Safety Companion

Submitted in partial fulfilment of the requirements

of the degree of

BACHELOR OF ENGINEERING

in

INFORMATION TECHNOLGY

(A.Y. 2018-2019)

by

Harsh Agarwal (Roll No.:03)
Priyank Dave (Roll No.:24)
Tanish Jain (Roll No.:39)

Under the Guidance of

Mrs. Purvi Sankhe

Assistant Professor, I.T Department, TCET

Estd. in 2001

Choice Based Credit Grading System with Holistic Student Development

(CBCGS-H 2019)

Jagdu Singh Charitable Trust's (Regd.)

THAKUR COLLEGE OF ENGINEERING & TECHNOLOGY

Autonomous College Affiliated to University of Mumbai

Approved by All India Council for Technical Education(AICTE) and Government of Maharashtra

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

Programmes Accredited by National Board of Accreditation (NBA), New Delhi*

Conferred Autonomous Status by University Grants Commission (UGC) for 10 years w.e.f. AY 2019-20

Amongst Top 200 Colleges in the Country where Ranked 193rd in NIRF India Ranking 2019 in Engineering College category

*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)

*3rd cycle NBA Accredited UG Programmes : • Computer Engineering • Electronics & Telecommunication Engineering • Information Technology (3 years w.e.f. 01-07-2019)

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

Certificate

This is to certify that Mr. Harsh Agarwal , Mr. Priyank Dave , Mr. Tanish Jain are bonafide students of Information Technology Department, Thakur College of Engineering and Technology, Mumbai. They have satisfactorily completed the requirements of PROJEC Under Project Based Learning as prescribed by **Thakur College of Engineering and Technology (An Autonomous College affiliated to University of Mumbai),** while working on "Safety Companion".

Signature :----------
Name : Mrs. Purvi Sankhe
AP-IT

Signature :----------
Name : Dr. Rajesh S. Bansode
HOD-IT

Signature: ----------
Name : Dr. B. K. Mishra
Principal,
Thakur College of Engineering and Technology.

Internal Examiner:

Signature :----------

:

External Examiner:

Signature :----------

Name :

Thakur College of Engineering and Technology, Kandivali(East) Mumbai.

Date:

Place:

DECLARATION

I/we declare that this written submission represents my/our ideas in my/our own words and where others ideas or words have been included, I/we have adequately cited and referenced the original sources. I/we also declare that I/we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my/our submission. I/we understand that any violation of the above will because for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

1. ----------
   Harsh Agarwal (Roll No. : 03)

2. ----------
   Priyank Dave (Roll No. : 24)

3. ----------
   Tanish Jain (Roll No. : 39)

Date:

ABSTRACT

Engineers and researchers in the automobile industry have tried to design and build safer automobiles, but traffic accidents are unavoidable. Today, traffic safety is one of the main priorities of governments. Road accidents may not be stopped altogether, but can be reduced. Driver emotions such as sad, happy, and anger can be one reason for accidents. At the same time, environment conditions such as weather, traffic on the road, load in the vehicle, type of road, health condition of driver, and speed can also be the reasons for accidents.

Considering the importance of topic, identifying the factors of road accidents has become the main aim to reduce the damage caused by traffic accidents. Data mining allows users to analyze data from many different dimensions or angles, categorize it and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. One of the key objectives in accident data analysis is to identify the main factors associated with a road and traffic accident.

Association rule mining is used to identify the various circumstances that are associated with the occurrence of an accident. Patterns involved in dangerous crashes could be detected if we develop accurate prediction models capable of automatic classification of type of injury severity of various traffic accidents. These behavioral and roadway accident patterns can be useful to develop traffic safety control policies.

List of Figures

C O N T E N T S

CHAPTER 1

OVERVIEW

1.1 Introduction

Today, traffic safety is one of the main priorities of governments. Considering the importance of topic, identifying the factors of road accidents has become the main aim to reduce the damage caused by traffic accidents. Road safety means development and management of roads, provision of safer vehicles, and a comprehensive response to accidents. The costs of fatalities and injuries due to traffic accidents have a great impact on the society. In recent years, researchers have paid increasing attention to determining factors that significantly affect severity of driver injuries caused by traffic accidents.

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. It allows users to analyze data from many different dimensions or angles, categorize it and summarize the relationships identified. Driver emotions such as sad, happy, and anger can be one reason for accidents. At the same time, environment conditions such as weather, traffic on the road, load in the vehicle, type of road, health condition of driver, and speed can also be the reasons for accidents.

Hidden patterns in accidents can be extracted so as to find the common features between accidents. Applying data mining techniques to model traffic accident data records can help to understand the characteristics of drivers' behavior, roadway condition and weather condition that were causally connected with different injury severity. We will use different machine learning algorithms to analyze those hidden patterns. This can help decision makers to formulate better traffic safety control policies.

1.2 Background

Road accidents severity is increasing at alarming rate. Regulating the traffic accidents on roads is a vital task. Road safety becomes a major public health concern when the statistics show that more than 3000 people around the world succumb to death daily due to road traffic injury. The huge econmic losses are an economic burden for developing countries. The road data are necessary not only for statistical analysis in setting priority targets but also for in-depth study in identifying the contributory factors to have a better understanding of the chain of events.

1.3 Importance of Project

In order to reduce the number and severity of traffic accidents, the reasons and the main factors should be known. Although, the driver' fault is the main reason in most of the accidents, physical environment can cause accidents. The capacity, type, material and surface of road, the presence of traffic light, lightening degree, lane, pedestrian-way, crossings and intersection can affect the occurrence of accident. Traffic accidents reduce level of traffic safety and cause traffic congestions. In order to increase the level of traffic safety, it is necessary to ensure that the causes of accidents are revealed and the risks are minimized.

1.4 Objectives and Scope of the Project

Road accident analysis can help in saving lives of many people. Road accidents has many reasons behind it. They include technical failure, driving unconsciously, etc. Thus, to study all these factors and provide solutions to those problems is the major objective of the project.

The following are the objectives of the project:

1. To collect data of a particular region on the globe.
2. To generate analysis report of those data using the KDD process
3. To generate solutions in terms of construction of road infrastructure that can help in preventing the crashes.
4. To help save life of people and infrastructure

Scope:

This analysis models has a great scope in current scenario and in future. The current scope is that we have a roadways highways that have steep turns and narrow roads and roots because of which there are chances of accidents i.e. there are more road accident prone regions, we can arrange a safety mechanism that protects the crashes of the vehicles. Future scope is that there can be a system with a similar framework can help in crime forensics .i.e. there can be a system that can help in crime investigation department and Intelligence Department.

1.5 Summary

The chapter consists of the introduction to the concept the background of the title selected and discuss about the why this project is important the scope of it etc.

CHAPTER 2

LITERATURE SURVEY

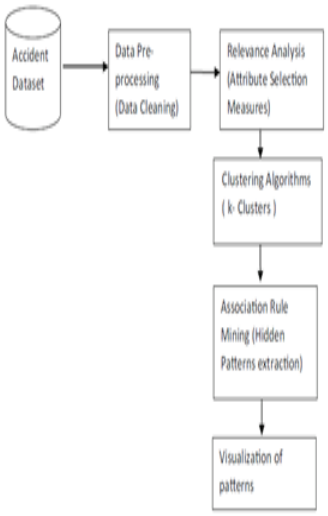
& PROPOSED WORK

2.1 Introduction

A writing literature review surveys overview insightful articles, books, papers, gathering procedures and different assets which are pertinent to a specific issue, zone of research, or hypothesis and gives a setting to an exposition by recognizing past research. Research recounts a story and the current writing causes us to recognize where we are in the story as of now. It is up to that composition a paper to proceed with that story with new research and new points of view yet they should initially be comfortable with the story before they can push ahead.

2.2 Literature Survey Table

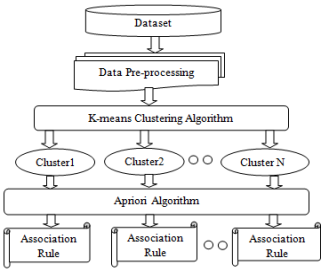
Table no.2.2.1: Literature Survey Table

Ref No.	**Year**	**Authors**	**Title**	**Methodology**	**Key findings**	**Research gaps**
[1]	2018	S.Vasavi	Extracting Hidden Patterns Within Road Accident Data Using Machine Learning Techniques		Used EM clustering and K-medoid method to cluster the data	The performance of the EM clustering is low compared to K-medoids clustering Algorithm,

						because it uses probability measures to cluster the data.
[2]	2018	Sannaila Praveen Kumar, M. Prashanti	Road Accident Data Analysis Using Machine Learning	Association rule mining is applied on dataset to find the feasible rules that are mostly contributing the road accidents. Feature selection is applied on the pre-processed dataset to find most effecting rules in road accidents by selecting the appropriate attributes	This paper proposed apriori algorithm and feature selection method to find the most effecting rules that are contributing the road accidents. This paper implemented machine learning algorithm to find the contributing factor of vehicle speed in road accidents	Apriori algorithm includes a lot of iterations

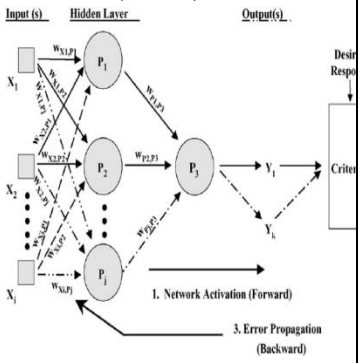
[3]	2018	Shrishti Sonal, Saumya Suman	A framework for analysis of road accidents	A. Data Collection B. Analysis and implementation	Use of linear regression model Tools-Python and Anaconda.	The machine learning algorithms can be optimized when more than one algorithm are used together to generate the required output with more accuracy. Hence a lack of accuracy can affect the quality of predicted results.
[4]	2017	Poojitha Shetty, Sachin P C, Supreeth V Kashyap, Venkatesh Madi	Analysis of Road Accidents using data mining techniques Published in International Research Journal of Engineering and Technology	Step1:Apply Association rule for generating pattern and to find which attribute value or item is bound with the other how strongly Step2: Classification of the accident	1. Used Apriori Algorithm for finding Association rule among the items of transitions in each individual dataset based on support and	Apriori algorithm require large amount of data storage as it generates candidates table first which consists of

			(IRJET)		confidence 2. Used Naïve Bayes algorithm for classification	analyzing the dataset for each data item(Time complexity=n^n where n is number of items or values in the dataset) hence it takes a lot of time to generate patterns for a dataset of 2MB Uses of FP growth can be made in order to perform association as it provides better latency. 2. Similarly instead of naïve bayes, Decision tree based classificati

						on can also be done in order to get the prediction of type of road accident
[5]	2017	Ms. Gagandeep Kaur, Er. Harpreet kaur	Prediction of the cause of accident and accident prone location on roads using data mining techniques	The Methodology is applied using R's IDE Integrated Development Environment (Rstudio) which is graphical and statistical computing tool on road accidental dataset for analyzation. R is free software, GNU package used for analyzing the datasets to vaticinate the hidden patterns and trends.	Regression and Co-relation models used for analysis of the accident data.	Neural networks are the better way to predict and analyze things in a better way and in a more accurate way with the help of various parameters .
[6]	2017	Priyanka A. Nandurge, Nagraj V. Dharwadhkar	Analyzing road accident data using machine learning paradigms.		Segmentation is used to reduce data heterogeneity using a number of measures such as expert knowledge, but there is no guarantee	The main difficulty of clustering algorithm is to estimate the number of clusters. In k-means clustering,

					that this will result in the best segmentation of the group including road accidents. Cluster analysis can helps to segment road accidents data.	value of k must be given by the user which is one of the limitations of this algorithm. If the value of k is incorrect then it may lead to incorrect clustering results.

[7]	2005	Miao Chong, Ajith Abraham and Marcin Paprzycki	Traffic accident analysis using machine learning paradigms	Step1: Gather the Dataset. Step2: Provide the dataset to different kinds of ML paradigms (such as ANN, Decision Tree and SVM. 1.Decision Tree: Based on the dataset certain rules are formed by scanning the dataset. Based on these rules the prediction takes place. 2. Support Vector Machines(SVM): SVM involves two key techniques, one is the mathematical programming and the other is kernel functions. Input vectors are mapped in a space linearly or non-linearly (depending on kernel function) A hyperplane is constructed for seperation of classes. Based on the intersection and distance between hyperplanes, the support vector algorithm increases the distance between the hyperplanes to occupy more positives. When the two classes are non-linearly seperable, SVM transforms the points into higher dimensional space.	1.Decision Tree: Classification and Regression Tree (CART) is use, in which each internal node in the tree specifies a binary test on a single variable, branch represents an outcome of the test, each leaf node represent class labels or class distribution. 2. Support Vector Machines(SVM): 3. Artificial Neural Network(ANN):	1. CART does not use combinations of variables. Also a small change in the data can lead to a large change in the structure of the optimal decision tree. 2. SVM require more speed and size requirements both in training and in testing. 3. If the neural network is large then the processing time required is high. The

				3. Artificial Neural Network(ANN):  A Multilayer Perceptron (MLP) is a feed forward neural network with at least one concealed layers. The network consists of an input layer of source neurons, at least one hidden layer of computational neurons, and an output layer of computational neurons. The input layer accepts input signals and redistributes these signals to all neurons in the hidden layer. The output layer accepts a stimulus pattern from the hidden layer and establishes the output pattern of the entire network.		neurons and layers for such ANN are difficult to understand. This affects learning phase as it may become slow.

2.3 Problem Definition

The main objective of this project is to investigate the role of human, vehicle, and infrastructure-related factors i.e. both internal and external parameters in accident severity by applying machine learning techniques on road accident data. The phases in which a part of problem will be solved.

Phase1: Data collection and Data gathering clearing the data and generating it in a single format so that it can be processed easily without any error.

Phase2: Understanding the dataset what all parameters it consists of and how the attributes in the dataset are interrelated to each other and to see in what way algorithms can be used to generate the frequent valuable pattern using certain algorithms like clustering, association rule mining, correlation and regression model (Big Data Analytics) to know how we can apply artificial neural network to it.

Phase3: Implementing the artificial neural network for generating the road accident pattern for the dataset. Understanding, how for a particular regions the internal and external parameters affect.

Phase4: Generating solutions to avoid the reasons that cause the accident

2.4 Features of the Project

The project is analysis based project hence the features of this project are as:

1. Better accuracy in regards to algorithms used to perform certain analysis
2. Analysis on real time data set (Original dataset)
3. Works on the historical data available hence effective

2.5 Methodology Used

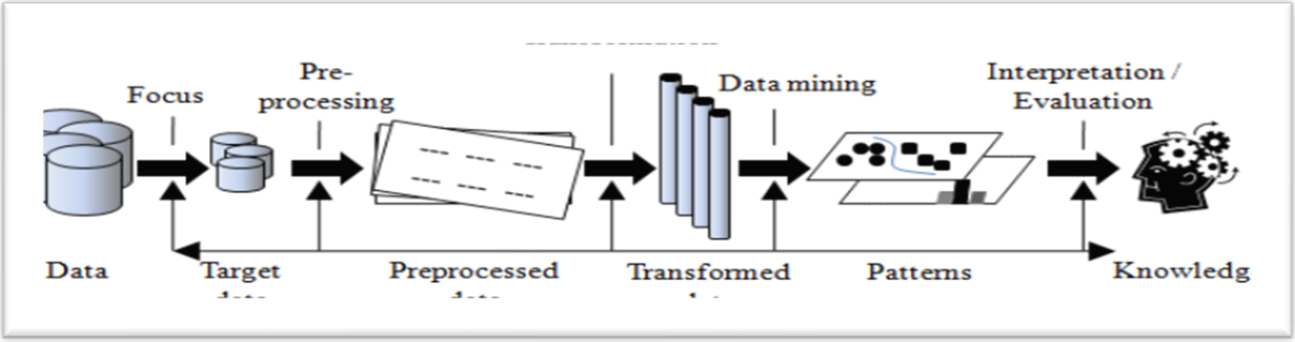


Fig.2.5.1: KDD process

For implementing the given problem definition, the classical KDD (Knowledge Discovery in Database) process which refers to non- trivial extraction of implicit, previously unknown and potentially useful information from the data stored in databases. This process is going to be used which involves the following:

1. Data Cleaning: Data cleaning is defined as removal of noisy and irrelevant data from collection.
    i. Cleaning in case of Missing values.
    ii. Cleaning noisy data, where noise is a random or variance error.
    iii. Cleaning with Data discrepancy detection and Data transformation tools.

2. Data Integration: Data integration is defined as heterogeneous data from multiple sources combined in a common source (Data Warehouse).
    i. Data integration using Data Migration tools.
    ii. Data integration using Data Synchronization tools.
    iii. Data integration using ETL(Extract-Load-Transformation) process.

3. Data Selection: Data Selection: Data selection is defined as the process where data relevant to the analysis is decided and retrieved from the data collection.
    i. Data selection using neural network.
    ii. Data selection using Decision Trees.
    iii. Data selection using Naive bayes.

iv. Data selection using Clustering, Regression, etc.

4. Data Transformation: Data Transformation is defined as the process of transforming data into appropriate form required by mining procedure.

   Data Transformation is a two-step process:

   i. Data Mapping: Assigning elements from source base to destination to capture transformations.
   ii. Code generation: Creation of the actual transformation program

5. Data Mining: Data mining is defined as clever techniques that are applied to extract patterns potentially useful.

   i. Transforms task relevant data into patterns.
   ii. Decides purpose of model using classification or characterization.

6. Pattern Evaluation: Pattern Evaluation is defined as as identifying strictly increasing patterns representing knowledge based on given measures.

   i. Find interestingness score of each pattern.
   ii. Uses summarization and Visualization to make data understandable by user.

7. Knowledge Representation: Knowledge representation is defined as technique which utilizes visualization tools to represent data mining results.

   i. Generate reports.
   ii. Generate tables.
   iii. Generate discriminant rules, classification rules, characterization rules, etc.

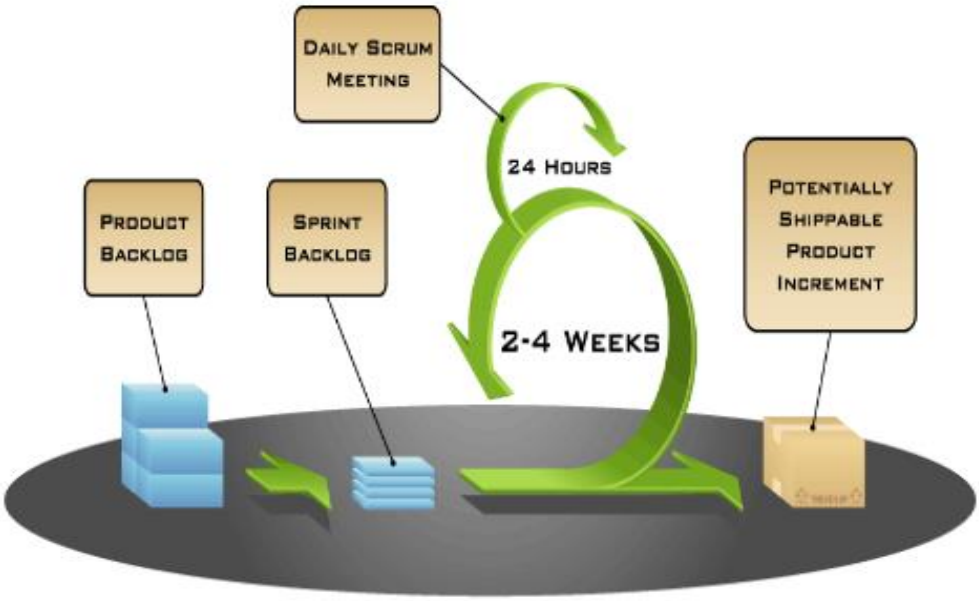


Fig.2.5.2: SCRUM

Scrum is an innovative approach to getting work done in efficient way. It is iterative & incremental agile software development method. These iterations are time boxed with various iterations & each iteration is called Sprint. The Sprint is basically 2-4 week long & each sprint requires sprint planning estimation. According to latest surveys Scrum is the most popular agile project management methodology in software development.

Scrum is ideally used where highly emergent or rapidly changing requirements. Scrum is basically worked on a self-organizing, cross-functional team. In the overall scrum team there is no team leader who assign the task to team rather whole scrum members work as a team & they decides the task on which they will work on. Also the problem will be resolve by team.

Each Agile Development Scrum team having three core scrum roles: Product Owner, Scrum Master & The Team.

1) Product Owner: The Product Owner is the person who represents the stakeholders and is the voice of the customer. Product owner writes the User Stories, ordered priorities and add in the

Product Backlog. It is recommended that Agile Scrum Master should not mix with Product Owner.

2) Scrum Master: The Scrum-Master is a facilitator, team leader who ensures that the team adheres to its chosen process and removes blocking issues to deliver the sprint deliverable/goal. Scrum Master is not a team leader but act as a shield for the team from external interference's & also removes barriers.

3) The Team: The scrum development team is generally size of 5-9 peoples with self-organizing and cross-functional skills who do actual work like Analysis, Design, Development, Testing, Documentation etc.

CHAPTER 3

ANALYSIS & PLANNING

3.1 Introduction

Getting a clear idea of the project title and doing research on it we will get our definition and after that then we will first create the Literature Survey of the project and do the whole documentation. After analysis, we will first study about it and do some research on it for our better understanding of the project and also get a rough picture about what would be our problem definition for the particular project.

3.2 Feasibility Study

1. Technical Feasibility - This assessment focuses on the technical resources available to the organization. It helps organizations determine whether the technical resources meet capacity and whether the technical team is capable of converting the ideas into working systems. It also involves the evaluation of the hardware, software, and other technical requirements of the proposed system.

2. Economic Feasibility - This assessment typically involves a cost/ benefits analysis of the project, helping organizations determine the viability, cost, and benefits associated with a project before financial resources are allocated. It also serves as an independent project assessment and enhances project credibility— helping decision makers determine the positive economic benefits to the organization that the proposed project will provide.

3. Legal Feasibility - This assessment investigates whether any aspect of the proposed project conflicts with legal requirements like zoning laws, data protection acts, or social media laws. Let's say an organization wants to construct a new office building in a specific location. A feasibility study might rev

eal the organization's ideal location isn't zoned for that type of business. That organization has just saved considerable time and effort by learning that their project was not feasible right from the beginning.

4. Operational Feasibility - This assessment involves undertaking a study to analyze and determine whether—and how well—the organization's needs can be met by completing the project. Operational feasibility studies also analyze how a project plan satisfies the

requirements identified in the requirements analysis phase of system development.

5. Scheduling Feasibility - This assessment is the most important for project success; after all, a project will fail if not completed on time. In scheduling feasibility, an organization estimates how much time the project will take to complete.

3.3 Project Planning

Phases of the Project:S

Fig.3.3.1: Phases of project development

1. **System Planning**

   The planning phase is the fundamental process of understanding why an information system should be built and determining how the project team will go about building it. It has two steps:

- During project initiation, the system's business value to the organization is identified how will it lower costs or increase revenues? Most ideas for new systems come from outside the IS area (from the marketing department, accounting department, etc.) in the form of a system request. A system request presents a brief summary of a business need, and it explains how a system that supports the need will create business value.

- Once the project is approved, it enters project management. During project management, the project manager creates a work plan, staffs the project, and puts techniques in place to help the project team control and direct the project through the entire SDLC.

2. **System Analysis**

   The analysis phase answers the questions of who will use the system, what the system will do, and where and when it will be used. During this phase, the project team investigates any current system(s), identifies improvement. This phase has three steps:

   1. An analysis strategy is developed to guide the project team's efforts. Such a strategy usually includes a study of the current system (called the as-is system) and its problems, and envisioning ways to design a new system (called the to-be system).

   2. The next step is requirements gathering (e.g., through interviews, group work-shops, or questionnaires). The analysis of this information—in conjunction with input from the project sponsor and many other people—leads to the development of a concept for a new system.
   3. The analyses, system concept, and models are combined into a document called the system proposal, which is presented to the project sponsor and other key decision makers (e.g., members of the approval committee) who will decide whether the project should continue to move forward.

3. **System Design**

   The design phase comes after a good understanding of customer's requirements, this phase defines the elements of a system, the components, the security level, modules, architecture and the different interfaces and type of data that goes through the system. The design phase decides how the system will operate in terms of the hardware, software, and network infrastructure that will be in place; the user interface, forms, and reports that will be used; and the specific programs, databases, and files that will be needed. The design phase has four steps:

1. The design strategy must be determined. This clarifies whether the system will be developed by the company's own programmers, whether its development will be outsourced to another firm (usually a consulting firm), or whether the company will buy an existing software package.
2. This leads to the development of the basic architecture design for the system that describes the hardware, software, and network infrastructure that will be used.
3. The database and file specifications are developed. These define exactly what data will be stored and where they will be stored.
4. The analyst team develops the program design, which defines the programs that need to be written and exactly what each program will do.

4. Implementation and Deployment

This phase comes after a complete understanding of system requirements and specifications, it’s the actual construction process after having a complete and illustrated design for the requested system.

In the Software Development Life Cycle, the actual code is written here, and if the system contains hardware, then the implementation phase will contain configuration and fine-tuning for the hardware to meet certain requirements and functions.

In this phase, the system is ready to be deployed and installed in customer’s premises, ready to become running, live and productive, training may be required for end users to make sure they know how to use the system and to get familiar with it, the implementation phase may take a long time and that depends on the complexity of the system and the solution it presents.

5. System Maintenance

In this phase, periodic maintenance for the system will be carried out to make sure that the system won’t become obsolete, this will include replacing the old hardware and continuously evaluating system’s performance, it also includes providing latest updates for certain components to make sure it meets the right standards and the latest technologies to face current security threats.

These are the main six phases of the System Development Life Cycle, and it's an iterative process for each project.

3.4 Scheduling

Gantt chart:

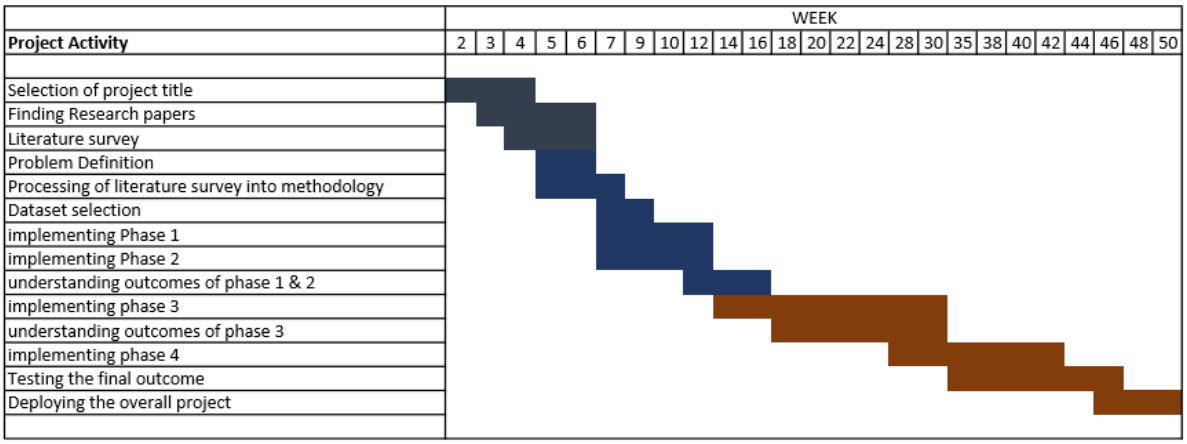


Fig.3.4.1: Gnatt Chart

It shows the representation of each day of the work from the Selection of Project Title to its Testing.

In this after understanding the topic the project feasibility was analyzed by performing different types of feasibility studies and by also planning the project tools, their project schedule, timeline charts, etc. Feasibility study will help in better understanding the various feasibilities associated with the project and helping to make the correct decisions and completing the project within the schedule, budget, etc.

The tools were specifically identified in this chapter stating which technology can be feasible and how conveniently the project can be completed. This helps to understand the technology and tools that can be used for the project. The Gantt chart helps us to track the project and see the schedule of the project and to see if the project is on the right track and on schedule and not behind the deadline

CHAPTER 4

DESIGN & IMPLEMENTATION

4.1 Model Requirements

The Model requirements for our project are as follows:

i. **Dataset:** Large data sets are first sorted, then patterns are identified and relationships are established to perform data analysis and solve problems.

ii. **Anaconda:** Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing, that aims to simplify package management and deployment. Package versions are managed by the package management system conda.

iii. **Algorithms:** In today's world of "big data", a large database is becoming a norm. Just imagine a database with many terabytes. As facebook alone crunches 600 terabytes of new data every single day. Also, the primary challenge of big data is how to make sense of it. Moreover, the sheer volume is not the only problem. Also, big data need to diverse, unstructured and fast changing. Consider audio and video data, social media posts, 3D data or geospatial data. This kind of data is not easily categorized or organized. Further, to meet this challenge, a range of automatic methods for extracting information.

4.2 Block Diagram

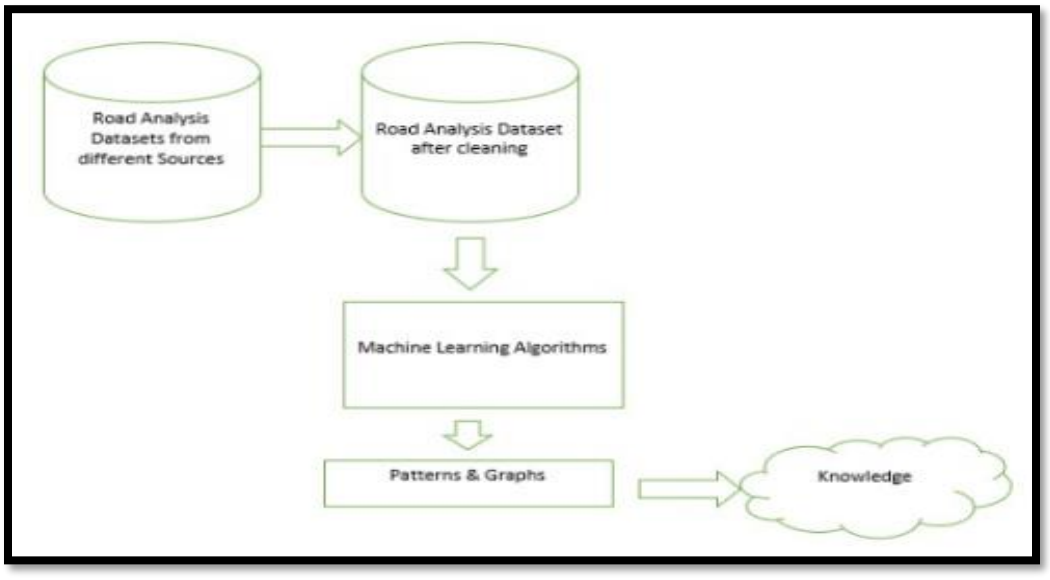


Fig.: Block diagram

4.3 Flow Chart

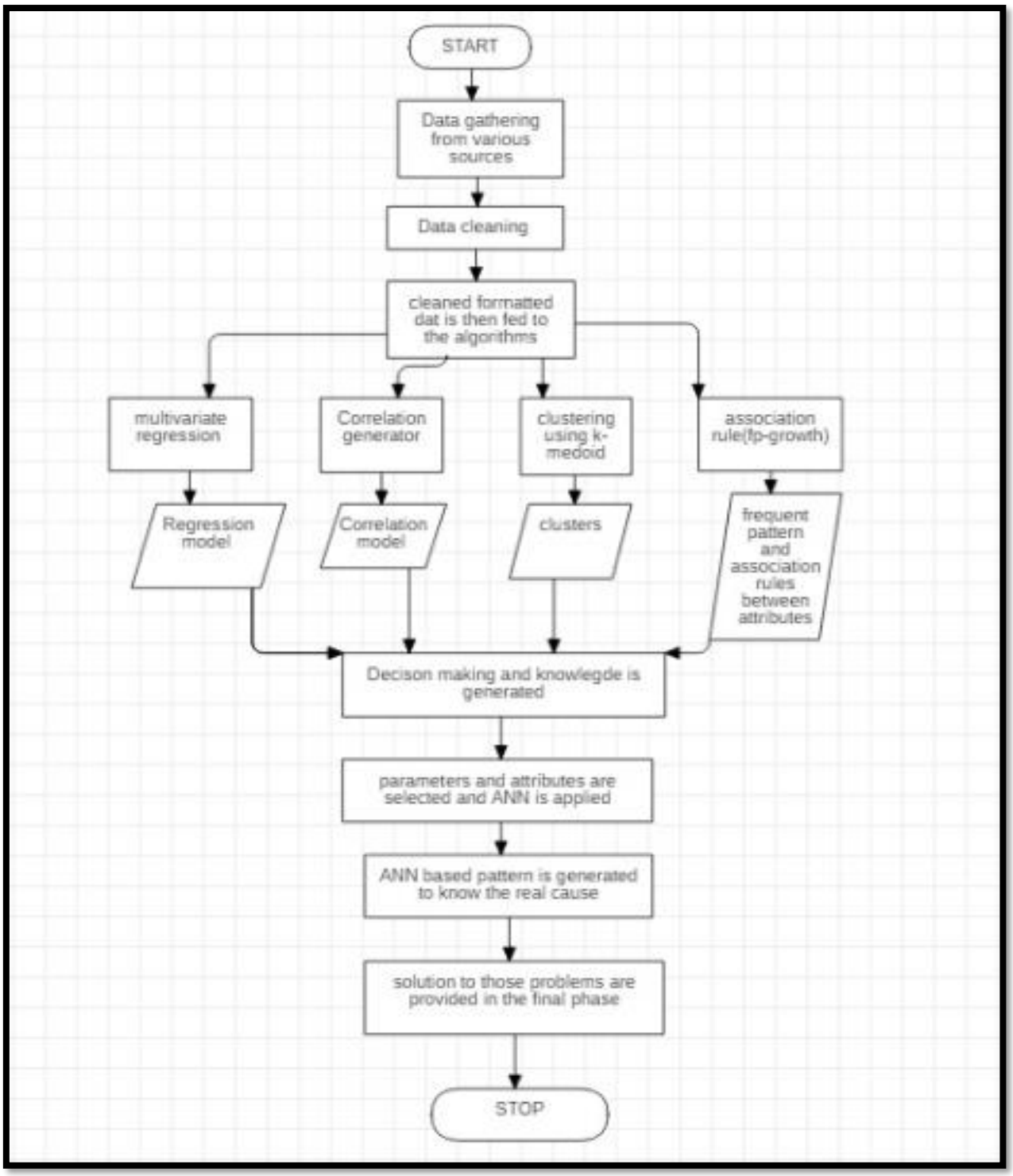
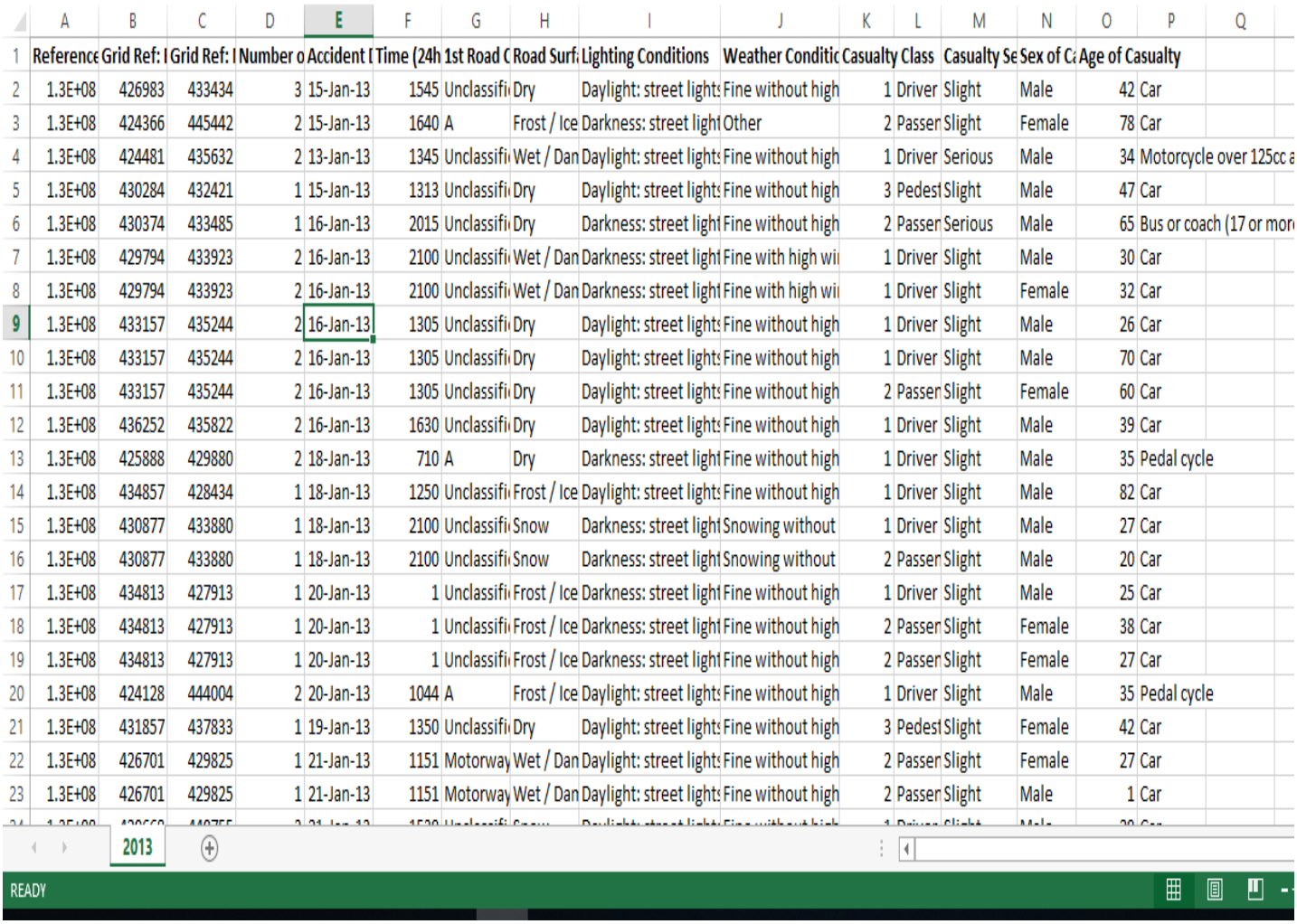


Fig.4.3.1: flowchart

4.4 Database Screenshot



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Reference	Grid Ref: I	Grid Ref: I	Number o	Accident I	Time (24h	1st Road C	Road Surfi	Lighting Conditions	Weather Conditic	Casualty	Class	Casualty Se	Sex of Ca	Age of Casualty		
2	1.3E+08	426983	433434	3	15-Jan-13	1545	Unclassifi	Dry	Daylight: street lights	Fine without high	1	Driver	Slight	Male	42	Car	
3	1.3E+08	424366	445442	2	15-Jan-13	1640	A	Frost / Ice	Darkness: street light	Other	2	Passen	Slight	Female	78	Car	
4	1.3E+08	424481	435632	2	13-Jan-13	1345	Unclassifi	Wet / Dan	Daylight: street lights	Fine without high	1	Driver	Serious	Male	34	Motorcycle over 125cc a	
5	1.3E+08	430284	432421	1	15-Jan-13	1313	Unclassifi	Dry	Daylight: street lights	Fine without high	3	Pedest	Slight	Male	47	Car	
6	1.3E+08	430374	433485	1	16-Jan-13	2015	Unclassifi	Dry	Darkness: street light	Fine without high	2	Passen	Serious	Male	65	Bus or coach (17 or mor	
7	1.3E+08	429794	433923	2	16-Jan-13	2100	Unclassifi	Wet / Dan	Darkness: street light	Fine with high wi	1	Driver	Slight	Male	30	Car	
8	1.3E+08	429794	433923	2	16-Jan-13	2100	Unclassifi	Wet / Dan	Darkness: street light	Fine with high wi	1	Driver	Slight	Female	32	Car	
9	1.3E+08	433157	435244	2	16-Jan-13	1305	Unclassifi	Dry	Daylight: street lights	Fine without high	1	Driver	Slight	Male	26	Car	
10	1.3E+08	433157	435244	2	16-Jan-13	1305	Unclassifi	Dry	Daylight: street lights	Fine without high	1	Driver	Slight	Male	70	Car	
11	1.3E+08	433157	435244	2	16-Jan-13	1305	Unclassifi	Dry	Daylight: street lights	Fine without high	2	Passen	Slight	Female	60	Car	
12	1.3E+08	436252	435822	2	16-Jan-13	1630	Unclassifi	Dry	Daylight: street lights	Fine without high	1	Driver	Slight	Male	39	Car	
13	1.3E+08	425888	429880	2	18-Jan-13	710	A	Dry	Darkness: street light	Fine without high	1	Driver	Slight	Male	35	Pedal cycle	
14	1.3E+08	434857	428434	1	18-Jan-13	1250	Unclassifi	Frost / Ice	Daylight: street lights	Fine without high	1	Driver	Slight	Male	82	Car	
15	1.3E+08	430877	433880	1	18-Jan-13	2100	Unclassifi	Snow	Darkness: street light	Snowing without	1	Driver	Slight	Male	27	Car	
16	1.3E+08	430877	433880	1	18-Jan-13	2100	Unclassifi	Snow	Darkness: street light	Snowing without	2	Passen	Slight	Male	20	Car	
17	1.3E+08	434813	427913	1	20-Jan-13	1	Unclassifi	Frost / Ice	Darkness: street light	Fine without high	1	Driver	Slight	Male	25	Car	
18	1.3E+08	434813	427913	1	20-Jan-13	1	Unclassifi	Frost / Ice	Darkness: street light	Fine without high	2	Passen	Slight	Female	38	Car	
19	1.3E+08	434813	427913	1	20-Jan-13	1	Unclassifi	Frost / Ice	Darkness: street light	Fine without high	2	Passen	Slight	Female	27	Car	
20	1.3E+08	424128	444004	2	20-Jan-13	1044	A	Frost / Ice	Daylight: street lights	Fine without high	1	Driver	Slight	Male	35	Pedal cycle	
21	1.3E+08	431857	437833	1	19-Jan-13	1350	Unclassifi	Dry	Daylight: street lights	Fine without high	3	Pedest	Slight	Female	42	Car	
22	1.3E+08	426701	429825	1	21-Jan-13	1151	Motorway	Wet / Dan	Daylight: street lights	Fine without high	2	Passen	Slight	Female	27	Car	
23	1.3E+08	426701	429825	1	21-Jan-13	1151	Motorway	Wet / Dan	Daylight: street lights	Fine without high	2	Passen	Slight	Male	1	Car	

Fig.: Database Screenshot

Dataset size: 2420*14

Attributes: 14

CHAPTER 5

RESULTS & DISCUSSION

5.1 Results

The study performed here consists of parameters analysis their associations and their correlation with each other in-order to reach out the major factors that are the cause of the road accident and find the appropriate solution in order to avoid it and incase unavoidable then how to provide maximum protection with minimum loss and damage.

5.2 Future Scope

The current scope of the project is up-to the analysis phase i.e. analysis can be at a better level of Machine learning and we have a future scope of the project are:

1. Neural Networks in Forensic Science.
2. Use of image processing in analyzing the accident has happen and giving the information of it to the nearby hospital and police station.
3. Using IOT based system to protect the car from the crash etc.

CHAPTER 6

CONCLUSION

6. Conclusion

The aim is to generate association rules that will analyze how to discover hidden patterns that are the root causes for accidents among different combinations of attributes of a larger dataset. Density histograms for visualizing Region wise such as fatal versus weather, fatal versus time, time versus day, fatal versus month, fatal versus traffic, and fatal versus age are performed. Percentage distribution of accidents on various criteria, speed limit and injury severity, distribution of accidents by time of accidents and deceased age, distribution of accidents by month and weather during the accident, distribution of accidents by lightness and speed limit, distribution of accidents by accident type (human factors), distribution of accidents by day of accident and deceased age, distribution of accidents by deceased emotions, distribution of accidents by hospital reported and ambulance used is also made. Future work is to make analysis on road accidents' dataset by considering more features and clusters and also to use deep learning techniques so as to better cluster the records.

References

[1] S.Vasavi, “Extracting Hidden Patterns Within Road Accident Data Using Machine Learning Techniques”,2018.

[2] Sannaila Praveen Kumar, M. Prashanti,”Road Accident Data Analysis Using Machine Learning”, 2018.

[3] Shrishti Sonal, Saumya Suman,” A framework for analysis of road accidents”, 2018.

[4] Poojitha Shetty, Sachin P C, Supreeth V Kashyap, Venkatesh Madi, “Analysis of Road Accidents using data mining techniques”, in International Research Journal of Engineering and Technology(IRJET), 2017.

[5] Ms. Gagandeep Kaur, Er. Harpreet kaur, Prediction of the cause of accident and accident prone location on roads using data mining techniques, 2017.

[6] Priyanka A. Nandurge, Nagraj V. Dharwadhkar, Analyzing road accident data using machine learning paradigms, 2017.

[7] Miao Chong, Ajith Abraham and Marcin Paprzycki , Traffic accident analysis using machine learning paradigms, 2005.

[8] Road safety and traffic management: Report of the committee Planning Commission, Government of India in February 2007.

[9] Rayle L, Pai M. Scenarios for future urbanization: carbon dioxide emissions from passenger travel in three Indian cities. Transportation Research Record: Journal of the Transportation Research Board, 2193:124–131 (2010).

[10] http://www.deccanchronicle.com/130629/news-current-affairs/article/andhra-pradesh-ranked- 3rd-road-accidents last accessed June 29th 2013. [online]

[11] https://towardsdatascience.com/train-your-neurons-on-artificial-neural-networks-featuring-keras-2757f51bb7ed. [online]

[12] https://www.geeksforgeeks.org/data-mining-kdd-process/ [online]