECHTURE DIAGRAM

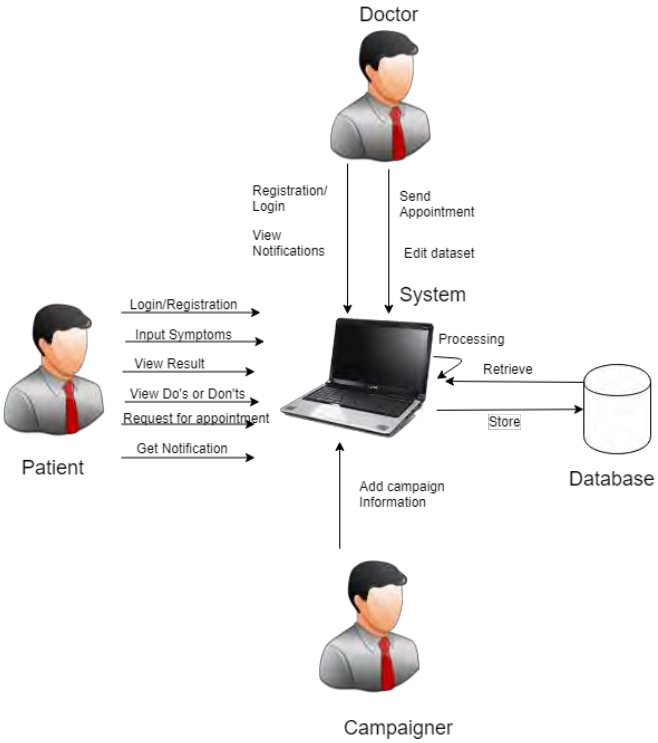


Fig. 3. Software Architecture Diagram

XI. APPLICATION OF PROJECT

- Use in Healthcare Premises.
- Patient can use it at home easily.
- Can be used by institutions to hold medical campaigns.

XII. CONCLUSION & FUTURE SCOPE

This project proposes a new model for developing an integrated disease prediction tool which uses available patient data. The data is obtained online. The system is valuable for patients to become aware of possible diseases and take important actions instantly when necessary. The potential limitation of the proposed prediction model is that it produces relatively few outcomes in the form of disease predictions due to a lower number of diseases. Also, it requires being evaluated by more external evaluators. Despite the few outcomes, our model performed quite well by showing promising results. These limitations highlight the need for further research in this important but neglected topic.

ACKNOWLEDGMENT

We thank our college Thakur College of Engineering and Technology for giving us the opportunity for being given a chance for presenting this research paper. Guided under the Computer Science Department,

we are thankful to our Mentor Mrs. Rashmi Thakur, Assistant Professor Computer Science Department, Thakur College of Engineering and Technology, for his patient guidance, enthusiastic encouragement and useful critiques of this research.

REFERENCES

[1] Diaz-Chao A, Torrent 1., La Casta D., Saigi F. Improving integrated care: Modelling the performance of an online community of practice. International Journal of Integrated Care, PMCID: PMC3952812, 2014.

[2] WebMD. Disease symptoms, 2015. [Online; accessed 22-January 2015].

[3] Wikipedia. List of medical symptoms, 2015. [Online; accessed 22-January-2015].

[4] A. R. Aronson. Effective mapping of biomedical text to the UMLS metathesaurus: the metamap program. In AMIA, Annual Symposium, pages 17–21, 2001.

[5] A. Babashzadeh, J. Huang, and M. Daoud. Exploiting semantics for improving clinical information retrieval. In SIGIR, pages 801–804,

Implementation Of Fake Fingerprint Image Detection System

Akshay Kumbhar
Computer Engineering
Thakur college of engineering and technology
Mumbai, India
akshay.kumbhar7@gmail.com

Akshay Jaiswal
Computer Engineering
Thakur college of engineering and technology
Mumbai, India
aj6242429@gmail.com

Saurabh Yadav
Computer Engineering
Thakur college of engineering and technology
Mumbai, India
saurabhy051@gmail.com

Mrs.Jesalkumari Varolia
Computer Engineering
Thakur college of engineering and technology
Mumbai, India
Jesaljoshi29@gmail.com

Abstract— **Digital imaging has experienced tremendous growth in recent decades, and digital camera images have been used in a growing number of applications. With such increasing popularity and the availability of low-cost image editing software, the integrity of digital image content can no longer be taken for granted. Here we introduce a new methodology for the forensic analysis of digital camera images. The proposed method is based on the observation that many processing operations, both inside and outside acquisition devices, leave distinct intrinsic traces on digital images, and these intrinsic fingerprints can be identified and employed to verify the integrity of digital data. The intrinsic fingerprints of the various in-camera processing operations can be estimated through a detailed imaging model and its component analysis. Further processing applied to the camera captured image is modeled as a manipulation filter, for which a blind deconvolution technique is applied to obtain a linear time-invariant approximation and to estimate the intrinsic fingerprints associated with these post camera operations. The absence of camera-imposed fingerprints from a test image indicates that the test image is not a camera output and is possibly generated by other image production processes. Any change or inconsistencies among the estimated camera-imposed fingerprints, or the presence of new types of fingerprints suggest that the image has undergone some kind of processing after the initial capture, such as tampering or steganographic embedding.In this project we are using two main algorithms named as ELA(Error Level Analysis)algorithm and Metadata cheker algorithm.ELA algorithm will find the error level occurred due to editing in the JPEG image.It checks the differences in the intensity of the RGB color of the image.Because JPEG is the popular image type used by many digital cameras and smartphones nowdays.And also JPEG is the lossy type of image ELA algorithm is very useful to find out intrinsic fingerprints remained by intruder after manipulation. By this project we can check for the originality of a digital image. It can be used for nonintrusive digital image forensics.**

Keywords—Digital camera,image editinfg softwar,ELA algorithm,Metadata checker, intrinsic fingerprints,steganography

I. INTRODUCTION

The past few decades have witnessed rapid advancements in digital photography. Due to the improved technology Digital images have been used in a variety of applications, from military and reconnaissance to medical diagnosis and consumer photography. With such high popularity, accuracy, the advent of low-cost and sophisticated image editing software, the integrity of image content can no longer be taken for granted and a number of forensic-related questions arise amidst such extensive use. For example, one can readily as indirectly or directly question asks how an image was acquired? Was it captured using a digital camera, an image scanner, or was it created artificially using image editing software? Has the image undergone any manipulation after capture? Is it authentic or has it tampered in any way? Does it contain any hidden information or

steganographic data? Many of these forensic questions are related to tracing the origin of the digital image to its creation process. Evidence obtained from such forensic analysis would provide useful forensic information to law enforcement, security, and intelligence agencies. Knowledge of image-acquisition techniques can also help answer further forensic questions regarding the nature of additional processing the image has undergone after capture. In this work, we develop a novel methodology for digital image forensics of color images. We present techniques to identify the inherent traces that are left behind in a digital image when it goes through various processing blocks in the information processing chain. We refer to these traces as the intrinsic fingerprints and use them to identify the source and establish the authenticity of the digital image. We have built software that can classify intrinsic fingerprints into two categories, namely in-camera and post camera fingerprint. Using a detailed imaging model and its component analysis, we estimate the intrinsic fingerprints of the various-camera processing operations. Further processing applied to camera outputs, if any, are modeled as a filtering operation, and its coefficients are estimated to obtain the post camera fingerprints. While the absence of in-camera fingerprints suggests that the test image is not a camera output and is possibly generated by other image production processes, any change or inconsistencies among the estimated in-camera fingerprints or the presence of new post camera fingerprints indicates that the image has undergone some kind of post-camera processing.
The goal of the project is:
1.To detect any modifications on the JPEG images using Error Lvel Analysis.
2. Detecting forgeries in internet uploaded images.
3. Preventing fake images from misleading the people.

II. SCOPE OF THE PROJECT

1.It checks JPEG format of image.

2.It checks smartphone captured image.

3.It checks DSLR captured image.

4.It checks all lossy type of images**.**

.

III. LITERATURE SURVEY

In Digital Image Forensics via Intrinsic Fingerprints System[1], a set of forensic signal processing techniques to verify whether a given digital image is a direct camera output.The proposed formulation is based on the
observation that each in camera and postcamera processing operation leaves some distinct intrinsic fingerprint traces on the final image. this paper demonstrates the effectiveness of ELA and CFA concept.

Development of Photo Forensics Algorithm by Detecting Photoshop Manipulation[5] Using Error Level Analysis System,researchers presented the design and development of image forensic application on JPEG images. Error level analysis was used to detect any modification on the JPEG images

Fake Colorized Image Detection System proposed by Yuanfang Guo[6], aimed to address a new problem in the field of fake image detection: fake colorized image detection. We observed that fake colorized images and their corresponding natural images possess statistical differences in the hue,saturation, dark and bright channels. We proposed two simple yet effective schemes, FCID-HIST and FCID-FE, to resolve this detection problem.

IV.EXISTING SYSTEMS VS PROPOSED SYSTEM

There are still no useful softwares to detect forged or manipulated images in recent years. So this software helps to detect forged images

IV. PROPOSED METHODOLOGY

1.Project Description:

There are not much Applications/Sites available in market which can detect the changes made in any pictures. Due to this Image tempering is growing in this technological world. Many crimes and blackmailing is being done by showing edited pictures,one can edit image and cause harm to the victim. So we are trying to detect those edited part in the image with the help of various algorithms.With the help of these algorithms we

can tell the difference between original image and the edited image.

2. Proposed Work:

The main objective of the system is to detect forged images.This system uses following algorithm:

1.Error Level Analysis Algorithm(ELA):

Error Level Analysis evaluates the quality level for grids squared within the images. They present an increased degree of error during successive resave operations. The phenomenon is obvious if images aren't optimized for a specified camera quality level. Subsequent resaves reduce the error level potential, producing a darker ELA. After a number of resaves, the grid square reaches its minimum error level.It means that ELA checks the RGB color intensity in the whole image like edges,surface of the image and texture etc.

Edges:Similar edges should have same coloring across the image.all high contrast and high bright edges should be similar to each other..And all low contrast and low bright edges should be similar to each other in the ELA analysis.

Surface : Regardless of the actual color of the surface, all flat surfaces should have about the same coloring under ELA.

2.Metadata Checking:

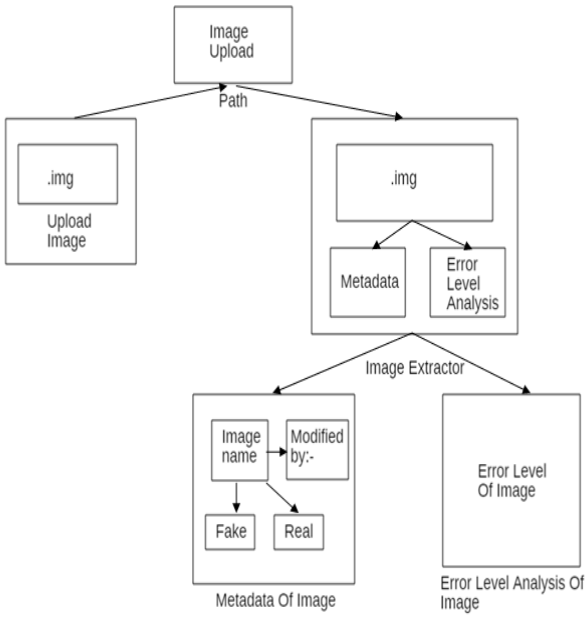
At first level, it checks the image metadata.Metadata is

all of the information of photos that taken on any

digital camera.Nowdays digital camera have computer inside them,we can get time and date of the photo,name of the camera or smartphone used,which lens was used,location of the photo etc.For example, if an image is edited with Adobe Photoshop, the metadata will contain even the version of the Adobe Photoshop used.Metadata remains same even if we forward the image to someone.

3. System Architecture:

The overall system architecture is shown below:



I . Economic Feasibility :

- **Cost Benefit Analysis:**

The Cost Benefit Analysis summarizes the revenues and costs involved with the proposed project. As the proposed model will be used for the benefits of society so no additional cost will be paid by them.Only camera and smartphone is included in our project, so the hardware cost gets minimized. Only minimal amount will be required by certain software. Hence, our system is Cost Efficient.

II. Technical Feasibility :

Resources Required: , Programmers,testers,debuggers
Software required:MATLAB
Hardware required:PC for development,smartphone or digital camera
Language used:MATLAB,Java

III. Managerial Feasibility :

Management support, employee involvement, and commitment are key elements required to gauge managerial feasibility in the proposed project. The success and the profitability of the project partly depend on managerial competence of the major ingredients of the proposed project which are the users.

IV. Operational Feasibility:

The proposed model is to detect morphed or edited JPEG images.To do that we will be using MATLAB software and some algorithms.First,we have to send the image to MATLAB software then software will run the codes and shows edited area of the given image.

V. RESULTS AND DISCUSSION

Originality and authenticity of an image is usually tough to determine due to various image editing softewares like Adobe Photoshop are nowdays easily available in market.In order to detect manipulated images some techniques introduced in last decade.Image forensics aims to detect the manipulation of digital images.Here,we aimed to address a problem in the field of Digital Image Forensics.Using ELA (Error Level Analysis) and Metadata Analysis photoshoped images can be identified which are not seen by naked eye.ELA algorithm will identify the error level in the photoshoped image in terms of color brightness of the edges ,texture of an image and the surface of an image.ELA will give the detailed information about edited part by checking each pixel of the image which contains subpixel of RGB color.ELA highlights the forged part in the image and tells that in that particular area manipulation is done and metadata analysis gives the all the information about that image for example.location of captured image,which device is used while capturing an image,modified bye which device etc.

Limitations:ELA algorithm only analyse JPEG image.
After every successive resave of an image some information ,colors are reduced.It will not work for PNG,GIF fomat images.

VI. CONCLUSION

The proposed system is to detect forgery in digital images. The original image captured by digital camera or smartphone normally is saved in the JPEG format due to its popularity. For resaving operation, each block should degrade at approximately the same rate due to the introduction of similar amount of errors across the entire image. For modified image, the altered blocks should have higher error potential compred to the remaining part of the image.We have studied Error Level Analysis which gives error introduced in resaved modified JPEG image . It is among the recent developments in the digital world.

VII. FUTURE SCOPE

The project can be further used to find whether the image is steganographed or not. And the steganographed message can be recovered. Since we have studied the various algorithms that could be used to identify patterns present in the digital image. The project will be implemented successfully.

VIII.ACKNOWLEDGEMENT

We sincerely thank our guide Mrs.JesalKumari Varolia for his/her guidance and constant support and also for the stick to our backs. We also thank the project coordinators for arranging the necessary facilities to carry out the project work.

We thank the HOD, Dr. Sheetal Rathi, Dean Academic, Dr. R.R Sedamkar, the Principal, Dr. B. K. Mishra and the college management for their support.

REFERENCES

[1] Ashwin Swaminathan, Min Wu and K.J.Ray.Liu, “DigitalImage Forensics via Intrinsic Fingerprints” on Information Forensic and Security, Vol. 3, No. 1, March 2008J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] [2]P. Ferrara, T. Bianchi, A. D. Rosa, and A. Piva, “Image forgery localization via fine-grained analysis of CFA artifacts,” IEEE Trans. Inf. Forensics Security,Oct2012

[3] G. Chierchia, G. Poggi, C. Sansone, and L. Verdoliva, “A
Bayesian-MRF approach for PRNU-based image forgery
detection,” IEEE Trans.Inf. Forensics Security, vol. 9, no. 4, pp.
554–567, Apr. 2014

[4] R. Zhang, P. Isola, and A. A. Efros, “Colorful image colorization,” in Proc.
Eur. Conf. Comput. Vis. (ECCV), 2016, pp. 649–666

[5] Teddy Surya Gunawan,. Siti Amalina Mohammad Hanafiah ,Mira Kartiwi ,Nanang

Ismail “Development of Photo Forensics Algorithm by Detecting Photoshop Manipulation Using Error Level Analysis” 10.11591/ijeecs.v7.i1.pp131-137,2017

[6] Fake Colorized Image Detection Yuanfang Guo , Member, IEEE, Xiaochun Cao, Senior Member, IEEE, Wei Zhang, Member, IEEE, and Rui Wang, Member, IEEE,2018

Business Intelligence using Internet of Things (IoT) : A Case Study

Avani Agrawal, Pooja Patil, Rajesh Bansode
Thakur College of Engineering & Technology University Of Mumbai Mumbai, India
avaniagrawal27@hotmail.com, ppatil879@gmail.com, rajesh.bansode@thakureducation.org

***Abstract*—The Internet of Things (IoT) is an emerging concept and thatis currently under creation and development. However, it hasalready made an impact on many research domains byproviding new solutions and ideas particularly in collecting various information from the various type of systems. The IoT concept has provided anew research path that is conducive with gathering massive data from the system and to manage the data by applying different techniques for making it intelligent for Business. This paper focuses on how you can gather information from IoT and use the information intelligently for your business by not making much effort.The 3 layers of the IoT model has been extended to 4-layersby splitting the application layer into knowledge managementand visualization layer respectively. A business intelligence using IoTconcept has been chosen, based on a casestudy of a variousbusiness where you have to analyses the data manually and make decision for the Business. This concept of Business Intelligence using IoT will help us understand how we can collect the data and use it for the business precisely for profit of the business and make important and useful decision quickly.**

Keywords—Business Intelligence, Internet of Things (IoT), Big Data, Data Mining, Knowledge Management, Decision Making, Data Analytics.

I. INTRODUCTION

The internet of things, or IoT, is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. A thing in the internet of things can be a person with a heart monitor implant, a farm animal with a biochip transponder, an automobile that has built-in sensors to alert the driver when tire pressure is low or any other natural or man-made object that can be assigned an IP address and is able to transfer data over a network.Increasingly, organizations in a variety of industries are using IoT to operate more efficiently, better understand customers to deliver enhanced customer service, improve decision-making and increase the value of the business.

An IoT ecosystem consists of web-enabled smart devices that use embedded processors, sensors and communication hardware to collect, send and act on data they acquire from their environments. IoT devices share the sensor data they collect by connecting to an IoT gateway or other edge device where data is either sent to the cloud to be analyzed or analyzed locally. Sometimes, these devices communicate with other related devices and act on the information they get from one another. The devices do most of the work without human intervention, although people can interact with the devices -- for instance, to set them up, give them instructions or access the data.

IoT helps collecting all type of information from the system or the environment for which we apply the IoT technology. But if we try to make the IoT work more efficiently like we can make systems intelligent by analyzing the data collected from the system and use that information for decision making.

The potential benefits of business intelligence tools include accelerating and improving decision-making, optimizing internal business processes, increasing operational efficiency, driving new revenues and gaining competitive advantage over business rivals. BI systems can also help companies identify market trends and spot business problems that need to be addressed.

BI data can include historical information stored in a data warehouse, as well as new data gathered from source systems as it is generated, enabling BI tools to support both strategic and tactical decision-making processes.

Initially, BI tools were primarily used by data analysts and other IT professionals who ran analyses and produced reports with query results for business users. Increasingly, however, business executives and workers are using BI platforms themselves, thanks partly to the development of self-service BI and data discovery tools and dashboards.

II. BACKGROUND

IoT is one of the biggest sources of big data, which are rendered useless without analytics power. IoT interacts with big data when voluminous amounts of data are needed to be processed, trans- formed, and analyzed in high frequency. This work specifically focuses on the big data context. First, we investigate the recent literature on big data processing and analytics solutions for IoT. Second, we identify the numerous requirements for big data and analytics in IoT. Third, we taxonomized the literature. Fourth, we determine the various opportunities that are brought about by big data. Fifth, we highlight the role of data analytics in IoT applications. Sixth, we present the open research challenges that must be addressed in the future. Seventh, we conclude that the existing big data solutions in the IoT paradigm are still in their infancy and the challenges associated with them must be solved in the future. [1]

The concept of IOT has taken born from the requirement of managing, automating and exploring all devices, instruments

and sources all over the world. To deal simultaneously with people and IOT devices, there arises the need of data mining techniques. Various techniques of IOT supports in decision making and optimization of any application. When fused with IOT, data mining basically deals with discovering useful and interesting patterns from large data. Then number of algorithms is applied to extract important hidden information. Data mining techniques includes clustering, classification, pattern mining. Adding more to the introduction of IOT and data mining, this paper focuses on challenges and open issues of data mining. Characteristics of big data, data mining algorithms and analysis on challenges are also described. At this stage when the development of IOT is at initial stage, focus is more on developing efficient pre-processing mechanism and consequent developing effective mining technologies for describing rules of IOT data. In the addition to research trends going on nowadays in the world are discussed in this paper. The research trends are: (i) Data will be uploaded the internet once connected, but IOT faces come problem such as, data abstraction, data summarization and data fusion. (ii) To mine data from the application controlling multi-media devices and controlling energy are also a topic of research. [2]

New, breakthrough technologies, in particular the Internet of Things, change the way knowledge is managed in organizations, forcing the implementation of innovative knowledge management systems and an open approach to supporting knowledge flows. Bearing in mind their importance already today, IoT and Big Data most definitely are key factors affecting societal development in the future. Private and public organizations have begun to gain critical insights from the Big Data and ubiquitous technology through various management systems. Basically, the issue at stake here is the fact that it is not just the question how to manage and control the technological possibilities. The development also concern leadership functions. Namely, taking seriously Internet of Things and ubiquitous technology may lead towards the revolution of digitalization which effects on knowledge management processes in organizations. The authors intend to conduct further, more detailed research on the impact of IoT on knowledge management processes in organizations, such as: locating knowledge, acquiring knowledge, developing knowledge, sharing, using and protecting knowledge. [3]

Prescriptive analytics goes one step further by recommending (or prescribing) an action to keep the machine from failing. This could mean scaling back the RPMs of the machine while informing the operator that a service call is necessary. If a catastrophic failure is predicted, the machine could be shut down prior to such failure to avoid higher replacement cost, injury or even death. However, Prescriptive Analytics does not just rely on what has occurred in the past or is likely to happen in the future, but takes desired outcomes, specific scenarios as well as all current and historic data into consideration when making a recommendation. This will allow us to prescribe the optimal way to handle a future scenario and goes far beyond the example of a maintenance issue or machine failure. [4]

III. PERFORMING BUSINESS INTELLIGENCE USING INTERNET OF THINGS

It is obvious that IOT may create data consisting of much useful information. But, recently technical issues and challenges are major research topic along with different methods of handling these data. It is possible to solve the problem of large IOT data is to design sensors capable of collecting only the interesting and useful data. To reduce the complexity of input data is the recent research trend. Distributed computing, cloud computing and feature selection are the popular data handling methods. It is assumed that big data will give born to more number of patterns from services and applications.
Moreover, how to handle large data obtained from IOT devices and finding hidden information from the data is important task. As described, data analysis for sensors and devices will be helpful in developing some useful system for the smart city or smart home. Many possible applications are possible to be developed from the large data analysis process. For fulfilling the task of finding hidden information from big data, Knowledge Discovery in Databases (KDD) successfully applied to different domains. Applied KDD has capability of finding “Something” or “interesting pattern” from IOT, with the help of following steps: data gathering, preprocessing, data mining and evaluation or decision making. Of these, data mining step play a main role in extracting interesting pattern or knowledge from the big data. Data fusion, data transmission, big data and data decentralized issues has an impact on the system performance and IOT service quality.

The proposed framework is composed of three components which are big data management, IoT sensors, and data analytics. The analytics are performed in real- time in order to be used in different parts of the smart building to manage the oxygen level, smoke/hazardous gases, and luminosity. The framework is implemented in Cloudera Hadoop distribution where the analytics is performed using PySpark. The results show that the framework can be utilized for IoT enabled big data analytics. The proposed framework is specifically designed for smart buildings that should be extended to make it generalize so that it can deal with other IoT applications including smart cities and smart airplanes. This system, which focuses on industrial big data analytics, integrates various data analytics components in the form of reconfigurable and interchangeable modules to meet different business needs. The authors also provide a new context intelligence framework that can help handle industrial informatics based on the sensors, locations, and unstructured data for big data mining. A case study is also performed to illustrate the design of the proposed cyber physical system.

Organizations employ information and knowledge both for the improvement of their decision-making processes and for the legitimization of decisions (including those issued on the basis of unreliable information). In addition, it seems valid to assume that modern organizations retrieve and store much more information than needed for their decision-making purposes, and still call for more. The overabundant supply of information may come as a result of various forces and trends at play. For instance, organizations tend to underestimate the cost of information retrieval in proportion to the benefits offered. Usually, decisions made with respect to information are delegated to the authority of units other than those directly involved in its retrieval. By maintaining such clear separation between information use and information retrieval/storage, decision makers are able to initiate information retrieval procedures they deem of potential value for them, but without taking heed of the real cost-benefit ratio as perceived from the

viewpoint of the organization as a whole. The rationality of information oversupply can also be viewed in the context of strategic value of information, particularly when such information is employed to enforce certain response rather than serve as basis for rational decisions. In organizational settings, information is rarely of neutral character. Most of the information feed is prone to misinterpretation. From the viewpoint of decision makers, it is better to have access to information deemed unproductive (for decision-making purposes) than to bear the risk of missing information that may ultimately prove valuable. Both the Internet of Things and cloud computing offer potential to simplify the identification of requirements to be addressed by knowledge management systems and to improve their effectiveness by providing real-time access to data/information ("just in time" and "right on point"). The Internet of Things is also perceived as another important factor to affect the transaction processes, most specifically: the methods for identifying and satisfying customer needs and requirements. Based on different recent reports and surveys, the number of objects connected to the Internet, including smart phones, is already in excess of the total Earth population and, according to the Gartner report, will reach 26 billion wireless devices by the end of 2020.

Knowing that there are hundreds, if not thousands of articles on IoT out there already, you may wonder why I've decided to add to the pile. I recently completed a consulting opportunity with an IoT SaaS company and experienced first-hand the challenges, opportunities and impact of the Internet of Things.

As businesses continue to generate more and more data over time, chances are as Business Intelligence Consultants, we will continue working with larger and larger datasets, which will require new techniques to efficiently store, process and analyze the data.

It's not like storing, processing and analyzing massive data hasn't been done before. In fact, it is done every day by companies all around the world, including recognizable household names such as Netflix, Facebook and Google.

However, the impact of IoT and the general growth in data will require more of us to become familiar with building, maintaining and tuning environments which can handle this load efficiently. Technologies will continue emerging to spur this innovation and as BI Consultants, this should excite us and make us want to continue developing our knowledge and skill set.

Data: Characteristics of data such as distribution, representation, size etc. plays a vital role in data mining. Various data normally need to be processed differently. For example, Ri and Rj may be similar or not, but they need to be analyze differently if the semantics of the data is various.

Mining Algorithm: Data mining algorithm can be designed easily with the specified needs of objective and data application. Here the question arises that, which data mining algorithm is suitable for increasing high performance of the system or for the finding of a better service in various IOT environments .There are different techniques for extracting big data from IOT. These techniques are termed as Classification, Clustering and Association Rule Mining.
Classification: It is used to classify labeled as well as unlabeled patterns. Labeled data basically considers the piece of unlabeled data with some information, class or tag related to it. For example, the picture of any animal is said to be the part of unlabeled data unless any information about it, its name, or tag like its sound is not clearly mentioned. Labeled data are achieved by making judgments about available piece of unlabeled data.

Clustering: It is used to classify unlabeled patterns. The example of unlabeled patterns are the samples of natural or human-created facts. In addition to this facts, the category of unlabeled patterns include photos, videos, audio recordings, x-rays, tweets, news articles, etc. There is no specific knowledge about the constituent of the unlabeled patterns, but contains only information of data.

Association: It is used to find an event from the input pattern that doesn't occur in particular order. Sequential pattern is a part of association rule mining and is used to find an event from the input pattern that occurs in particular order.

IV. CONCLUSION AND FUTURE WORK

Business intelligence (BI) is a technology-driven process for analyzing data and presenting actionable information to help executives, managers and other corporate end users make informed business decisions. BI encompasses a wide variety of tools, applications and methodologies that enable organizations to collect data from internal systems and external sources, prepare it for analysis, develop and run queries against that data, and create reports, dashboards and data visualizations to make the analytical results available to corporate decision-makers, as well as operational workers. This is why IoT helps making the work effort less. By collecting data and analysing it automatically we can reduce the task of the managers to make decision. Future work of our paper will be the visualization of the data collected by IoT.

REFERENCES

[1] Ahmed, E., Yaqoob, I., Hashem, I.A.T., Khan, I., Ahmed, A.I.A., Imran, M. and Vasilakos, A.V., 2017. "The role of big data analytics in Internet of Things". Computer Networks, 129, pp.459-471.

[2] Chovatiya, F., Prajapati, P., Vasa, J. and Patel, J., 2017, March. "A research direction on data mining with IoT". In International Conference on Information and Communication Technology for Intelligent Systems (pp. 183-190). Springer, Cham.

[3] Rot A, Sobinska M. "The Potential of the Internet of Things in Knowledge Management System". InPosition Papers of the 2018 Federated Conference on Computer Science and Information Systems (p. 63).

https://www.analytics8.com/insights/iot-predictive-prescriptive-analytics/

Smartphone Sensors based Rubik's Cube Timer

Sanam Kazi
Department of Information Technology
M. H. Saboo Siddik College of Engineering
Byculla, Mumbai – 400008, India.
sanam.kadge19@gmail.com

Aliza Shaikh
Department of Information Technology
M. H. Saboo Siddik College of Engineering
Byculla, Mumbai – 400008, India.
alizashaikh29@gmail.com

R. Ihsaan Muhiyadheen
Department of Information Technology
M. H. Saboo Siddik College of Engineering
Byculla, Mumbai – 400008, India.
ihsaanm97@hotmail.com

Tazeen Ansari
Department of Information Technology
M. H. Saboo Siddik College of Engineering
Byculla, Mumbai – 400008, India.
ansaritazeen1234@gmail.com

***Abstract* — The craze of solving the Rubik's Cube has taken over the world in the 21st Century with students, mathematicians and researchers taking to it to find ways to solve the puzzle in almost no time. The pace with which a person solves the cube is often considered as a measure of the person's IQ Level. The Rubik's Cube is also proven to increase the aptitude of people, especially students who start learning it from a very young age and practice it regularly. When this paper was being written, the World Record for the fastest cube solve was held by an Australian, Feliks Zemdegs who clocked just 4.221 seconds during the First Round of Cube for Cambodia 2018 held on May 6th, 2018. The timer used to solve these solves in an official WCA competition is called a 'StackMat timer' which has a precision of 1/1000th of a second. This device, however, is not available worldwide and the cost according to its official website is $24.99, at the time of writing this paper. So much for just a stopwatch that you find in the simplest of smartphones! In this paper, we are proposing a system that uses the different sensors available in today's smartphones to trigger an event that will stop the timer when it is in use, thus simulating the working of the WCA-recognized existing manual system.**

In this paper, we are proposing a system which includes a mobile application that makes use of sensors in the mobile phone to trigger an event that will stop the timer when it is in use. Also, there will be a statistical representation of the timings periodically. It will show a consistent increment of the performance monthly, thus simulating the working of the WCA-recognized existing system.

***Keywords* — cube timer, StackMat, Rubik's Cube, cubing, World Cube Association.**

I. Introduction

World Records are always exciting to watch in any sport and it reaches a whole new level in the cubing community when a cuber sets a new one. A World Record solve requires rigorous practice, an unparalleled muscular memory and a bit of luck. The other major factor is a perfectly functional, near-to-ideal cuber's kit: a speedcube, a cube timer (Speed Stacks StackMat G4 Pro Timer), a StackMat mat and a stopwatch. The timer is very an important aspect of any cubing competition where participants are judged based on how fast he/she is able to solve the cube. This paper proposes a novel idea where there is a sensor in the mobile phone. Whenever the user taps a button in the application, the timer starts and it flashes a green light after a delay of seconds giving an indication that the timer started. When the user finishes solving the cube, he/she will tap on the desk and the sensors in the phone will detect the sense and the accelerometer will track the acceleration value. Detection of change in acceleration value will produce vibration in the phone and this will stop the timer and show the exact time with the precision value up to 4 digits after the decimal. It will also show the consistent increment of the timings of the user along with the statistical representation of the timings of the user.

II. Current System

The World Cubing Association (WCA) is the regulatory body for cubing. It sets standards for cubing, lays down the regulations and strives to promote cubing by organizing competitions around the world. In the current system, the WCA competitors are ranked based on their timings with respect to other competitors. According to WCA Regulations Article 8a7, competitions must 'Use the authentic Speed Stacks StackMat timer (Generation 2, Generation 3 Pro Timer, or Generation 4 Pro Timer) for time measurement.' These timers are very expensive`. When the user takes the cube in his hands. He will be given 15 seconds to evaluate the cube. This is tracked by the WCA judge manually. After 8 and 12 seconds, the volunteer gives an indication to start solving the cube. The solver starts solving. As soon as he/she has solved the cube the individual needs to tap on the right and left side of the timer and the time elapsed is displayed with a precision up to 4 decimals.

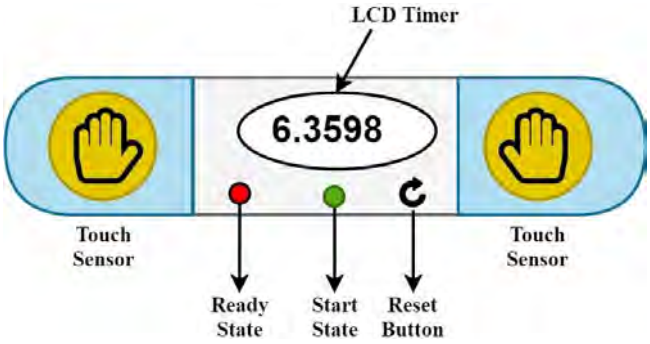


Figure 1: Representational diagram of StackMat Pro Timer (Generation 4)

A StackMat timer consists of two touchpads on either side of a timer module. The touchpads generate a trigger that is sent to the timer module. The timer consists of a Readout Display and two LED indicators (green and red). The StackMat timer displays the accurate time of the solve with a precision of 1/1000th of a second. This is really important as World Records are set and competitions are won even by a small fraction of a second such as 0.003s.

A competitor performs various prior to and after he solves the cube. A cube solve consists of the following phases:

Action 1: Inspection Time – The solver is given 15 seconds to inspect the scrambled cube and plan out his solve. He can rotate the cube in any direction but is not allowed to make any face turns. This timing is monitored by the judge and he/she intimidates the solver after 8 and 12 seconds respectively. The solver must start his solve latest by 13 seconds to avoid any penalties.

Action 2: Starting the Timer – Before those 15 seconds run out, the solver has to start his timer which eventually marks the start of his solve. He has to do this by placing his cube on the StackMat mat and then starting the timer. To start the timer, the solver has to place both his sets of fingers on each of the touchpads of the StackMat timer. A green light glow will on the timer module after a brief duration of 0.55 seconds which indicates that the timer is ready. The solver can now pick up his cube and start solving while the clock ticks.

Action 3: Stopping the Timer – When the solver has completed solving the cube, the timer needs to be stopped. He does this by tapping on the touchpad as fast as he can. The time displayed is in the precision of 4 numbers after the decimal point.

Action 4: Verifying and Entering the score – The WCA judge notes down the solve time and gets it verified by the solver. This marks the end of his solve.

III. Problems in the Existing System

- Very Expensive.
- No mechanism for keeping track of the timings of the previous solves.
- Requires that the timer device be kept in a fixed orientation.
- No support for different puzzles like 4x4, 5x5, Pyraminx, etc.
- No generation of scrambles.
- Manual entry of data.
- Manual monitoring of inspection times by the judges. This makes the inspection times to often be inaccurate.

IV. Users of the System

The system will be used by

- Cubers
- WCA Competitors
- WCA Judges

V. Proposed System

The proposed system is a smartphone application which makes use of the different sensors available in smartphones. A cube solve in the proposed system consists of the following actions:



Figure 2: Starting the Timer

Action 1: Starting the Timer – User starts the timer with a screen tap. He holds the screen for 0.55 seconds, after which the screen color changes.



Figure 3: Solving the Cube

Action 2: Solving the Cube – The user picks up the cube and starts solving it.

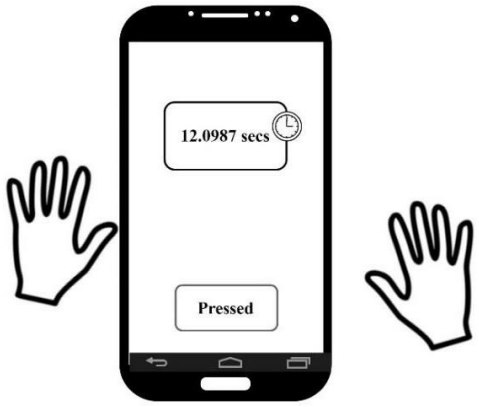


Figure 4: Stopping the Timer

Action 3: Stopping the Timer – After solving the cube, the solver taps on the table to stop the timer. The smartphone captures the physical vibration and stops the timer.

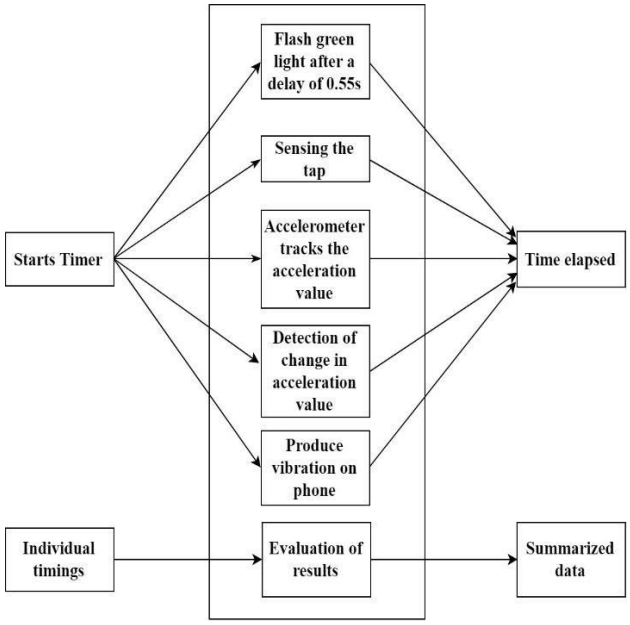


Figure 5: Block Diagram

The above figure shows that the first input block is the start timer because of which a green light is produced after a delay of 0.55 seconds and the timer starts. Then the user solves the cube. When the user taps on the desk,there is a sensor which senses the tap and the accelerometer tracks the accleration value and produced vibration in the phone. Detection of change in acceleration value stops the timer. A proper acceleration of 9.8 m/s^2 is expected. The second input is the individual timings of the participants which is given as a input. Then the results of various participants are compared to give a summarised data as the output. This summmarised data will present the user with his progress as per the different puzzles, timespan and

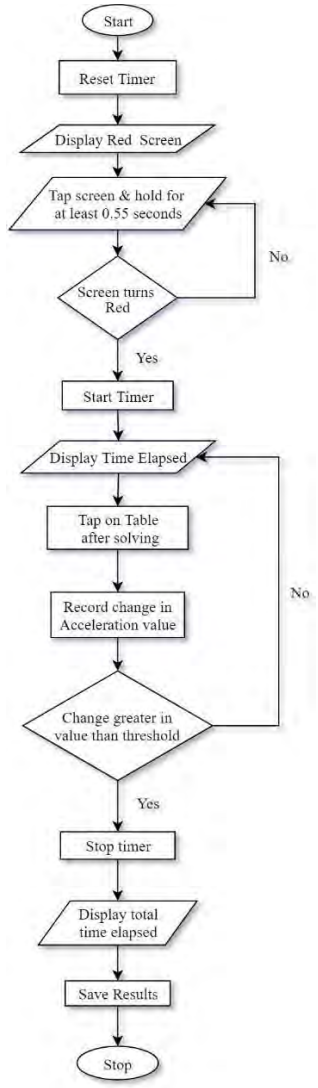


Figure 6: Flow Chart

The figure above shows the flow of the timer when the timer is reset it displays red screen then tap the screen and hold for atleast 0.55 seconds. When the screen turns red the timer is ready to start. It will start recording the time and then user has to solve the cube and tap on the desk after solving. Due to tapping change in acceleration value will be recorded. If the change in value is greater than the threshold then timer should stop and display the total time elapsed and save the results and stop the timer.

VI. Features of the Proposed System

1. Stopping timer using mobile phone sensors.
2. Adjustable threshold value for sensing the table tap.
3. Scrambles generator (using the WCA approved TNoodle program).
4. Keeping track of the timings of the last solves.

5. Personalised results page that displays the fastest and slowest timings; averages of the last 5, 10, and 100 solves in graphical format.
6. Support for different puzzles like the 3x3, 4x4, 5x5, Pyraminx, Megaminx, etc.

VII. Technology to be Used

1. Accelerometer in Smartphones

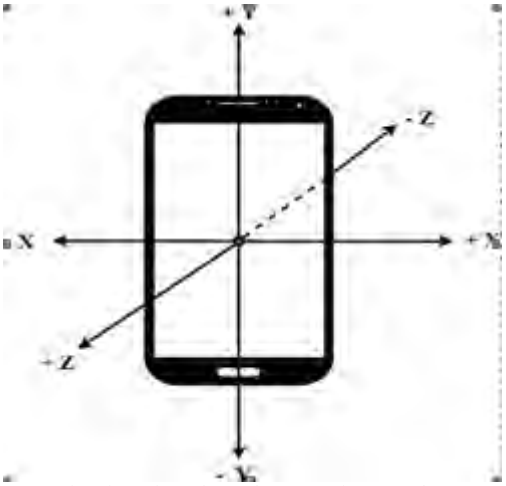


Figure 7: Axis orientation of Accelerometer in smartphones

To detect the sudden change in the acceleration value of the phone (usually when you strike the table on which the phone lies or to detect an unusual physical shock).

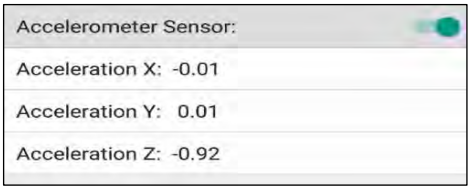


Figure 8: Accelerometer in action in a smartphone

These values were captured when the phone was lying flat on a table. In mobile devices, accelerometer is used as a sensor for measuring tilting motions and orientation changes of the mobile device. Accelerometers generate voltages as it experiences changes in acceleration the same phenomenon can be used to experience vibrations or sounds. It uses microscopic crystals that go under stress when vibrations occur, and from that stress a voltage is generated to create a reading on any acceleration.

2. Gyroscope in Smartphones

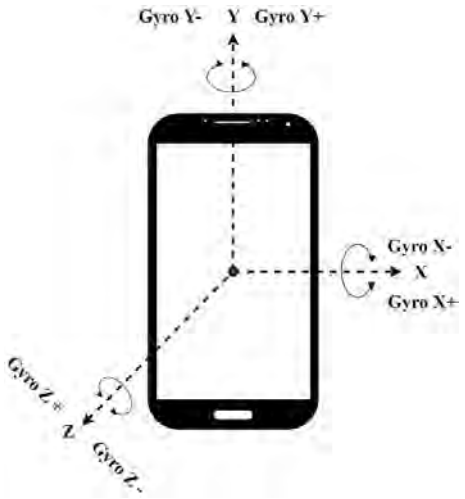


Figure 9: Axis orientation of Gyroscope in smartphones

Gyroscopes are used to keep a track of the rotational velocity values of a smartphone. This aids in determining the current orientation of the device which is vital for determining the threshold value of the accelerometer. If the phone is lying flat on its back, the acceleration threshold trigger value will be set with respect to the Z-axis. If the phone is resting on either of its side face, then the acceleration threshold trigger value will be set with respect to X-axis.

Conclusion

This system aims to give a simple and cost-effective alternative to the timers used in cubing. It also provides a wide range of possibilities to explore in the mobile application development domain. The system has the potential to revolutionize the way cubing as a sport is carried out and the way timings are recorded in cubing. The use of sensors in an application is a novel idea which can prove as a driving force for the cubing community where an individual can evaluate one's self which becomes possible due to this cost-effective system.

References

[1] World Cube Association 3x3x3 Cube Rankings webpage on WCA website. Available: https://www.worldcubeassociation.org/results/events.php?eventId=333&regionId=&years=&show=100%2BPersons&single=Single

[2] Mat for timing competition, R. W. Fox and J. L. Goers, (2002, Jun. 14). *US6940783B2.* Accessed on: Jan. 2, 2019. [Online]. Available: https://patents.google.com/patent/US7042806

[3] World Cube Association Cube for Cambodia webpage on WCA website. Available: https://www.worldcubeassociation.org/competitions/CubeforCambodia2018

[4] V. Kulkarni, Using Rubik's Cube for Developing the Aptitude of B-Schools Students, EXCEL International Journal of Multidisciplinary Management Studies ISSN 2249- 8834 EIJMMS, Vol.6 (4), APRIL (2016), pp. 10-17. [Online]. Available: zenithresearch.org.in

[5] Umek, A. & Kos, A. (2016) Validation of smartphone gyroscopes for mobile biofeedback applications. Pers Ubiquit Comput (2016) 20: 657. https://doi.org/10.1007/s00779-016-0946-4

[6] StackMat Pro Timer Official Webpage on Speed Stacks website. Available: https://www.speedstacks.com/store/retail/speed-stacks-stackmat-pro-timer/

[7] WCA Regulations webpage on WCA website. Available: https://www.worldcubeassociation.org/regulations/

[8] How a 15-year-old solved a Rubik's Cube in 5.25 seconds webpage on Vox's YouTube channel: Available: https://youtu.be/XwNUmnDu1r8

Biometric Identification and Real Time Control and Security System in Smart Car (BIRCS3)

Riya Parikh
Department of Information & Technology
Thakur College of Engineering & Technology
Mumbai, India
rparikh1996@gmail.com

Manasi Kadam
Department of Information & Technology
Thakur College of Engineering & Technology
Mumbai, India
Manasikadam6@gmail.com

Pooja Kammala
Department of Information & Technology
Thakur College of Engineering & Technology
Mumbai, India
kammalapooja@gmail.com

Neha Patwari
Assistant Professor
Department of Information & Technology
Thakur College of Engineering & Technology
Mumbai, India
neha.patwari83@gmail.com

Abstract
In the era of globalization the vehicle market is booming. There is rapid increase in the number of vehicles. In the last two decades number of vehicles running on road has increased in multiple of population. As the number of vehicle increases, the number of theft attempts has also increased all around the world, hence security and protection of the vehicles has gained a great importance nowadays. So to protect vehicles from the theft we have designed a security system for the vehicle and named it as Biometric Identification And Real Time Control and Security System in Smart Car (BIRCS3). This includes i) Authentication (Face Detection System), ii) prevention (Buzzer ringing and Ignition On and off) and iii) Android Application (Engage with vehicle). The system has Mobile Phone that is installed in the vehicle. The vehicle stuff stolen can be stopped by using GPS full-length of mobile phone & this information is used by owner of the vehicle for remoter processing. The owner sends a notification to mobile which is used in the vehicle that has stolen which in turn controls the engine of vehicle by locking it immediately. The engine can be unlocked by only by the sending the password to microcontroller by Authorized person only. The goal behind the design is to develop security for vehicle and Embedded system to communicate with engine of vehicle.

Keywords

Vehicle, Security, Owner, Theft, Identification, Android, PS

I.INTRODUCTION

A huge number of cars are lost every year in the state and a large number of automobiles are additionally recuperated by the Police from when they get the thief or not withstanding when the offenders leave the vehicles they have stolen after they have utilized them. The typical issue with the recouped vehicles achieving the real proprietors is that the vehicle require not be found in an indistinguishable ward from one in which the protest was propelled. Along these lines, when a vehicle is recouped, typically the Police attempt to follow out the real proprietor of the vehicle from the RTO in view of the permit and case number. In any case, this is a protracted and tedious process for the RTO to follow out the real proprietors from the records and advise back to the Police stations. In light of these deferrals, vehicles that are recouped all lengthy time-frame to really achieve their proprietors.

In spite of the different advancements that have been acquainted as of late with distinguish auto burglaries and following it, It was accounted for that the same number of as autos were stolen yearly on the planet. This venture comprises of an android based remote vehicle framework will give viable, ongoing vehicle area, mapping and detailing this data esteem and include by enhancing the level of administration gave. A vehicle following framework will advise where your vehicle is and where it has been, to what extent it has been. The framework utilizes geographic position and time data from the Global Positioning Satellites. The framework has an "On-Board Module" which dwells in the vehicle to be followed and a "Base Station" that screens information from the variegated vehicles [5].

II. LITERATURE SURVEY

Vehicle security [3][4] is unchangingly been an important priority in the automobile industry. Various techniques like inside locking system with watchtower were one of the security parameter, which could only protect versus thefts only when the

vehicle was stationery. However, to alimony in touch with a remote vehicle and track its other aspects like speed and location are stuff ripened and tested Today ‘s generation phones are not only capable of sending mails, making phone or video calls but moreover have the sufficiency to tenancy other smart phones. In this project we introduce a new efficient technique to dis-engage a remote vehicle using android technology. The parameters used for measuring the result was the squatter image placid from the users and stored them in database and then using these images as a way to demonstrate the users [1].

A.EXISTING SYSTEM

Vehicle security is always been an important priority in the automobile industry. Various techniques like central locking system with alarm were one of the security parameter, whichcould only protect against thefts only when the vehicle was stationery. Untracking Vehicle Tracking Unit has the worthiness to integrate the GPS tracking system with existing vehicle watchtower or provide watchtower features when someone is tampering with owner vehicle. It allows detecting the security threat surpassing the vehicle is driven yonder and gives the worthiness to track the vehicle over the internet. The existing system has, Car watchtower techniques are used to prevent the car theft with the help of variegated type of sensors like pressure, tilt and shock & door locking system, and GPS systems which provides a wholesale layout of geographical address. If Owner vehicle has been stolen then he/she must be file FIR to the police station.

B. DRAWBACKS OF EXISTING SYSTEM

The existing system does not provide any security to two wheelers vehicles. The forfeit and maintenance of the system is high. Owner must requite statement to the police and wait for whoopee that will be taken from the police. The GPS system does not provide street-wise address. Owner doesn‘t have knowledge well-nigh his vehicle until the police take action.

C. PROPOSED SYSTEM

The proposed system is Biometric Identification And Real Time Control and Security System in Smart Car (BIRCS3) a new and self-contained product that gives complement to the existing system. The product combines the installation of an electronic device in a vehicle, with purpose-designed android software (application) at least at one operational base to enable the owner or a third party to track the vehicle's location, collecting data in the process from the field and deliver it to the server using GPS technology for locating the vehicle and also lock/unlock the engine of the vehicle by sending a notification to the electronic device.

III. MOTIVATION

Face recognition has been a sought without problem of biometrics and it has a variety of applications in modern life. The problems of squatter recognition attract researchers working in biometrics, pattern-recognition held and computer vision. Several squatter recognition algorithms are moreover used in many variegated applications untied from biometrics, such as video compressions, indexing’s etc. They can moreover be used to classify multimedia content, to indulge fast and efficient searching for material that is of interest to the user. A squatter recognition system can be of unconfined help in forensic sciences, identification for law enforcement, surveillance, hallmark for financial and security system, and giving preferential wangle to authorized users i.e. wangle tenancy for secured areas etc. The problem of squatter recognition has gained plane increasingly importance without the recent increase in the terrorism related incidents. In today’s fast pacing world increasingly and increasingly people are moving into the smart phone technology. This is due to various applications in the smart phones that allows users to tenancy and automate stuff virtually them at the printing of the button, reducing the time they spend on the work they consider unproductive.

IV. METHODOLOGY

The proposed solution, Biometric Identification and Real time Control And Security System in Smart Car (Bircs3) is mainly used for securing car in case of theft. If the person who enter in the car is owner then the security system will be enabled that is the scanner will scan the image of a person because as he/she is an owner his image will be already stored in database. So the system will consider him/her as a authorized person then the car will start automatically.

If the person is relative, by using android application the owner will give access by locking or unlocking the car. Then scanner will scan the image of a relative and store that information in database. Whenever in future that person wants to use car he will be considered as authorized person by system. If the person is unknown (stranger/thief), the security system will be enabled i.e., the scanner will scan the image, if the scanned image is not matching with the images stored in the database the system will consider the person as unauthorized and then immediately the control system will enabled which will lock the doors and windows of the car and the alarm will ring and the scanned image of the unauthorized person will be send to the owner to his mobile app and to the police. Android application is for authorized person that should be installed in user mobile phone.

The person is authorized then following features will be provided by using android application. Such as a navigation application system where a user can search for a location that he or she wants to visit, messaging system where an application

will offer the colleague/friend contact number to send the estimated delay time due to traffic.

Hardware:

a) Power Supply
b) USB To TTL Type USBMOD3
c) Arduino Atmega328
d) Motor driver IC L293D
e) Dc motor
f) Buzzer

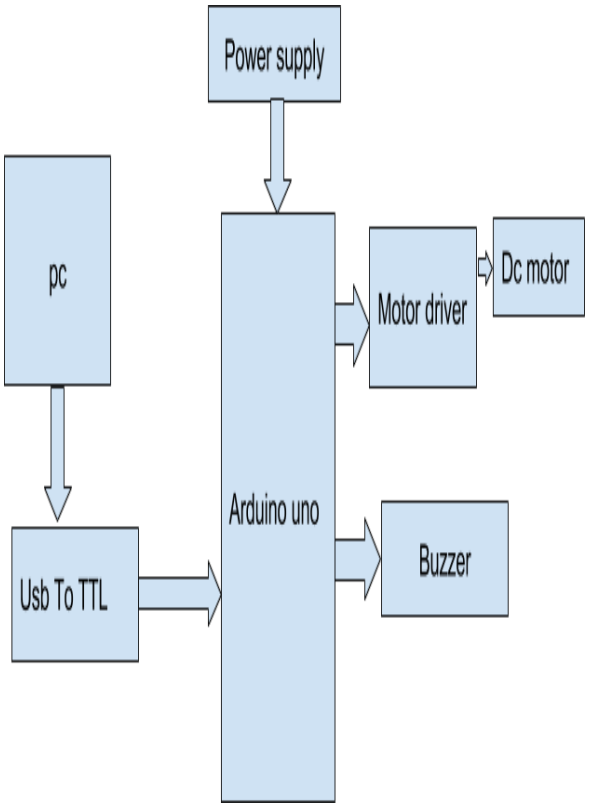


Figure no.1 Block diagram of hardware kit

The Hardware kit is connected in following steps:

1. The Arduino uno is connected to PC using USB To TTL.
2. The power supply is connected to Arduino uno.
3. The buzzer and Dc motor is connected to Arduino for authorization
4. The DC motor is connected to Motor Driver for direction i.e. to move in clockwise or anticlockwise.

Software:-

a) Android Studio
b) MATLAB

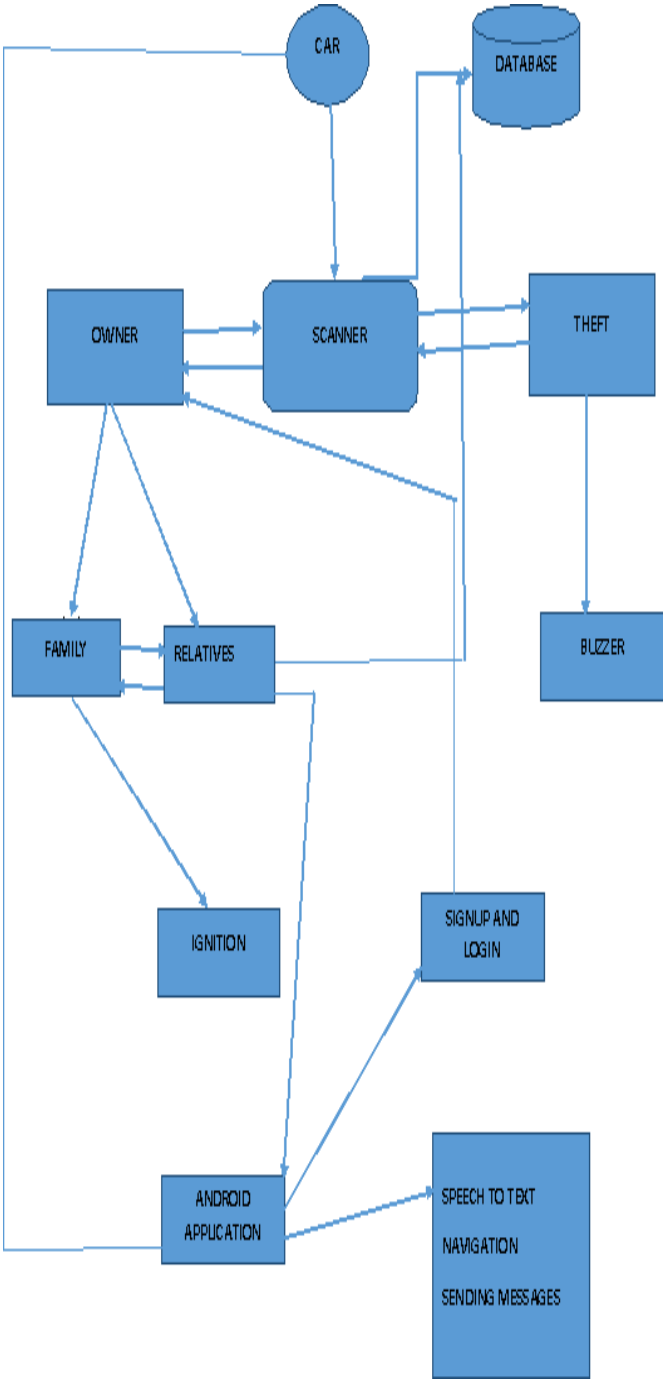


Figure no. 2 Flow chart for Biometric Identification And Real Time Control and SecuritySystem in Smart Car (BIRCS3)

a) GPS Technology

The Global Positioning System (GPS) is a satellite based navigation system consists of a network of 24 satellites located into orbit. The system provides essential information to military, starchy and commercial users virtually the world and

which is freely wieldy to anyone with a GPS receiver. works in any weather circumstances at anywhere in the world. Normally no subscription fees or system charges to utilize GPS. A GPS receiver must be locked on to the signal of at least three satellites to estimate 2D position (latitude and longitude) and track movement. With four or increasingly satellites in sight, the receiver can determine the user's 3Dposition (latitude, longitude and altitude). Once the vehicle position has been determined, the GPS unit can determine other information like, speed, loftiness to destination, time and other. GPS receiver is used for this research work to snift the vehicle location and provide information to responsible person.

b)ATMEGA328

The ATmega328is a single tweedle microcontroller created by Atmel and belongs to the megaAVR series. The Atmel 8-bit AVR RISC-based microcontroller combines 32 KB ISP wink memory with read-while-write capabilities, 1 KB EEPROM, 2 KB SRAM, 23 unstipulated purpose I/O lines, 32 unstipulated purpose working registers, three flexible timer/counters with compare modes, internal and external interrupts, serial programmable USART, a byte-oriented 2-wire serial interface, SPI serial port, 6-channel 10-bit A/D converter ,programmable watch dog timer with internal oscillator, and five software selectable power saving modes. The device operates between 1.8-5.5 volts. The device achieves throughputs unescapable 1 MIPS per MHz

c) Android

ANDROID (Automated Numeration of Data Realized by Optimized Image Detection) Android is an operating system for mobile devices such as Smartphone and tablet computers. It is ripened by the Open Handset Alliance led by Google. Android consists of a kernel based on the Linux kernel, with middleware, libraries and Computer Engineering and Intelligent Systems APIs written in C and using software running on an using framework which includes Java-compatible libraries based on Apache Harmony. Android uses the Dalvik virtual machine with just-in-time compilation to run Dalvikdex-code (Dalvik Executable), which is usually translated from Java byte code.

d) Matlab

MATLAB is a high-level language and interactive environment that enables you to perform computationally intensive tasks faster than with traditional programming languages such as C, C and FORTRAN. The Image Processing Toolbox is a hodgepodge of functions that proffer the capabilities of the MATLAB's numeric computing environment. The toolbox supports a wide range of image processing operations, including: Geometric operations, Neighborhood and woodcut operations, linear filtering and filter design, Transforms, Image wringer and enhancement, Binary image operations – Region of interest operations

V. RESULT



Figure no.3 Image processing

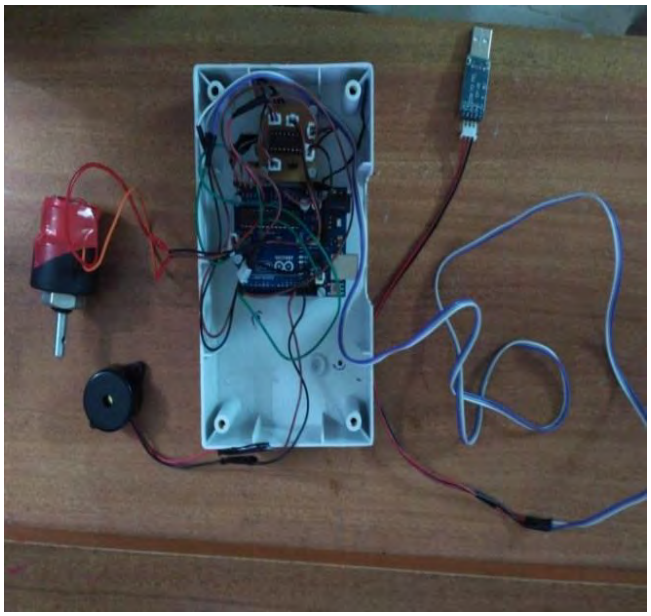


Figure no.4 Hardware kit

The Figure no:3 shows that If the image in the trained database is matched with the image in the test database then the ignition will start and if the image doesn't match then the ignition will stop and buzzer will ring. The Figure no:4 represent the hardware kit where the Arduino Atmega328 is connected to the computer using the USB To TTL Type USBMOD3 and programming has been implemented in the matlab.

VI. .CONCLUSION

Thus the low cost proposed system acts as an anti theft and stolen vehicle recovery system which can be used to track a vehicle by simply sending a notification to the proposed embedded device in the vehicle. The android app both in case of personal as well as business purpose improves safety, security, communication medium, performance monitoring and recovery of vehicle . Upgrading this setup is very easy which makes it open to future requirements without the need of rebuilding everything from scratch, which also makes it more efficient.

VI .ACKNOWLEDGMENT

The people we would to thanks are many for their help in presenting this project. We specially want to thanks our Principal Dr.B.K.Mishra, our HOD-IT Dr. Rajesh Bansode and our project guide Mrs. Neha Patwari whose support have inspired and shaped much of the content of this project moreover guiding us on this project work and spending hours working to modernize the content of this project. Moreover we want to thanks to all the staff members who help us to improve.

VII. REFERENCES

[1] i. Jian Xiao, Haidong Feng; ―A Low-cost Extendable Framework for Embedded Smart Car Security System‖, *Proceedings of the 2009 IEEE International Conference on Networking, Sensing and Control,* March 26, 2009, pp: 2320–088X.

[2] ii. Raquib B., Soumyajit R., Parthib M., Subhajit B., Abhishek M., Sauvik D., ―Implementation of MATLAB based object detection technique on Arduino Board and iROBOT CREATE‖, *IJSRP*, Vol. 4, Issue 1, Jan 2014.

[3] Joseph A. O'Sullivan, Robert P., ―Advances in Security Technologies: Imaging, Anomaly Detection, and Target and Biometric Recognition‖, *Microwave Symposium IEEE/MTT-S International Volume,* 2012.

[4] Jayashri. M. Ahire, Swapnali.S.Baste, Krutika. S. Bhadke , Pratibha. G. Shelake -*Android App for Stolen Vehicle Tracking and Engine-Disengaging System*IJRITCC | February 2015, Available*http://www.ijritcc.org*

[5] J.-P. Hubaux, S. C Apkun, and J. Luo, ―The security and privacy of smart vehicles,‖ IEEE Security & Privacy Magazine, vol. 2, no. 3, pp. 49–55.

[6] W. Stallings, ―Cryptography and Network Security‖, Prentice-Hall, Englewood Cliffs, NJ, USA, 4th edition

[7] http://www.androidauthority.com/best-car- apps-android

Security in Hadoop

Omkar Shinde, Sweta Mishra
Vistaar Mumbai, Barclays Pune, India
omkarshinde254@gmail.com, sweta.dl07@gmail.com

***Abstract*— Hadoop is an excellent distributed processing framework that ace's in data processing and big data storage by using the power of clustered system. The main aim of this paper is to enhance Hadoop capabilities by providing a layer security called EA^3 (Encryption, Authentication, Authorization, Auditing). With an additional layer of security, Hadoop services can now be more easily adopted thus providing more flexibility, scalability, availability and performance. Along with EA^3 this paper also contains various security protocols like OAuth that can be implemented with Hadoop. This paper will discuss on implementation of these ideas, give an approach on problems and challenges that we came across and some solution. It is a blend of many exquisite technologies integrated together in a marvellous way to achieve capabilities of Hadoop without risking anything.**

Keywords—Hadoop Security; OAuth; Authentication; Authorization, Auditing, Encryption.

I. INTRODUCTION

Today is an era of technology, in this fast-paced world computational power is everything. Almost in all fields such as Medical, Artificial Intelligence, Constructions, Aerospace Engineering computers are the main source of productivity. But due to lack of resource and technologies it is very hard to satisfy all those needs without sacrificing our pockets. This is where Hadoop comes into picture. Hadoop provides high computational power that at low cost that makes it more preferable and easiest source of computational power, with the power of cluster of computer Hadoop can divide any task into small chucks and provides result in least possible time.

However, the problem here lies in reliability. As Hadoop is developed on a network of computers, it becomes an easy target for attackers. The reliability of Hadoop is very poor as its dependents on multiple sources thus having multiple points of failure.

With recent advancement in Security, today we can introduce various protocols on a distributed network in such a way that it can prevent itself without relying on third party entities and also providing high level of performance without degrading the performance

II. LITERATURE REVIEW

This paper focuses on providing security to Hadoop system using a node-based security protocol called as EA^3. With help of EA^3 we can improve the security of Hadoop system without affecting the performance. Also, this algorithm can increase the efficiency of Hadoop as it requires less time to process rather than Kerberos which is currently used as a Security method to provide security to Hadoop [5]. It will also eliminate the risk of trusting a third party as a key entity of security.

The main 4 phases here are Authentication which will check if the user is valid or not, Authorization which will only allow the user to perform those actions that they are allowed to perform, Auditing will keep track of all actions of the user and encryption will encrypt the data so that there won't be any risk even if an unauthorized user get access to data [2].

The main topic of this research totally focuses recent development in Security and numerous ways, making it feasible to assists Hadoop develop into a better and secured system. With high computation power, high storage capabilities and strong security protocols Hadoop can be the most demanding system in today's era and can satisfy all the needs of technology [4].

This paper will also focus on various security algorithms like OAuth which were introduced later after incorporating Kerberos in Hadoop, OAuth is a very powerful security algorithm used in almost all the smartphones app's and can be very efficient and powerful and powerful when used as a security medium for Hadoop [5].

III. SOLUTION APPROACH

A. Kerberos

Kerberos is a network authentication protocol designed to provide security and authentication for client and server applications by using secret-key cryptography terminology.

The Kerberos protocol uses encryption and decryption procedures so that a client can prove its identity to a server (and vice versa) over internet (Insecure Network). After a client and server has used Kerberos to establish a connection and prove their identity, they can encrypt their communication to ensure privacy and data integrity.

This will help them make sure that other people can't eavesdrop the communication, and even if they did, data confidentiality is still maintained due to encryption

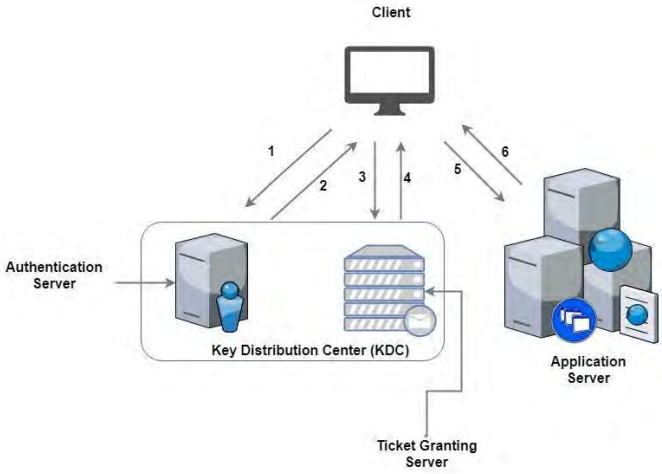


Fig 1. Kerberos Flow Chart

There are the main operations that occur in Kerberos security algorithm

First Step is client will send an authentication request to the Authentication Server.

The Authentication Server will check if the client is an authenticated user to have server resource. If the client is authenticated then the Authentication Server will respond with a Ticket Granting Ticket (TGT).

After receiving the TGT, Suppose clients want to perform some operations like [ls / :- Listing the contents of the root directory]. Then the client will us the TGT and go to Key Distribution Centre (KDC) or Ticket Granting Server (TGS) and request for accessing a server with its name node.

- The TGS will reply with a service ticket for that specific name node and the client will catch the service ticket.
- Using this Service ticket, the client will communicate with Server so that server can know the client is authenticated.
- The server will send a response to client, so that client can know that it is talking to an authenticated entity.

In this way a client can communicate to a Server (name node) on a Hadoop network using Kerb ores as its security protocol

Disadvantages of Kerberos: -

- Kerberos has a single point of failure i.e. if the KDC becomes unavailable then the entire system will shut down
- If an attacker compromises the KDC, then authentication information of every client and server on the network would be revealed.
- A minimum 3 number of transactions are required to perform even a small operation like listing directory [ls /].
- Requires third party trusted server KGS [Key Generation Server]

B. EA^3 Algorithm

EA^3 is a simple algorithm that uses blockchain technology to provide security in a distributed Hadoop network.

There are 4 basic operation that all client and server do, like Authentication, Authorization, Auditing, And Encryption

a) Authentication: - Whenever a client wants to access a resource on server the client has to prove itself, this is accomplished by authentication.

Authentication Process is further enhanced by using hashing and salting techniques.

- The password from user is altered with help of salting to increase the security
- The output of salting is then passed through hashing algorithm. Salting prevents leakage of password if rainbow tables are used for cracking hash
- The final output is then encrypted and stored in Database, thus adding a three-level security.

a) Authorization:

Authorization is a basic concept where the client is allowed to perform only those actions that they are allowed to. In this way an unauthorized user cannot access the files that He/she is not meant to use

c) Auditing:

Each action performed by client is audited, so that all server can keep a track of all user operations, it is an important step of blockchain algorithm to keep all the server in sync and prevent single point of failure. By single point of failure, we mean, if all one of the server crashes or its security gets compromised then rest of servers can still be able to run the system.

d) Encryption:

The data in server is encrypted with Server's encryption key, this data can only be accessed by those people who are authenticated at first step. After authentication the client will receive a Service ticket which would be used to decrypt the server's data.

As we can see this algorithm use very basic step, so how can it overcome Kerberos? Well Kerberos is a network authentication protocol so it provides security over a network, this make the system vulnerable to a single point of failure. If the network goes down, or the third-party Kerberos server's security is compromised then the clients and servers connected to it also gets affected by its consequences like security breaches, system down time, information leakage.

EA^3 on the other hand provides node level security, this prevents single point of failure but consumes more processing power from the node.

Each server is connected to rest other servers like a node in blockchain. All the actions of a server are audited and synchronized with other servers to maintain log. After getting authenticated successfully, the client can perform any number of operations with recursively raising the ticket.

A simple operation like [ls /] Listing the contents of root directory can be done easily if user is authenticated and has permission to perform it. Comparing this with Kerberos, only one step is needed to perform that is authentication thus increasing the performance and reducing the wait time.

C. OAuth

O Auth is also one more Authorization that can be used to provide security in Hadoop.

While introduction Kerberos to Hadoop, the reason why OAuth was not used is because, it wasn't invented at that time. O Auth could be a good alternative algorithm that can be used to protect distributed networks.

OAuth (Open Authorization) is an open standardized security algorithm that uses token-based authentication and authorization to provide security on Internet.

Thus, algorithm is commonly in android apps to provide security. The main feature of this app is that it identifies the authenticated user once and then uses a token to get authorized contents and connect to Service API's instead of prompting the user for credential again and again.

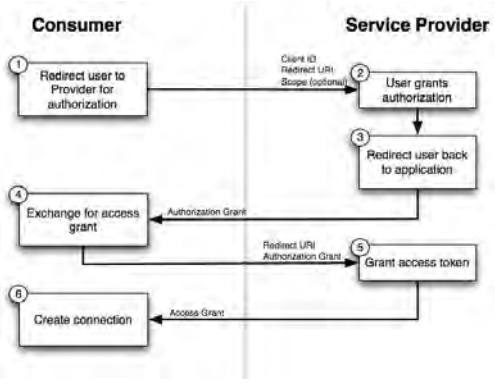


Fig 2. OAuth Authentication

IV. RESULTS AND DISCUSSION

As you can see, there is a back-and-forth conversation that takes place between the application and the service provider to grant access of providers account to the application. This exchange, commonly known as the "OAuth Dance"

Following are the steps that takes place in Open Authentication:

1) Everything starts when user opens the app. The application redirects the user to a URL where he/she has to provide to prove its identity. This URL is generally the Service API that user want to access. Here the Service provider asks the

user if he or she want to grant access to application to read, updated their data.

2) It's up to user to grant the request or not, if the request is granted then Step-3 is performed, if not then the application will not proceed further until request is granted.

3) The Service provider will redirect the user back to the application. Generally using a redirect URL. The provider also passes an authorization code as a parameter that is used in later stages

4) The application uses the authorization code for an access grant request.

5) After getting the grand request, the service provider grants access to the application. This grant includes an access token and a refresh token. The application receives these tokens

6) The application uses the Access Grant token to establish a connection between local application user's account and external Service API. With the connection established, the application now gets the authority to refer the Service API and invoke the provider and other procedures on behalf of the user. This makes the application a trusted medium that is recognized by user and can be used further to communicate with the server without any intervention to the user. [5]

V. CONCLUSION

Thus, EA^3 can be a better algorithm to provide security in Hadoop as it is work on node bases instead of overall network. It also doesn't rely on a third-party server thus preventing it from trust issues. It also requires minimum number of transactions than Kerberos which makes it faster thus improving the performance of Hadoop system.

With a 3-layer authentication process, it becomes impossible for the attacker to get user credentials in the case where data is compromised.

Authorization at every step ensures confidentiality and authenticity of data, this also prevents unauthorized users to perform unnecessary action on the system.
Auditing keeps logs of each and every activity of user at all the nodes, this help in recovering and identifying the attacker in case of an attack
Encryption of data makes sure that even when data is accessed by other user, its confidentiality is still maintained

EA^3 is a pretty simple yet powerful algorithm that provides node level security to Distributed Hadoop Network. This algorithm is powerful and tremendous and can be easily implemented in a distributed network system. Although each node would be required to have enough processing power to perform necessary action. A further deep-down research is required to increase its capabilities. No doubt this algorithm can be improved in lots of ways according to the needs and this is the reason that makes it flexible and adaptable.

ACKNOWLEDGMENT

The Authors would like to thanks our friend and family for their support. The constant dedication of our members and enthusiasm helped us to move forward and encouraged us to write this paper. Last but not least the Authors would thank all the author of the IEEE paper that we have referred and appreciate their efforts for showing such marvelous piece of work.

REFERENCES

[1] Konstantin S., Hairon Kuang,"The Hadoop Distributed File System",
IEEE Symposium on Mass Storage Systems and Technologies (MSST), Jun 2010 pp.1–10.

[2] Bao Rong Chang ; Hsiu Fen Tsai ; Zih-Yao Lin ; Chi-Ming Chen,"Access Security on Cloud Computing Implemented in Hadoop System", 2011 Fifth International Conference on Genetic and Evolutionary Computing Year: 2011, pp. 77-80.

[3] Johnu G., Chien-An Chen, "Hadoop MapReduce for Mobile Clouds,"
IEEE Transactions on Cloud Computing, Vol.3, No.1, Jan 2014, pp.1-14.

[4] S. Huchton, G. Xie, and R. Beverly,"Building and Evaluting a k-resilient mobile distributed file system resistant to device compromise," in Proc. MILCOM, 2011

[5] https://blog.cloudera.com/blog/2012/03/authorization-and-authentication-in-hadoop/

[6] Apache Hadoop. http://hadoop.apache.org/

Event Management and Analysis System for College Fests

Prof. Mariyam Jawadwala, Priyank Patel, Hinal Patel, Priyanka Patil
Dept. of Information Technology Vidyavardhini's College of Engineering and Technology University of Mumbai, India
mariyam.jawadwala@vcet.edu.in, prp.priyank@gmail.com, hinup2123@gmail.com, patilpriyanka982@gmail.com

***Abstract*—This document is a model for Event Management system which will help in reducing paper work and implement processes like registration, login, payment gateway, poll system, chat room for the events over Digital platform using android.**

Keywords—Android, Event Management, Firebase, Android studio, Java.

I. INTRODUCTION

The objective of this project is to develop an android application which provides information about events. Manual work has to be reduced which will be done with the help of this application. Users will be able to manage their event participation, such as reserving their seats in events, registering at the event site, and so on. The organizing committee can solve participants query using chat room feature. This application uses QR code to provide an easy way to verify participants' identity in an event. Only event organizers and volunteers can upload event content to the site

II. PROBLEM STATEMENT

In the current scenario when large number of population is using smart phones in their day to day life in order to speed up working, increase efficiency, reduce efforts and paper work we are still working in the old fashioned way doing publicity, registering participants over a piece of paper, maintaining all records filing up the data. If any kind of analysis needs to be done like checking which event was successful or which event was a failure the organization team needs to talk to each and every events head collect the data manually and then come to a conclusion. Similarly if any event needs to be brought, looking for participants interests gets difficult using such old fashioned methodologies.

III. SERVEY

In order to work with the events taking place in our colleges we use web applications to advertise the events get registration and provide information about the same through e-mails and calls. There do exist some mobile application to provide event details and registration but when it comes to analysis of the data that has been collected over the time there exists no such system.

IV. IV. PROPOSED SOLUTION

As we already are into in an era of automation it is easier to look for a solution which will be easy to use by everyone and easy to handle by the participant and the organization team. An android application for the event management system would be the best solution looking at the growth in demand of smart phones and android applications which will help overcome issue's such as change in schedule, venue and Display of winner's list.
Similar overcoming problems like coordinators need to make call to each participant and tell them about the changes in schedule, event has started, venue of event has changed, or declaration of winners. During an ongoing event records need to be maintained about participants who have qualified for next rounds, for this purpose traditionally they are asked for their name which is recorded on a new piece of paper and again so on till winner is found.

A. Registration and Log-in

This module will do Basic registration which has further fields such as user name, e-mail address, mobile number, college name and a submit button. Further log-in page will have email address and password to validate the user using the online database i.e. firebase for authentication. In this module other features like password regeneration, validation for lost password will also be used. The registration forms will be specific to the events user is participating, on the basis of which the user will be given payment and Qr options.

B. Qr code

Qr code implantation is for the participants who have registered and made a payment through Paytm and Qr code is basically a validation and a proof that the user has successfully registered paid and participated. This Qr code generated in the application can be used at the time of participation by the user to authorize themselves. The Event Managers or the Event Heads may posses the authority to validate the user based on the administration rules. Once validated the user is said to be a eligible to participate in the event.

C. Payment gateway

Payment Gateway is an important part in this project as the events are not free to registered. Every event has different price to be payed but the account to which amount has to be transfered will be same. By implementing Paytm as the payment system we stay flexible with different accounts easily for different fests. The Qr code and the number ill be provided to the user to pay to the said user's Paytm account.
The transaction id will be then reverted back in order to validate the transaction and if the said amount of data is transferred. Once the payment is done the user gets the Qr code as a proof of payment and registration. Some users might not be using an such resource for payment. So for such users an option for on-day payment is also made available which gives them the flexibility to pay cash.

D. Poll module

This module is implemented to understand the demands of users and work on future development I those sectors to bring up new ideas and increase the popularity of the overall fests. An activity will be used in which all the events will be listed and the users and give their response a if they liked or disliked the event and on the basis of that the administrator will understand the response of users.

The users can take poll for all the events weather participated or not. As they are valuable audience in the events as well. The polls will then be measured with a threshold set by the organization team to understand which events are dragging them down and which events are responsible for successful events.

E. Chat room module

Chatroom is a very revolutionary tool in this project. As we know there cannot always be a person to coordinate with every user. So instead user can simply add his/her query onto the chat room and the users can help each other as well as the organizers will be willing to help them with their queries.
Sometimes the event venues change due to certain mishaps or misguided time scheduling. The organization team will notify the users using the chatroom, it helps in decreasing the communication gap between the participants and the technical and cultural committees.

F. Locations module

This module is like a guide book for participants visiting from different colleges as they are unaware of the college campus, they are in. This module provides college location as well as a 2D college map showing all the floors the rooms with their room number, the user can easily go through it and locate itself in the campus and find its way to the event location.

G. Notification

Events are not always expected to be on time, so instead of continuously waiting for the event to start at the location we can rather wait for it anywhere else or complete our task we have on pending as soon as the event is live a notification shall pop up saying that this event is live please assemble at this particular location, and so the time is saved and unwanted crowd is avoided.

V. DIAGRAMS

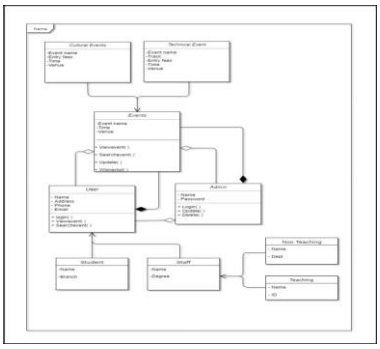


Fig 1. Use Case Diagram



Fig 2. Class Diagram



Fig 3. Screenshot of navigation activity

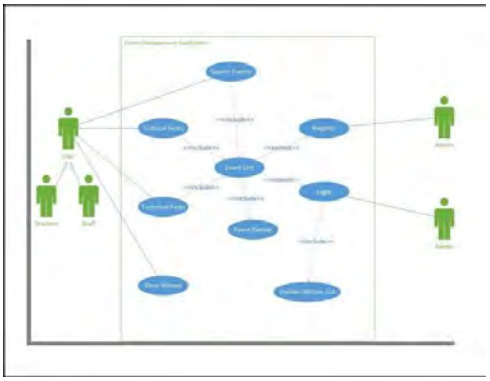


Fig 4. Screenshot of Splash screen

VI. ALGORITHMS

In order to do analysis we need to have a proper algorithm to do the thinking part. Apriori algorithm is an algorithm which would help in finding out frequent patterns and will in the process also provide results like which is the most popular events participants prefer and what are events might be more popular if brought together.
Apriori is a mining algorithm for frequent data set mining. It uses bottom up approach which means it starts with single values moving on with multiple values

Confidence(A− > B) = SuppCount(AUB)/SuppCount(A)
(1)

The Apriori Algorithm is an influential algorithm for mining frequent item sets for Boolean association rules.

Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time

VII. SCOPE

The Event Management Application can reduce the manual paper work of organizing committee includes registration, maintaining records of participants, transaction, display winners list. Considering the live updates of events, participants will come to know of the same using chat room feature. QR code will be scanned at the time of event to cross check information about participant.

Chat room and poll feature will be the highlight as this platform will help participants to contact committee members and organizing committee will be having results of event being successful or not on the basis of polls given. Overall this project will help the organizing committee to do their work putting less efforts.

VIII. CONCLUSION

Basically to help the organizing committee to reduce human efforts. The aim of this paper is to introduce platform to participants for registering, giving feedback for event, transaction purpose, solving queries using chat box and also getting live updates about the event.

This application will solve the problems faced in human interactions related to money, improper paper work. The application helps in improvising and the use of technologies for the benefit of the organisation by easily publicising the events reaching out to more user then ever and making the events more and more successful and better than ever.

REFERENCES

[1] M.Mahalakshmi, S.Gomathi and S.Krithika, ”Event Management System”, Coimbatore: International Journal of Trend in Research and Development, 2016.,

[2] J. Nagarathina and Dr.Y.Kalpana, INTERCOMMUNICATION INFORMATION SYSTEM USING ANDROID, International journal of current engineering and scientific research (IJCESR), Chennai, 2017.,

[3] S. Misal, S. Jadhav, T. Jore and A. Ugale, ”Review on College Event Organizer”, Nashik: International Research Journal of Engineering and Technology (IRJET), 2017.

The Alternate Method of Agriculture: Aquaponics Using IOT

Prof. Maryam. Jawadwala
Information Technology
Mumbai University
VCET
Mumbai, India
mariyam.jawadwala@vcet.edu.in

Prof. Yogesh. Pingle
Information Technology
Mumbai University
VCET
Mumbai, India
yogesh.pingle@vcet.edu.in

Soumya. Kulal
Information Technology
Mumbai University
VCET
Mumbai, India
kulalsoumya851@gmail.com

Bhumika Mate
Information Technology
Mumbai University
VCET
Mumbai,India
bhumikamate90@gmail.com

Prof. Yogesh. Pingle
Information Technology
Mumbai University
VCET
Mumbai, India
yogesh.pingle@vcet.edu.in

Soumya. Kulal
Information Technology
Mumbai University
VCET
Mumbai, India
kulalsoumya851@gmail.com

Manasi. Pilankar
Information Technology
Mumbai University
VCET
Mumbai, India
manasipilankar29@gmail.com

***Abstract*— Agriculture provides food for most of the population in the whole world. Also the agriculture has many problems related to soil fertility, soil erosion, maintaining the climatic conditions for the crops and so on. To overcome all these problems related to soil and growth of natural food we have this Aquaphonics system which will provide natural food for the country. It will provide fish as well as plants too for humans. Aquaponics is a mixture of aquaculture and hydrophonics.In aquaculture there is growth of fish and in hydrophonics we have growth of soiless plants.Hence combining the two methods and making use of IOT technology we have designed a system in which soiless plants are grown with the help of fish extract which will be fed as a good and natural manure for plants. Also the user will just make use of the designed system and nodeMCU and thingspeak cloud due to which the pH and temperature conditions of water will be sent on an app called pocketIOT by connecting it with thingspeak cloud.Indoor aquaponics system will help the farmers to depend on both at a same time and also it is a soiless plantation so it naturally doesn't incur much cost. Indoor aquaponics system will help to save time and money in future as the fishes are nurtured at home and plants are also grown simultaneously. This will also reduce the loan and sucide rates of the farmers across the world. The economic condition will be better in a country if this is adapted by every homes. Naturally the GDP of the country will also improve in a good ratio and will also help many of them for natural food and in monetary funds.**

Keywords—Aquaponics, IOT, Soiless plantation

I. INTRODUCTION

Agriculture has provided the most of the population with natural food but nowadays the use of pesticides has been increased for fast growth of the crops because of which it has caused damage to human health and the problems related to soil also increased in the past few years. Because of which we have introduced aquaponics system for the growth of the plants naturally with the help of fish extract. Also most of the people prefer fish in their diet so its beneficial for them too as growth of fish also takes place in this system with proper and natural care.Years before everything was natural so to bring up the same natural food we have introduced this aquaponics system that resolves soil and natural food issues together.

II. EXISTING SYSTEM

In todays time the world has faced many health issues because of unavailability of natural food. Also the artificial fertilizers and chemicals are used for the fast growth of plants due to which the soil loses its original nature and it takes years to regain that back. Farmers have no alternative than to grow plants artificially with the help of chemicals and fertilizers available.

III. PROBLEM STATEMENT

Due to artificial use of fertilizers and chemicals the soil has lost its originality. The farmers have no idea or either forced to use this for growing crops. It has become difficult too for farmers too to waive off loans with the help of crops as they are dependant on nature of soil and climatic conditions. Always this has created problem so it is necessary to find out solution for this.

III. LIMITATIONS

The system is being developed humanly with the help of various components. It is possible that some components may fail to give the accurate result over the period of time so it is necessary that when such failure of components fail to work then it needs to be replaced. Also the feeding of fish needs to be on time which is not automated so it will consume small amount of time for feeding fish.

IV.SCOPE

The scope of the project is in the places where soil erosion and fertility of the soil is the main problem .Also the monetary issues of the farmers get resolved. In this way we acquire natural food by this system, also the use of pesticides and other chemicals are avoided so anyway this also keeps a proper health among the people. So in the end we acquire natural food, good health and extra source of income with the help of this system.

V. REVIEW OF LITERATURE

The aquaponics system[1] makes use of various components like pH sensor, temperature sensor , grow bed and nodeMCU. The conditions of water like the pH and temperature with the help of sensors is being sent on pocketIOT app with the help of thingspeak and nodeMCU. This system is going to take fish extract as a nutrient for plant with the help of tubes and water. There is an aerator to pull out the extract with the help of water.

VI. SOLUTION OF APPROACH *A.ALGORITHM:*

- Collect the information of the aquaponics system from various available sources.
- Study of all hardware components and software requirements for project from IOT point of view.
- Finding the estimated cost of components for the project.
- Designing a basic outline for implementation purpose.
- Preparing an aquaponics system:
    - Gathering of all hardware components.
    - Establish model of Aquaponics system.
    - Interfacing of sensors, components (relay, led light, air pump, water pump, grow light etc.) with microcontroller.
    - Testing of connection.
    - Achieving Readings and data.
    - Processing it to finding out trends.

B.BLOCK DIAGRAM

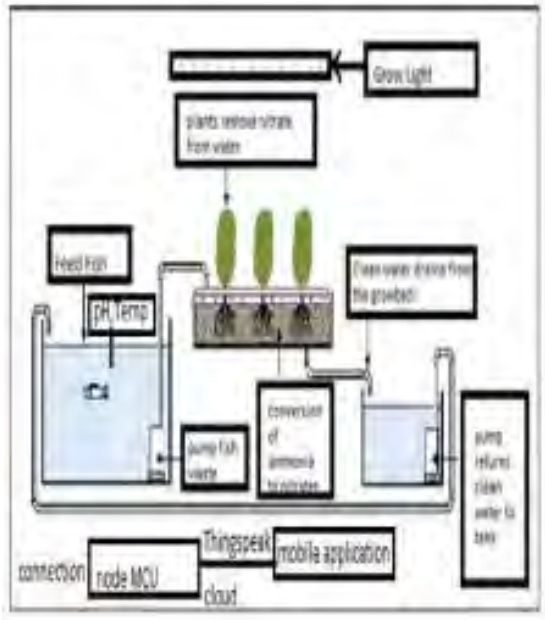


Fig.1. Working of AQUAPONICS

C. IMPLEMENTATION

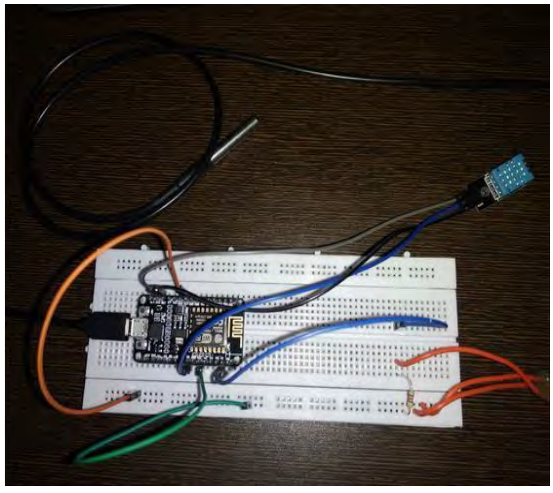


FIG.2 WORKING OF AQUAPONICS

VII.RESULTS AND DISCUSSION

The result of this aquaponics system is that we get the fish and plants together naturally without the use of any artificial fertilizers and chemicals. The system has been developed considering the pH and temperature conditions of the water in the tank and also the amount of extract of needed to grow plant will be in some great

amount . Also ultimately at last the system is been developed to solve the problem of farmers who grow on soil with the use of chemicals and fertilizers which degrade the soil.

VIII. CONCLUSION AND FUTURE SCOPE

The conclusion of the project is that it is developed to avoid the use of artificial fertilizers and chemicals that is used to grow the crops. The farmers who face this issue of growing natural food will get solved by the use of this system. The fish extract used here is natural nutrient for the growth of the plants. In future the farmer will make profit out of this and also the loan issues might get resolved to certain extent.

IX. References

[1]Wanda Vernandhes, N.S Salahuddin, "Smart Aquaponic with Monitoring and Control System Based On IOT", In IEEE, pp. 2017.,

[2]Analene Montesines Nagayo, Caesar Mendoza, Rodrigo S. Jamisola, "An Automated Solar-Powered Aquaponics System towards Agricultural Sustainability in the Sultanate Oman" ,In IEEE, pp. 2017.,

[3] N Hari Kumar ,Sandhya Baskaran ,Sanjana Hariraj , Vaishali Krishnan, "An Autonomous Aquaponics System using 6LoWPAN based WSN" In IEEE,pp. 4,2016 .,

[4] Aquaponics: retrieved from http://en.wikipedia.org/wikiiAquaponics

[5]Megumi U. Leatherbury "Vegilab and Aquaponics Indoor Growing System " In IEEE ,pp.2014.,

[6] M.F. Saaid, N. S. M. Fadhil, M.S.A. Megat Ali, M.Z.H. Noor "Automated Indoor Aquaponic Cultivation Technique" In IEEE, pp.2013.,

[7]M.N. Mamatha1 and S.N. Namratha2 " Design and implementation of Indoor Farming using Automated Aquaponics System" In IEEE ,pp.2017.,

Svara – The Music Therapy App

Swapnil H. Mistry, Ashok R. Patel, Sridhar Subramanian, Prof. Yogesh P Pingle

Department of Information Technology
Mumbai, India

***Abstract* — Stress on human behavior has been widely recognized. Stress must be there in human life at certain level to go ahead and take challenges in life and solve it. But sometimes if stress exceeds the certain level, then it causes negative impact in human body. Due to stress, human gets chronic diseases such as headache, insomnia, hypertension, high BP, diabetic and so on. There is no stress test in pathology. To reduce the stress, people do Yoga therapy. They do certain meditation. They do listen music which they like. This paper introduces Music Therapy which is based on Indian classical music. There are majorly seven "Swaras" in Indian classical music. Depending on these seven and five other swaras, thousands of ragas have been developed by many people. These ragas have impact on time as well as the body of human being. As people listens to these ragas which have certain day and night timings, the happiest molecules are being flowed during the listening time. Thus we have come up with an app which run on both the platforms IOS and Android systems. This project generates the report of the user for further usage and benefits.**

I. INTRODUCTION

There is vast increase in stress level with growing society. The young generation lay with stress and the old generation people suffering from mental disorder of dementia and Alzheimer's disease. The Music Therapy act as one of the substitute medication to stress, anxiety, dementia and Alzheimer's disease. The Researcher have found that the Music Therapy act as a stress buster. The important need for this music therapy is, because of the increase in the consumption of antibiotic by people to reduce stress. Whenever the person is in tension or stressed, the heart palpitates at a faster rate at that moment. The music at that particular time will rejuvenate the mood and will reduce the heart beat rate. The person will relaxed after listening to the music. The music can give immense pleasure to the person by enhance songs played of different RAGA, which suits that person according to the time and disease. There is no age bound for this therapy. So our project would use music therapy to reduce stress and also measure the effect of music on the person on a real time basis. These sensor would sense the heart rate and according simulate the RAGA. The scope of the project is wide.

The Scope of the project is wide as it will serve all the people of all ages according to their need and wish. As the App will serve as remedy and an assistant to most of the patients.

II. PROBLEM STATEMENT

This project solves the purposes on studying the effect of music therapy on the anxiety, dementia and Alzheimer's disease. The recorded music will be played and the effects of the music is been analyzed. There are many music therapy sites which plays music without the reference to the time and the effect on human. The music therapy app will help the patients and also the old aged people to overcome their difficulties. Even by this app the work load of the assistance of the patients is reduced. The feedback will be taken and the contacts of nearby therapist is been provided.

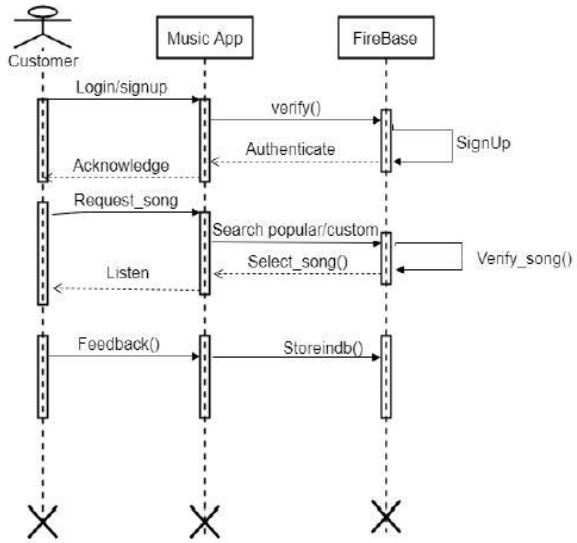
III. METHODOLOGY

We want to serve people who undergoing such diseases. This motivated us to create this Music Therapy App came from an anime named Svara. There is need of assistants for all the dementia and Alzheimer's diseased patients. This app will help them as an assistant. The Music Therapy App act as the substitute medication to stress, anxiety, dementia and Alzheimer's disease. Researchers have proved that music has shown its effect over 21 patients.

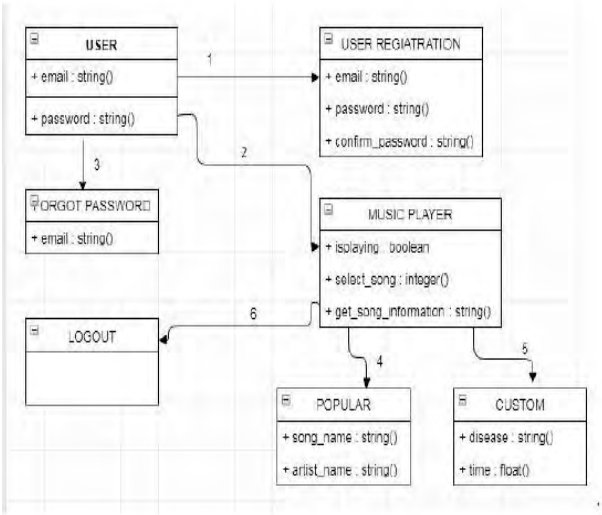
IV. DISEASES ANS IT'S RAGAS

RAGA	DISEASE
Ahir Bhairav	Indigestion,Rheumatic Arthritis,Hypertension
Asavari	To build confidence- Low BP
Bageshri	Insomnia
Bhairavi	Rheumatic Arthritis sinusitis encourages detachment
Bhim palas	Anxiety , Hypertension
Chandrakauns	Anorexia – heart ailments
Darbari kanada	Headache Asthama
Gunakali	Rheumatic Arthritis, constipation headache, piles
Malkauns	Intestinal Gas- low BP
Bhupal Todi	High blood pressure
Shree	Anorexia, common cold, cough

V. SEQUENCE DIAGRAM



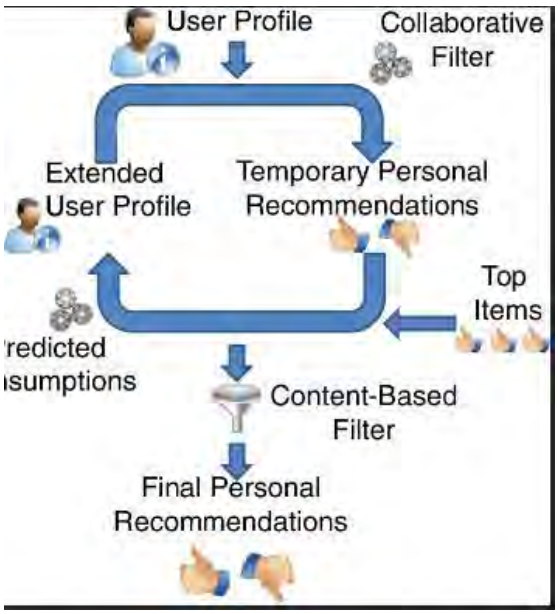
VI. CLASS DIAGRAM



VII. TECHNOLOGIES USED

1) Front End – React Native
2) Back End – Firebase (Database and Authentication)
3) Music Recommendation Algorithm – Collabrative filtering

VIII. WORKING OF COLLABRATIVE FILTERING ALGORITHM



IX. FUTURE WORK

The project "Svara-The Music Therapy App" ,can be further developed by following methodologies.

1) The project can be upgraded to take multiple stress parameters which would help to provide more specific music to a user.
2) Multiple databases can be used for different types of user.
3) There will be GPS tracker which will help to find the nearby Therapists.
4) The person will be providing with the more good music as per the requirements.

REFERENCES

[1] Yogesh Pingle,IOT for Music Therapy, IEEE Conference, October 2016.

[2] Deepika R. Chavan, M.S. Kumbhar, Rohit R.Chavan,The Human Stress Recognition of Brain, using Music Therapy, IEEE Conference, Septem-ber 2016.

[3] Masahari Yamamoto, Shinobu Naga, Jun Shimizu,Positive musical e_ect on two types of negative stressful conditions, IEEE Conference,

[4] S. Koger, K.Chapin, and M. Brotons, “Is music therapy an effective intervention for dementia - A meta-analytic review of literature”, International Journal of Music Therapy, 36(1):215, 1999.

[5] A. Konig, C. Crispim J., G. Bensadoun, P.D. P., F. Bremond, R. David P. Aalten,” Validation of an automatic video monitoring system for the detection of instrumental activities of daily living in dementia patients,” International Journal of Alzheimers Disease, 44(2):675685, 2015

[6] Antitza Dantcheva, Piotr Bilinski, Hung Thanh Nguyen,Expression Recognition for severely Demented Patients in Music Reminiscence-Therapy, IEEE Conference, October 2017.

BEST OUT OF WASTE(RECYCLING WASTE)

Milind Purohit, Jaspreet Kaur, Hitakshi Patel, Aditi Pandit, Prof. Yogesh Pingle
DEPARTMENT OF INFORMATION TECHNOLOGY
VIDYAVARDHINI'S COLLEGE OF ENGINEERING AND TECHNOLOGY, MUMBAI,INDIA

***Abstract*—The Radical idea (BEST OUT OF WASTE) is about to make natural products from the temples waste i.e. waste flowers. The idea is to collect waste flowers from the temples and recycle them in to products like charcoal-free incense stick, Vermicompostand florafoam -A Biodegradable Thermocol.**

I. INTRODUCTION

Reuse, Reduce, Recycle and Residual Management is the most essential mantra of the time for a greener, cleaner, healthier and happier earth. Reducing the most possible amount of waste. Reusing the resources to its optimum use. Recycling the products as many times as possible. Recovering the energy or material from solid waste stream to the fullest using technology. Lastly safe and effective waste management after the solid waste stream has been reduced through application of technology.

India is known as the "Land of Gods". Every year tonnes and tonnes of flowers are offered to the deties which finally makes their way to the "Ganges" resulting in contamination of our rivers. Using the concept of reuse, reduce, recycle, recoveryandresidualmanagementaproduct

"Vermicompost" is introduced which is made primarily by used flowers which are then organically treated and made into charcoal-free incense sticks which are Non Hazardous to health unlike others are available in the market. This white paper mainly focuses on the idea (HELPUSGREEN), which converts flowers waste collected from temples and mosques into patented organic manure and scent sticks.

Also now makers of "Vermicompost" have comeup with the idea of organic Thermocol i.e. Florafoam (An Eco-Friendly and Biodegradable Alternative to the Toxic Thermocol) which is again made out of waste flowers and is 27% cheaper than the existing one and is completely degradable.

II. NEED

India being one of the holiest place in the world, offering or carrying flowers, garlands and other quintessential religious materials is considered necessary. Many famous holy places such as Shirdi, Tirupati, Rameswaram, Vaishno Devi, Varanasi etc produces many tonnes of flowers on daily basis. These flowers are dumped on a dumping ground and no action is taken to dispose them properly. According to a survey, it was found that India has have 638,000 villages, 5000 towns and 400 big cities, so on an average we have around 2 million temples in India. A survey found that Kanpur city generates approximately 20 tonnes per day. Even if one temple produces 1000 kg of waste flowers per day, so calculating total wastage per day is 20,00,000 * 1000 = 2,00,00,00,000 (2 billion) flowers which is very gigantic and monumental. Treating and recycling this huge number of flower waste is very necessary since they are not treated properly and all this waste is dumped on a land or rivers which further pollute our environment so recycling waste flowers is extremely critical.

HEALTHY PRODUCTS:

This natural products not only emphasises on environmental and sociological benefits, but also focuses on personal benefits for our end users. Our values include making product of the maximum quality which are 100% organic, chemical free and promote a healthy living.

Incense Stick – Charcoal free to

Vermicompost – Organic soil for pesticide free healthy growth of plants

Florafoam – eco-friendly, bio-degradable instead of petroleum based products

Bio-leather – Made from flower waste instead of animal skin saving fauna

Soap – 100% natural for skincare

How natural products are created using waste flowers; converting them in to charcoal-free incense stick, Vermicompostand florafoam-A Biodegradable Thermocol ,

Converting flowers to incense sticks:

The waste flowers are recycled to incense sticks by first sun drying them to demoisturize the flowers completely. These dried flowers are then crushed in a mixer and turned into powder form. Wood saw dust is added to this powdered form to make it abrasive. To increase the efficiency and burning of sticks and also absorb oils , charcoal is added to the mixture. To bind the mixture properly, gum-powder is added to a blank stick so that the mixture sticks properly. To obtain good consistency of mixture, the mixture is hand rolled until the required consistency is obtained. The sticks are then dipped in

essential oils such as sandalwood, tea tree oil or lavender for the smell. One kilogram of incense sticks requires about 300 grams of essential oil. From around 500 grams of flowers, 150 grams of dried flower can be obtained.

INCENSE DIFFERENCE

Incense difference –

Natural incense –

- No charcoal
- Hand rolled
- Made from temple-flowers and plant resins
- Contains Milk and honey
- Hand dipped in Natural essential oils
- Smoke contains-
- Has therapeutic properties
- No sulphur content

Ordinary incense –

- Made form Charcoal
- Machine made
- Artificial Perfume
- Contains Potassium Nitrate(Saltpeter)
- Smoke contains –
- High Particulate matter
- Volatile compounds, such as benzene, toluene, and xylenes
- Produces aldehydes and polycyclic aromatic hydrocarbons(PAHs)
- High Sulphur Di oxide









Converting flowers to biodegradable thermocol:

Thermocol is basically a polymer of styrene or phenylethane which is extremely dangerous to environment. Many disadvantages of thermocol is that they are highly toxic, they turn soil infertile, extremely difficult to recycle and decompose, it clogs the ground and many more. But in todays world, it is used extensively ignoring the problems created due to thermocol. Many online things and delicate things are packed in thermocol and thrown away after unpacking of the item. Around 91% of thermocol is used only once. Biodegradable thermocol is an alternative to thermocol. Floraform is made using these waste flowers and biodegradable thermocol is made out of this floraform. An additive advantage of biodegradable thermocol is that it is 27% cheaper than traditional thermocol.

III.PROBLEM STATEMENT

Flowers are colourful, conspicuous structure associated with angiosperms, frequently scented and attracting various insects. Flowers are useful for many purposes they are used in scents, perfumes, decoration, medications, cosmetics etc. But the major percentage of flowers are used in our worship. India has two millions Gods and we worship them all. Tonnes and Tonnes of flowers are used for decoration, offerings, garlands, to our dietes. Flowers are such inseparable part of our culture that we cannot think of its alternative but problem arises when these flowers make way to our solid waste stream where they contaminate mosquitoes breed on them. Every year 80,000 MT of waste flowers are dumped into river Ganges. Toxic Arsenic, Lead and cadmium from the harmful farm run-off, pesticides & insecticides used to grow flowers mixes with the river water making it highly poisonous. When they are left in our holy rivers our marine life gets affected. If not used purposefully these floral waste lead to many diseases.

IV. CURRENT SYSTEM

The current system is that every day tonnes and tonnes of flowers are offered to GOD during worship. This generates huge amount of waste on daily basis and causes many problems which leads to our main stream solid waste, water pollution and land pollution.

V. PROPOSED SYSTEM

This system we proposed will reduce the problems related to solid waste, water pollution and land pollution. The amount of waste generated will be recycled and reused leading to lesser pollution, greener earth and resources used effectively. It will also lead to mass employment specially women empowerment as many women generate their source of income from this stream.

VI. SCOPE

Sources are limited demand is unlimited due to growing population we are more likely to fall scarce of diminishable resources in our near future. Only effective use of available resources can help us to restore them leading to safer and cleaner future for us and our future generation.

And other features of this policy are:

- Decline in cost of production.
- Growth in GDP.
- Growth in small scale industries.
- Mass employment in rural areas leading to their better lifestyles.
- Reduction in diseases .

VII. CONCLUSION

More and more industries should venture into this sector leading to its growth, preservation of diminishable resources . Not only this, we should try and reuse most possible resources in our life so that our future generation gets a better life and we and our world gets a better future.

VIII. ACKNOWLEDGEMENT

The author would like to thank Prof.YogeshPingle for his continued support and guidance. He would also like to thank all his friends and family, without their enthusiasm this project would not have been possible.

IX. REFERENCES

[1] Medvedev Alexey, Fedchenkov Petr, ZaslavskyArkady, AnagnostopoulosTheodoros, Khoruzhnikov Sergey, Waste Management as an iot-Enabled Service in Smart Cities, 2015.
[2] Hong Insung, Park Sunghoi, Lee Beomseok, Lee Jaekeun, JeongDaebeom, Park Sehyun, iot-Based Smart Garbage System for Efficient Food Waste Management, 2014.
[3] SharholyMufeed, Ahmad Kafeel, Mahmood Gauhar, Municipal solid waste management in Indian cities - A review, 2007.
[4] HimadriNathSaha, SupratimAuddy, Subrata Pal, Shubham Kumar" Waste management using Internet of Things (IoT)" Bangkok, Thailand, 2017

Flora Export

Jaspreet Kaur Thethi
Vidyavardhini's College of Engineering and Technology,Vasai
Mumbai, India

Hitakshi Patel
Department of Information Tech.
Vidyavardhini's College of Engineering and Technology,Vasai
Mumbai, India

Prof. YogeshPingle
Assistant Professor
Department of Information Tech.
Vidyavardhini's College of Engineering and Technology,Vasai
Mumbai, India

***Abstract*—Where India has two million gods and people worship them all in religion all other countries are pauper. India is the only millionaire. Badrinath, kedarnath, Konark Sun, Somanth, Rameshwaram, Golden Temple, Hemkunt Sahib, Shirdi and many more worship houses where tonnes and tonnes of flowers are being used. Approximately 40 tonnes of flowers are used for decorating purposes at Lord Venkateshwar and such tonnes and tonnes of flowers in all other worship houses which are then thrown. Pollution caused by flowers is unlikely dangerous compared to industrial waste, but throwing of flowers into river and its rotting may release pesticides used in them which can affect marine life. So in this case these wasted flowers can be reused in making biodegradable thermocol, incense sticks and many more useful biodegradable products. So an application can be created which will help contacting temple or local village people to industrial people to recycle and reuse the waste sent to them so that different biodegradable products can be made out of it. Using this app, any temple authority will register and give the buyer the information about tonnes of flower which is being wasted and polluting the environment so that the buyers can directly talk and receive his order.**

Keywords—NodeMCU, PIR sensor, LM393 LDR sensor

I. INTRODUCTION

In terms of population and pollution INDIA is in the diplomatic state. According to the study of general register of India census report of 2011 said that population of farmers or cultivators in India is around 118.7 million and 144.3 million of agricultural workers which consists of 31.55 to million count of farmer's chunk of populations of farmers who have all been into floriculture. West Bengal is known as the flower state of India. In fact 45% of flowers in India are produced in West Bengal. Flowers are the colorful conspicuous structure associated with angiosperms, frequently scented and attracting various insects. These flowers are used for medicinal purpose, perfumes cosmetics bouquets and mainly in puja. Thus INDIA is one of the holiest place in the world where offering and carrying flowers, garlands and other religious materials is considered as fortunate and prosperous.

II. CURRENT SYSTEM

As flower waste recycling is a new concept that few people are aware about, but how to recycle and where to give it to recycle is not known. People may have queries related to recycling process or to buy different recycled products to help in "GO GREENER" and "HELP US GREEN" activities

.There is no such website or mobile application that helps user to interact about the waste and send it for recycling process. This kind of application is in extreme need since the waste production at every temple is too large and needs to be managed or disposed properly.

III. PROPOSED SYSTEM

This software will make user to contact associations that recycle flower waste. Also it will save time for the temple member to go and meet the recycling association personally. This software can be made with any of the front-end language for presentation like JAVA, PHP,
.NET, VISUAL BASICS, JAVVASCRIPT (NODE JS), PYTHONetc.and a back-end server such as SQL, JAVA, C SHARP, RUBY, MY SQLITE, ASP.NET, DJANGO, MONGO etc. This application will also have a feature named navigation which will allow the user to share the location.

IV. SOLUTION APPROACH

According to a survey done by United Nations on 16th November 2018, current population of India is 1,359,561,937 i.e. one billion three hundred fifty-nine million five hundred sixty-one thousand nine hundred thirty-seven [1]. That says India is a country with very large population and People in India are extremely religious. Approximate waste production in India is about 62 Million Tones including degradable and non-degradable waste per year and this value increases approximately 4 percent every year out of which solid waste consists of 36.5 percent roughly. If scenario doesn't change then it will be very difficult to manage pollution produced due to rapidly increase in waste generation. Thus waste management is extremely important at today's date. The first step that should be taken is to segregate the waste at homes or at source so that the waste that can be recycled are sent for its recycling purpose. This step can help reducing the waste allocation or we can say that while dumping wastes to landfills or other places, waste will be reduced. Now-a-days People are aware about this a little and also managements started taking care about it a little.

India ranks 4th on being the most religious country. Hundreds of people visit temple daily. People take many kind of stuff to offer in temple which mainly consists of flowers, garlands and many more. People believe that flowers that are offered in temples cannot be thrown in garbage since it divine or one can say it is sacrosanct. According to a survey approximately total of 2,00,00,00,000 that is two billion flower waste is generated per year. This flower waste generated is not decomposed properly instead they are thrown in rivers and further water pollution increases. There were noappropriate measures taken to decompose or recycle these flora waste, until Kanpur city started recycling this waste into many different biodegradable products such as incense stick, biodegradable thermocol, biodegradable leather, fertilizers , Potpourri, Homemade Natural Soap, scented candles , sans coal etc. Thus recycling flora waste can be useful in two ways. Firstly the waste reduces and secondly the jobs are increased which is again one of the issues in India. Increase in number of jobs will help India develop faster.

Now that the awareness for flower waste recycling has raised among people of India, there is still no proper measures taken. There is no system for someone to call the respective officials and to ask to collect the flowers. Even though

people have learned that recycling flower waste is a “GO GREENER” method but implementation is still not done in proper manner. Currently there is no way a person can communicate to the institute that recycles this waste. The institute should get proper message about the collection of waste so that they can pick and collect them and send it for recycling process. Even if there is collection of flora waste somewhere on road or any landfill, a person can inform the respected institute and respective measures can be taken. An application which can help users contact the organization/institute that recycles the waste will help people a lot. “FLORA EXPORT” application can be built using any language. Maybe HTML, CSS,ASP DJANGO for front-end and making it user interactive/friendly and MySQL, php, oracle for back-end to store the information about sender and receiver or to track previous history between the sender and receiver. Also it will support different viewports so that the application can be used on mobile phones or laptop and tablets.

The application “FLLORA EXPORT” may help the people to communicate about it. Suppose for a temple, the temple priest or any of the temple authority or temple member can register on the application and they will be provided with the information about the institutes that recycle ate flowers and convert them into biodegradable products. So the user can reduce the loads of waste flowers and instead of throwing them into river or any other water bodies, it can be recycled.

On religious terms, a user may also perform a good deed by doing this as it reduces water pollution saving hundreds and thousands of marine life and by providing employment to many people. In simple terms, this will reduce the work load of the user in contacting the institutes recycling waste and also it will help the institute in a way such that it need not go everywhere and search for the waste. Also user can buy recycled products at the same website and can help nature by using biodegradable products and at a low rate. This application will make communication between the institute and temple members convenient and easy. Other advantages of using this application are as follows:

1. Convenience

A sender who wants to send the information about gathered flora waste can be given at anytime, anywhere by just using a simple application. Also biodegradable products can be purchased anytime and anywhere according to his/her convenience.

2. Time saving

Time of a person can be saved as he/she would not need to go to the place where products are recycled. Just by sharing the location, the user can inform the association that recycles flower waste so that they can collect it from that specified location. Also purchasing 3. Promotes Flexibility

User can inform the association from anywhere and call the association to collect the waste according to their time and associations working hours.

4. Easy Contacting

The temple authority or any lay man can contact the association for any kind of detail.

5. Reduced water pollution:

Waste flowers which were thrown in water bodies will reduce and hence the endangered marine life can be saved.

6. Increases Employment

As many associations are emerging for recycling of waste flowers, there will be increased job opportunities. Women’s of rural area can work for such organizations and earn some credits.

7. Easily Keeping a Record

Previous records can be tracked thus letting the user to have full knowledge about how much he/she had send waste for recycling process and to which association the waste was send and all other details .

8. Nature friendly

The products that are recycled are biodegradable and recycled. Thus it is nature friendly. Also a user can purchase any biodegradable products at his/her own convenient time, place and it will save time as well.

V. CONCLUSION AND FUTURE WORK

This application will help user or temple members to conveniently communicate directly with the associations that recycle flower waste thus saving precious time of and to get the details about recycling process. Also the user can buy or purchase biodegradable products thus going GREENER and helping nature. Even if 15-20 percent of people will use this app and help the nature, it will help reducing waste for around 35 percent. Also if the users regularly update and use this app for temple flower waste disposal, the flowers will be disposed properly and recycled, many biodegradable products can be made out of it and also water pollution can be reduced at a great extent saving thousands of marine life. As more and more people come to know about this technology and will use our application effectively huge amount of waste can get disposed easily. Using this app users can conveniently communicate with the associations and also the association can contact temple authorities on regular basis. This application will even help user to buy or purchase any biodegradable items recycled out of these waste flowers.

ACKNOWLEDGMENT

The author would like to thank Prof. YogeshPingle for his continued support and guidance. He would also like to thank all his friends and family, without their enthusiasm this project would not have been possible.

REFERENCES

[1]HimadriNathSaha, SupratimAuddy, Subrata Pal, Shubham Kumar”
Waste management using Internet of Things (IoT)”Bangkok, Thailand, 2017

[2] SharholyMufeed, Ahmad Kafeel, Mahmood Gauhar, *Municipal solid waste management in Indian cities - A review*, 2007.

[3] Hong Insung, Park Sunghoi, Lee Beomseok, Lee Jaekeun, JeongDaebeom, Park Sehyun, *iot-Based Smart Garbage System for Efficient Food Waste Management*, 2014.

[4] S. Abdoli, *RFID Application in Municipal Solid Waste Management System*, 2008.

[5] Medvedev Alexey, Fedchenkov Petr, ZaslavskyArkady, AnagnostopoulosTheodoros, Khoruzhnikov Sergey, *Waste Management as an iot-Enabled Service in Smart Cities*, 2015

Blockchain For Electronic Health Records(EHR)

Sarthak Taneja, Yashvardhan Tewari, Gaurang Tyagi
Computer Science Department, *Symbiosis Institute Of Technology,* Pune, India
sarthak.taneja@sitpune.edu.in, yashvardhan.tewari@sitpune.edu.in, gaurang.tyagi@sitpune.edu.in

Abstract- The Electronic Health Record (EHR) is a digital version of a patient's medical history maintained by a healthcare provider over the course of their visits. The record includes patient information on demographics, diagnosis, vital signs, past medical history, progress over time, lab tests and more.

The biggest challenge that is being faced by health care systems throughout the world is how to share medical data with known and unknown stakeholders for various purposes while ensuring data integrity and protection of patient privacy.

The blockchain solution for healthcare industry drastically reduces the time required to access patient's information, enhances system interoperability and improve data quality. Once a blockchain is deployed to manage EHRs, it becomes the unified and common backbone for digital health. Expanding on the consequences of this development on the future of specialized back-end systems, the implication of using this backbone is that each hospital or care provider no longer needs a specific version of databases or software to access patient data.

The purpose is to build a prototype that replaces currently prevalent intermediary methods for information exchange, such as a state-designated HIE with blockchain powered DAPPS.

Keywords

EHR- Electronic Health Records, **DAPPS**- Decentralized Applications, **HIE**- Health Information Exchange

I. INTRODUCTION

A peer-peer version of electronic health records would help the interoperability and improved participation of various stakeholders like the patients and the doctors. EHR contains digital information of patients medication history, the diagnosis prescribed and their lab records. It provides the solution to the problem of sharing medical data with stakeholders and also ensuring data privacy. Currently, the medical industry is relying on client-server architecture. Every health centre has its own database and faces various technical challenges like maintenance, security and more. The sensitive nature of health data along with the challenge of information exchange has created opportunities for a blockchain approach.

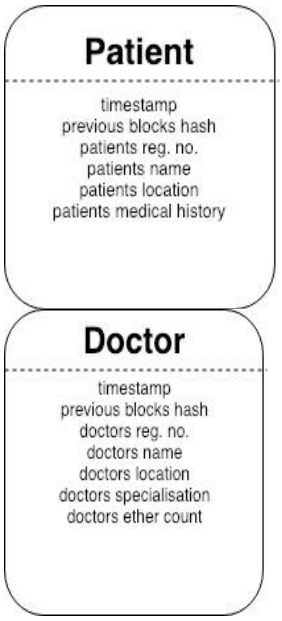
II. ARCHITECTURE

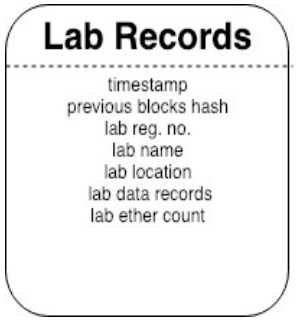
A blockchain is a mesh network of computers which are linked to each other and not a central server. It is a simple distributed data structure, which is built linearly, over time and is independently verified that is audited by the various stakeholders of the network. Various architectural components of a blockchain are:-

A. Block:

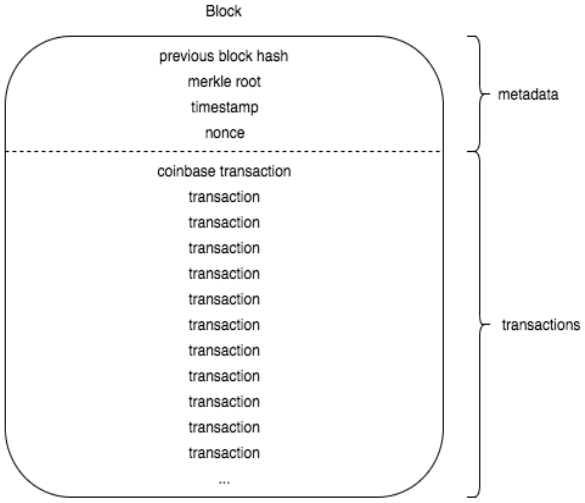
Blocks are data structures whose purpose is to bundle sets of transactions and be distributed over nodes in the network. Blocks contain a block header, which is the metadata that helps verify the validity of a block.
Typical block metadata contains:

- version - the current version of the block structure
- previous block header hash - the reference this block's *parent block*
- merkle root hash - a cryptographic hash of all of the transactions included in this block
- time - the time that this block was created
- nBits - the current *difficulty* that was used to create this block
- nonce ("number used once") - a random value that the creator of a block is allowed to manipulate however they so choose.

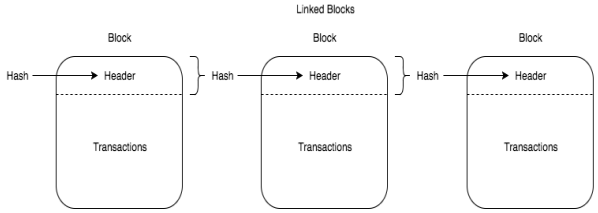




The first block of the blockchain is called the genesis block



A blockchain is a growing list of records, called blocks, which are linked using cryptography.



B. Transactions

Transactions are building blocks of a blockchain system. Transactions generally consist of a recipient address, a sender address, and a value. A transaction *changes the state* of the agreed-correct blockchain. Transactions are bundled and delivered to each node in the form of a block. As new transactions are distributed throughout the network, they are independently verified and verified by each node. It is this simple value-transfer mechanism using "transactions" over time that create a blockchain. As each transaction occurs, it gets added to the blockchain ledger.

C. Consensus

Consensus rules dictate that only valid changes to the blockchain will be accepted by everyone else. This results in a system that economically guarantees that only valid blocks will be worked on, submitted to the network, and accepted by the greater community. The bigger the blockchain network the better it is as the number of miners and nodes increases as the number of people enforcing their own rules increases.

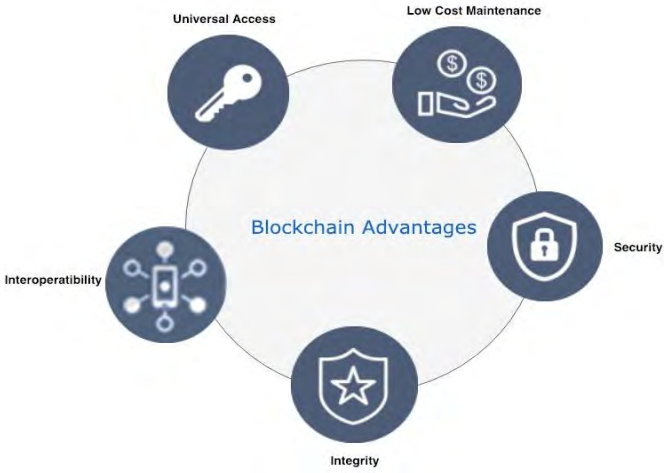
D. Mining

Mining is the process of putting in the effort for the creation of blocks that is accepted by the rest of the network.

The role of the miners is to verify the cryptographical authenticity of a transaction and then packaging them into blocks which are stored on the blockchain.

Benefits of blockchain architecture:

- Decentralization - each member of the blockchain structure has access to the whole distributed database. As opposed to the central-based system, a consensus algorithm allows for control of the network
- Anonymity- each blockchain network participant has a generated address, not user identity. This keeps users' anonymity, especially in a public blockchain structure
- Transparency - the blockchain system cannot be corrupted. This is very unlikely to happen, as it requires huge computing power to overwrite the blockchain network completely
- Immutability - any records made in a blockchain cannot be changed or deleted



III. STAKEHOLDERS

Our project focuses on presenting a prototype model for getting the basic know-how of the POC blockchain apps.

This project aims to establish a secure and cost-effective interaction between stakeholders via blockchain.

Various stakeholders are:

- Patients

  Patient are the primary stakeholder of this system.Patients comprehensive and immutable log of information gives easy access to their health information over providers to doctors and other participating stakeholders.

- Doctors

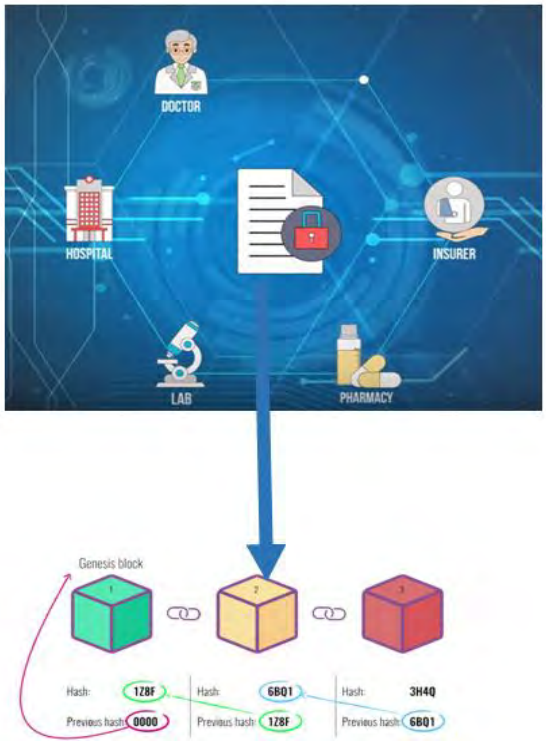
  The main purpose of this system is to bring patients into direct contact with doctors, protecting the privacy of patients data at the same time. Hence inclusion of doctors limits in smart contracts is a must. The doctors will be further benefited by the controlled sharing of data on a shared platform.

- Lab Scientists

  They buy patient's data in terms of ether value points. Hence, patients can earn credibility points in return for their own data which they chose to share or hide. The data purchased would be put into use for research work and diagnosis purposes, thus benefiting all stakeholders.

This will enable the stakeholders to access and analyze records as well as acknowledge transactions based on the value of the work offered or proposed by one party to other and common consensus on a distributed ledger via blockchain.

The interaction between stakeholders and their transaction terms using smart contracts is depicted :



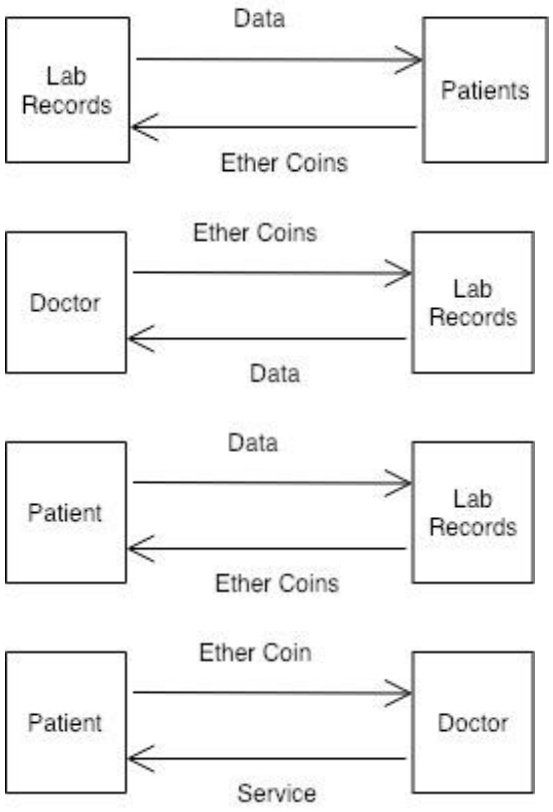
Interactions of various stakeholders

IV. SMART CONTRACTS

This smart project criterion helps to establish a project. It then helps in setting a funding goal and starts collecting funds from the stakeholders according to the guidelines framed in the contract. All the stakeholders mentioned in the contract had to agree upon the conditions about the rules and regulation of the project, monetary policy etc and had to abide by them.

The result of the contract is validated by everyone listed on the contract. Tempering with the contract is almost impossible.

Ethereum is the biggest platform for smart-contract. It was specifically created and designed to support smart-contract.



V. SOFTWARE TOOLS AND TECHNIQUES

A. *Ganache* - Ganache allows you to create a private Ethereum blockchain for you to run tests, execute commands, and inspect state while controlling how the chain operates. It gives you the ability to perform all actions you would on the main chain without the cost.

B. *Solidity*- Solidity is an object-oriented language used for writing smart contracts.

C. *Ethereum*- Ethereum is the biggest platform for smart-contract. It was specifically created and designed to support smart contract.

D. *MetaMask*- Its extension turns google chrome into Ethereum browser to retrieve data from the blockchain and users can manage identities and sign transactions.

E. *Truffle*- Truffle is a developer environment, testing framework and asset pipeline for blockchains. It allows developers to make changes to the smart contract and provides with a project structure, files, and directories that make deployment and testing easy.

F. Testing Tools

- Mocha: Tool used for asynchronous testing. It is accurate and flexible for reporting and identifying uncaught code exceptions.

- Chai: Chai is a BDD / TDD assertion library for node and the browser that is paired with any javascript testing framework.

We use these testing frameworks because for every "WRITE" transaction gas is deducted from the ether account, instead of testing your code after every code modification you can use a framework like chai or mocha for testing your smart contract without wasting ether gas.

VI. CONCLUSION

Instead of relying on a designated intermediary for information exchange, such as a state-designated HIE(Health Information Exchange) or a private network established between local hospitals, the decentralized nature of the blockchain would allow any approved participants to join an exchange community, without the need to build data exchange pipes between certain organizations.

VII. REFERENCES

[1] MedRec: A case study for blockchain in healthcare- MIT

[2] Bitcoin: Peer to Peer Electronic Cash System- Satoshi Nakamoto

[3] Real blockchain use cases for the health industry- John. D Halamka

[4] Blockchain technology in healthcare: The revolution starts here- IEEE



IC-ISCN - 2019

International Conference on Intelligent Systems and Communication Networks



IC-TELCON - 2019

International Conference on Trends in Electronics & Communication



IC-AMCE - 2019

International Conference on Advances In Mechanical & Civil Engineering



CHSTE - 2019

Conference on Humanities, Science and Technical Education



EC - 2019

Engineering Colloquium Post Graduate Students & Doctoral Scholars



EWPPC - 2019

Engineering Workshop & Paper Presentation Contest



IOTW - 2019

Industry Oriented Technology Workshop



PBW - 2019

Professional Body Workshop

A - Block, Thakur Educational Campus, Shyamnarayan Thakur Marg, Thakur Village, Kandivali (East), Mumbai - 400 101
Tel.: 022-6730 8000 / 8106 / 8107 Telefax: 022-2846 1890
Email: tcet@thakureducation.org Website: www.tcetmumbai.in www.thakureducation.org