IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138
INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG
An International Open Access, Peer-reviewed, Refereed Journal

ROAD ACCIDENT PREDICTION USING ARTIFICIAL INTELLIGENCE

Submitted by
SRIVIDHYA K, PAVANI G, VARSHITHA K
COMPUTER SCIENCE AND ENGINEERING
ADHI COLLEGE OF ENGINEERING AND TECHNOLOG,SANKARAPURAM.
ANNA UNIVERSITY: CHENNAI 600 025

ABSTRACT:

The traffic has been transformed into the difficult structure in points of designing and managing by the reason of increasing number of vehicles. This situation has discovered road accidents problem, influenced publichealth and country economy, and done the studies on solution of the problem. Large calibrated data agglomerations have increased by the reasons of the technological improvements and data storage with low cost.Arising the need of accession to information from this large calibrated data obtained the corner stone of the data mining. In this study, assignment of the most compatible machine learning classification techniques for road accidents estimation by data mining has been intended.

There are several problems with current practices for prevention of the accidents occurred in the localities.The database we will use is available officially by many institutes and government websites. The data collected will be analyzed, integrated, and grouped together based on different constraints using the best suited algorithm.This estimation will be helpful for collecting data and it shows the risk in a particular area is high or low.
We have made use of data mining techniques in developing an accident prediction model using logistic regression algorithm and gradient boosting algorithm. Bangalore road accident datasets for the years 2014 to 2017available in the internet have been made use for this study.

The results from this study can be advantageously used by several stakeholders including and not limitedto the government public work departments, contractors and other automobile industries in better designing roadsand vehicles based on the estimation will be helpful for collecting data and it shows the risk in a particular area ishigh or low.Chapter 1

PREAMBLE

1.1 Introduction

Road accidents are unquestionably the most frequent cause of damage. It's one of the most significant causes of the fatalities. The reasons for this are the extremely denseroad traffic and the relatively great freedom of movement given to drivers. Accidents that involve heavy goods vehicles (like Lorries, trucks) and even the commercial vehicles with the public transportations like buses are one of the most fatal kinds of accidents that occur, claiming the lives of innocent people.

Highways are always a soft spot for these accidents with injuries and deaths. Various weather conditions like rain, fog etc. play a role in catalysing the risk of accidents. Havinga proper estimate of accidents and knowing the hotspots of accidents and its factors will help to reduce them. Providing timely emergency support even when the casualties have occurred is needed, and to do that a keen study on accidents is required.

In spite of having set regulations and the highway codes, negligence of people towardsthe speed of the vehicle, the vehicle condition and their own Negligence of not wearing helmets has caused a lot of accidents. These accidents wouldn't have turned fatal, and claimed innocent lives if people had governed by the rules. Prevention of roadaccidents is significantly important and will be fortified by strict laws, by technical and police controls, tougher training for drivers to issue the driving license and creatinga sense of awareness among people as to how important it is to take these rules seriouslyby imposing penalties and legalities for people responsible.

The number of deaths due to road accidents in India is indeed a cause for worry. The scenario is very dismal with more than 137,000 people succumbing to injuries from road accidents. This figure is more than four times the annual death toll from terrorism. Reports show that a person dies in a road accident every four minutes.Existing System:

Bad Roads: India is said to be the fastest developing country after China. The nation is performing extremely well in fields such as education, industrialization and fashion. However, the conditions of the roads be it the metropolitans, towns or villages, act as a major pothole in India's complete progress. According to the Road Accident Report for 2014 prepared by the road transport and highways ministry, 75,000 people have been killed because of the killer roads of India.

Speeding, Helmets and overloaded vehicles: Another major factor contributing tothe increased number of road accidents is speeding, not wearing helmets and over loaded vehicles. The public fails to follow the speed limits, especially on the highway. This has resulted in 41% of the total deaths due to road accidents in India in 2014.Wearing a helmet can reduce the risk of severe injury by 72% and the risk of death by 39%, according to the WHO.

1.1.1 Drawbacks

What the victim suffers- Road accident victims pay for the accident with life and or financially. Disability for life & a full stop to future income are some fall outs.

What the family suffers- If the family losses an earning member of the family, they suffer emotionally as well as financially.

Problem for commuters- Accidents lead to heavy jams and the reason is very silly. People stop or slow down to witness the accident site! Many times they do not help the victim, makeno effort to contact the family but just stand there to witness an accident. Most of the times, it is not the accident but these passers-by who cause the accident.

Loss to the country- Based on data, 34% of all the people who lose life are in the age group18–34 yrs. This is the most productive generation. It is estimated that road accident cost 3%of India's GDP.

1.2 Proposed System

1.Our aim is to develop an application that can be used to predict the risk of accidents.

2.The data sets are collected from various websites such as Kaggle, data towards scienceetc.…

3.The next step is to clean the data and transform it into the desired format by featureselection method.

1.The data sets are trained and tested to produce the sophisticated results and henceforthapply the suitable algorithms .

2.The evaluation and interpretation of results is performed .

1.3 Plan of Implementation

1.The project can be broken down into 4 Major states.

2.The first stage of project is to take the support and confidence values from the user andprovide the rules to the driver.

3.The second stage is to plot graphs based on the input selected by the user.

4.The third stage is to predict the risk of accidents in a particular area chosen by the user.

5.The Final stage is to report new accidents encountered by the people which will be usedfor the collection of data sets in future.

1.4 Problem Statement:

There are several problems with current practices for prevention of the accidents occurred in the localities. The database we will use is available officially by many institutes and government websites. The data collected will be analyzed, integrated and grouped together based on different constraints using the best suited algorithm. This estimation will be helpfulto analyze and identify the flaw and the reasons of the accidents.

It will also be helpful while making roads and bridges as a reference to avoid the same problems faced before. The predictions made will be very much useful to plan the management of such problems.

1.5 Objective of the system:

Providing information to the people regarding the factors causing the accident will act as guidance to the travelers.

Government authorities will know the causes of accidents, the dominant factors such as weather conditions, transport infrastructure etc. that are inflicting the various accident prone regions and to provide assistance in sketching out the association between the various factors that directly or indirectly have a part in causing the accident. Information is provided to the Regional Transport Office regarding the accident prone regions where

alcoholic intoxication, distracting circumstances like making a phone call while driving, aggressive and careless driving, and disregard safety rules, fatigue of driver is the main cause of accidents. This information will help the RTO to impose strict actions such as checking the license of the driver, conducting alcohol check or always placing a traffic police in such areas. Our aim is also to assist them in organizing the traffic.

Chapter 2

LITERATURE SURVEY

Road safety becomes a major public health concern when the statistics show that more than 3,000 people around the world succumb to death daily due to road traffic injury. In addition, road crashes lead to the global economic losses as estimated in road traffic injury costs ofUS$518 billion per year.

In order to get required knowledge about various concepts related to the present application, existing literature were studied. Some of the important conclusions were made through thoseare listed below

An Asian Development Bank country report focused on the seriousness of the road accident problem which is shown in Figure 3.1 with an upward trend of injuries per accident whereas fatalities per accident remained constant with small fluctuations from 1993-2002.

Road safety is very important because the number of fatal accidents recorded in India is very high.People do not follow traffic rules and endanger public safety by zigzag driving, jumping red lights,parking in public spaces, drunken driving and not showing concern for pedestrians. Most offenders get away by just paying a fine A life each 3.5 minutes is lost dueto road crash in India at present. Around 3% GDP loss to nation is estimated due to the same.

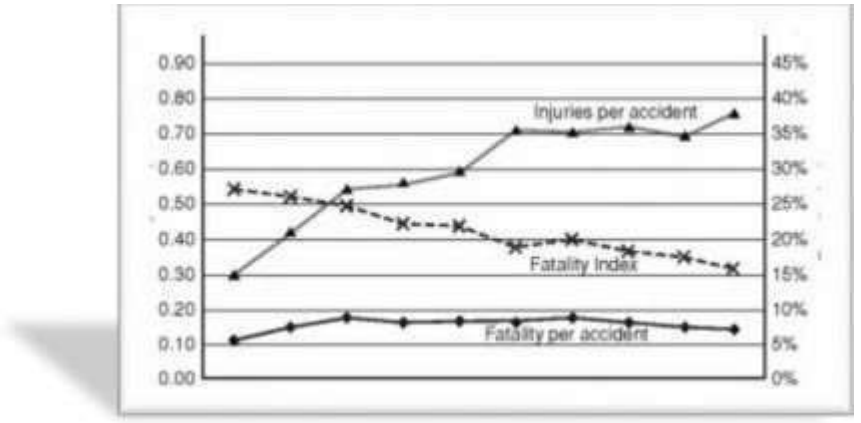


Figure 1 Trends in casualties per accident and fatality index

Ayushi Jain, Garima Ahuja, Anuranjana, Deepti Mehrotra[1]were presented analysis of road accidents based on Data Mining technique.

Author has made use of dataset which is selected from data.gov.in.The data includes from allthe union territories and states and analyzed for 58 attributes like total number of accidents, number of people killed and number of people injured due to various factors like alcohol, speeding, driver's fault, type of vehicles etc. [3]. This paper makes use of K-Means Clustering to group similar object off of the heterogeneous data (four clusters are formed by writers to analyze the road accident).

Ms. Gagandeep Kaur predicted accident prone locations on roads by using DataMiningTechniques .

This paper sheds light on predicting the probability of accidents on roads with special emphasis on State Highways and Ordinary District Roads by estimating the severity of accidents based on the type of accident, type of spot using the R tool.Simulation is performedby using R-Studio which is an Integrated Development Environment (IDE) for R tool. Correlation analysis and exploratory visualization techniques has been appliedFrom this paper we can substantiate that Correlation analysis examines the road conditions that help to derive the relation between two numerical variables i.e. length and progress of road which is negative that shows the inverse relation. From the study of Exploratory visualization techniques, we can justify that accident on State Highways occur on Straight roads and on Ordinary District Roads the accidents can occur on other type of spots such as Cross- intersection, R-intersection and Straight road but majorly on Cross-intersection.Mainly the type of accident that occurs is Head on collision type on both Roads.

Authors Irina Makarova KseniaShubenkova, Eduard and Mukhametdinov were highlighted the importance of road quality and infrastructure

The article provides an analysis of the global trends in the field of city transport systems' safety. It is shown that the decrease in the safety of traffic is one of the consequences of the growth of motorization. The efficiency of measures to prevent traffic accidents is analyzed from the viewpoint of their role in the process to ensure safety and sustainability of the urbantransportation system. In this article authors have illustrated the application of Haddon matrixto improve Road Safety.

Chapter 3

THEORITICAL BACKGROUND

Road accident data analysis is mainly done based on two categories: statistical techniques anddata mining techniques. Data mining techniques have certain advantages over traditional statistical techniques. Data mining techniques do not require certain assumptions between dependent and independent variables which are required in traditional statistical techniques. Also, data mining techniques is capable of handling large dimensional data whereas statisticaltechniques have some limitations.

Data collection

To create a successful machine learning model, it is imperative that an organization has the ability to train, test, and validate them prior to deploying into production.

Websites

Kaggle,UCI ,Data towards science

3.1 Feature Selection Techniques

3.1.1 Data Set Selection

Data is the most import part when you work on prediction systems. It plays a very vital role your whole project i.e., you system depends on that data. So selection of data is the first and the critical step which should be performed properly, For our project we got the data from thegovernment website. These data sets were available for all. There are other tons of websites who provide such data. The data set we choose was selected based on the various factors andconstraints we were going to take under the consideration for our prediction system.

3.1.2 Data Cleaning and Data Transformation

After we have selected the data set. The next step is to clean the data and transform it into thedesired format as it is possible the data set we use may be of different format. It is also possiblethat we may use multiple data sets from different sources which may be in differentfile formats. So to use them we need to convert them into the format we want to or the type that type prediction system supports. The reason behind this step is that it is possible that thedata set contains the constraints which are not needed by the prediction system and including them makes the system complicated and may extend the processing time. Another reason behind data cleaning is the data set may contain null value and garbage values too.So the solution to this issue is when the data is transformed the garbage values are replaced. There are many methods to perform that.

3.1.3 Data Processing

After the data has been cleaned and transformed it is ready to process further. After the data has been cleaned and we have taken the required constraints. We divide the whole data set into the two parts that can be either 70-30 or 80-20. The larger portion of the data is for the processing.The data set obtained will now be subjected to various data mining techniques.

Data Preprocessing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

Importing the required LibrariesImporting the Dataset

Handling the Missing DataEncoding categorical data

Splitting the Data set into Training set and Test SetFeature Scaling

Training:

► The idea of using training data in machine learning programs is a simple concept,but it is also very foundational to the way that these technologies work. The training

data is an initial set of data used to help a program understand how to applytechnologies like neural networks to learn and produce sophisticated results.

Testing:

► A test dataset is a dataset that is independent of the training dataset, but that follows the

same probability distribution as the training dataset. If a model fit to the trainingdataset also fits the test dataset well, minimal overfitting has taken place.

3.1.4Data flow diagrams

A data-flow diagram (DFD) is a way of representing a flow of a data of a process or a system(usually an information system). The DFD also provides information about the outputs and inputs of each entity and the process itself.

DFD level-0

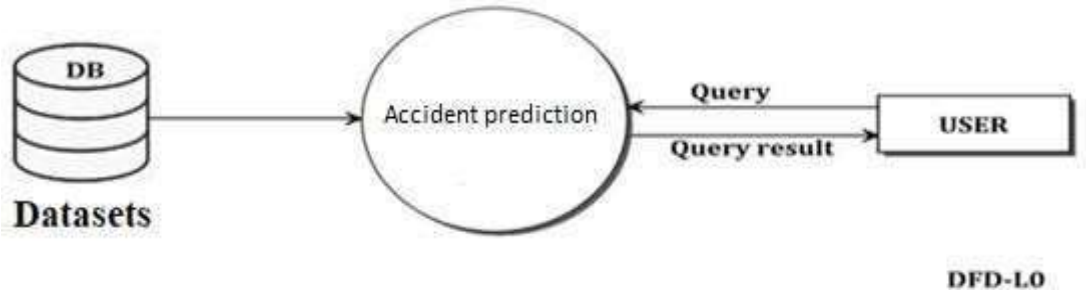


Figure 2 : Dfd level 0

DFD level -1

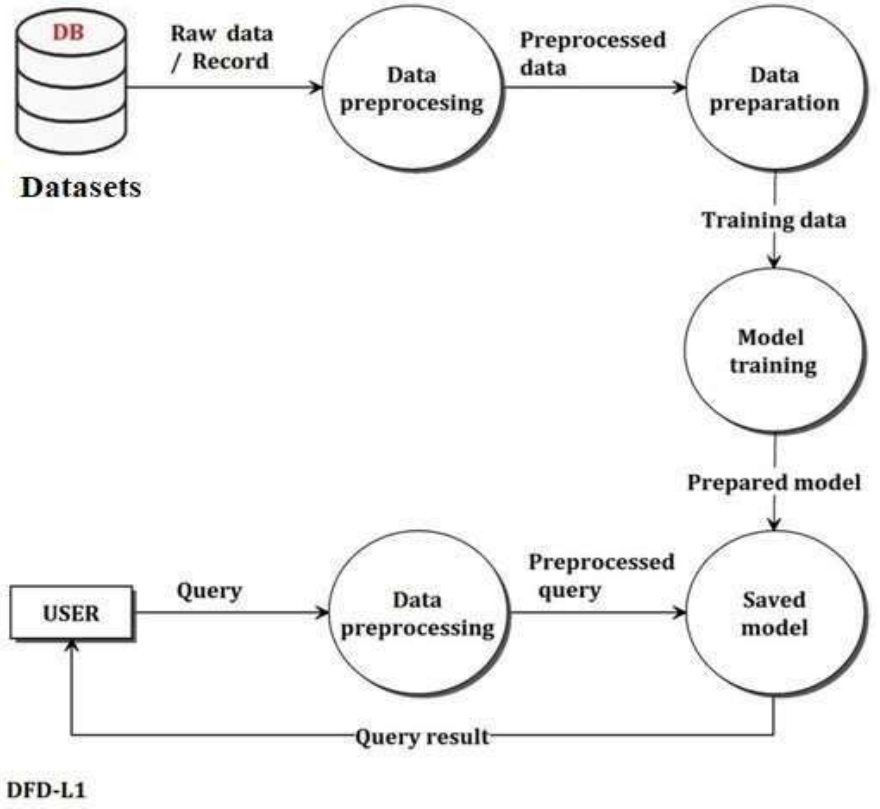


Figure 3: Fdf level 1

DFD level -3

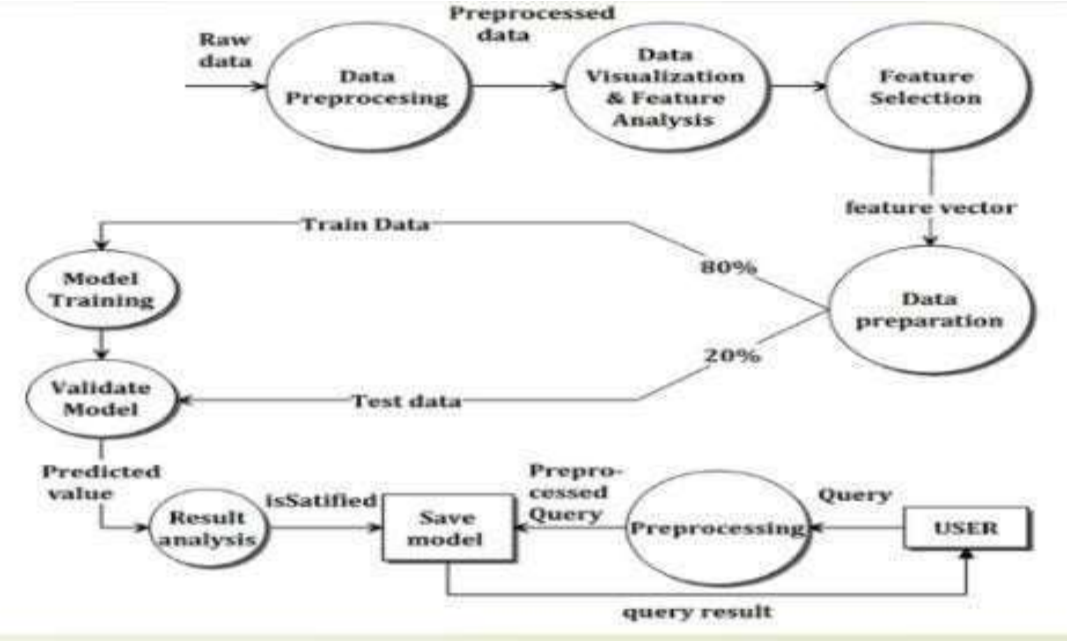


Fig:4 Fdf level 2

3.2 Clustering

Clustering will be performed on the given data set. The main aim of performing clustering isto divide the data into different clusters or groups such that the objects within a group are similar to each other whereas objects in other clusters are different from each other. There are several clustering algorithmsavailable: Hierarchical clustering technique like Ward method, single linkage, complete linkage etc, K means and latest class clustering (LCC).Other clustering algorithms like K-modes clustering is an enhanced version of K means clustering.The clusters are then subjected to other algorithms like Association rule mining and trend analysis.

3.2.1Association rule mining

Association rule mining is a very popular data mining technique that extracts interesting andhidden relations between various attributes in a large data set. Association rule mining produces a set of rulesthat define underlying patterns in the data set. The associativity is known by the frequency of their occurrence together in the data set.

3.2.2Trend analysis

The trend analysis is performed to determine the upcoming trends based on the total countof accidents for each cluster. The trends analysis can also be applied on the entire dataset as well as the clusters. The trends can show a positive or negative trend for the future based on the current and past trends. There can be difference in trends for various clusters as there mightbe different factors dominant in causing accidents for that particular cluster. This trend will help us analyse the extent to which the measures taken across the years to reduce accidents has contributed in reducing the accident rate.

3.2.3Simulation

The simulation is performed by using R tools. Various data mining techniques and exploratoryvisualization techniques is applied on the accident dataset to get interpreted results. The R tools help to develop and interactive user interface. Thus we can analyse the various factors contributing for the accidents by plotting various graphs, charts and other statistical and graphical representations.

These analysis and research helps in providing solutions in order to reduce the accident rate and decrease the fatality in the number of deaths occurring due to these accidents as this analysis will help in understanding the overall causes of accidents, the degree to which they play a role in the accidents and how they can be reduced.

3.2.4Time series merging:

Cluster analysis results in homogeneous segments of the time series data. Each cluster consists of various time series objects that are similar in nature. Hence there is a need to forma representative time series which can represent the entire time series. To find the time series that can represent the entire cluster, a time series merging algorithm is formed, that takes DTW (Dynamic time warping is type of similarity measure for time series data which can measure distance between two-time series objects even if their length is not similar) distanceto calculate the closest

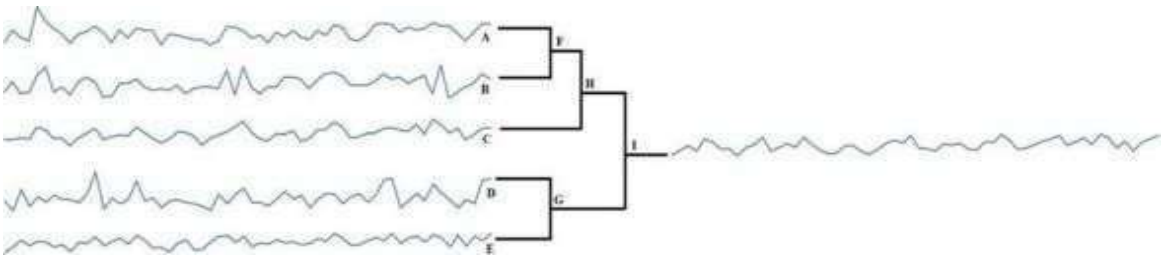


Figure 5: Time series merging

3.3 Algorithms

3.3.1Logistic Regression algorithm:

Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of certain classes based on some dependent variables.In short, the logistic regression model computes a sum of the input features and calculates the logistic of the result.

3.3.2Advantages of Logistic Regression algorithm:

The main advantage of logistic regression is that it is much easier to set up and train than other machine learning and AI applications. Another advantage is that itis one of the most efficient algorithms when the different outcomes or distinctionsrepresented by the data are linearly separable

3.4 Gradient Boosting algorithm:

Gradient Boosting is a functional gradient algorithm that repeatedly selects a function that leads in the direction of a weak hypothesis or negative gradient so that it can minimize a loss function. Gradient boosting classifier combines several weak learning models to produce apowerful predicting model.

3.4.1Advantages of Gradient Boosting algorithm:

- It is Lots of flexibility and can optimize on different loss functions.

Chapter 4

SYSTEM REQUIREMENT SPECIFICATION

A software requirements specification (SRS) is a description of a software system to be developed.It lays out functional and nonfunctional requirements, and may include a set of use cases that describe user interactions that the software must provide. In order to fully understand ones project,it is very important that they come up with an SRS listing out their requirements, how are they going to meet it and how will they complete the project. SRS also functions as a blueprint for completing a project with as little cost growth as possible. SRS is often referred to as the parent document because all subsequent project management documents,such as design specifications, statements of work, software architecture specifications,testing and validation plans, and documentation plans, are related to it. Requirement is a condition or capability to which the systemmust conform. Requirement Management is a systematic approach towards eliciting, organizing and documenting the requirements of the system clearly along with the applicable attributes. The elusive difficulties of requirements are not always obvious and can come from any number of source.

1.1 Functional Requirements

Functional Requirement defines a function of a software system and how the system must behave when presented with specific inputs or conditions. These may include calculations, data manipulation and processing and other specific functionality. Following are the functional requirements on the system:

- Collecting Date sets and data pre-processing is performed for that data set
- The date set will be subjected to various date mining techniques ,Clustering willbe performed on the given data set
- The clusters are then subjected to other algorithms like Association rule miningand trend analysis.

1.2 Non Functional Requirements

Non functional requirements are the requirements which are not directly concerned withthe specific function delivered by the system. They specify the criteria that can be use to judge.

the operation of a system rather than specific behaviours. They may relate to emergent systemproperties such as reliability, response time and store occupancy. Non functional requirementsarise through the user needs, because of budget constraints, organizational policies and the need for interoperability with other software and hardware systems.

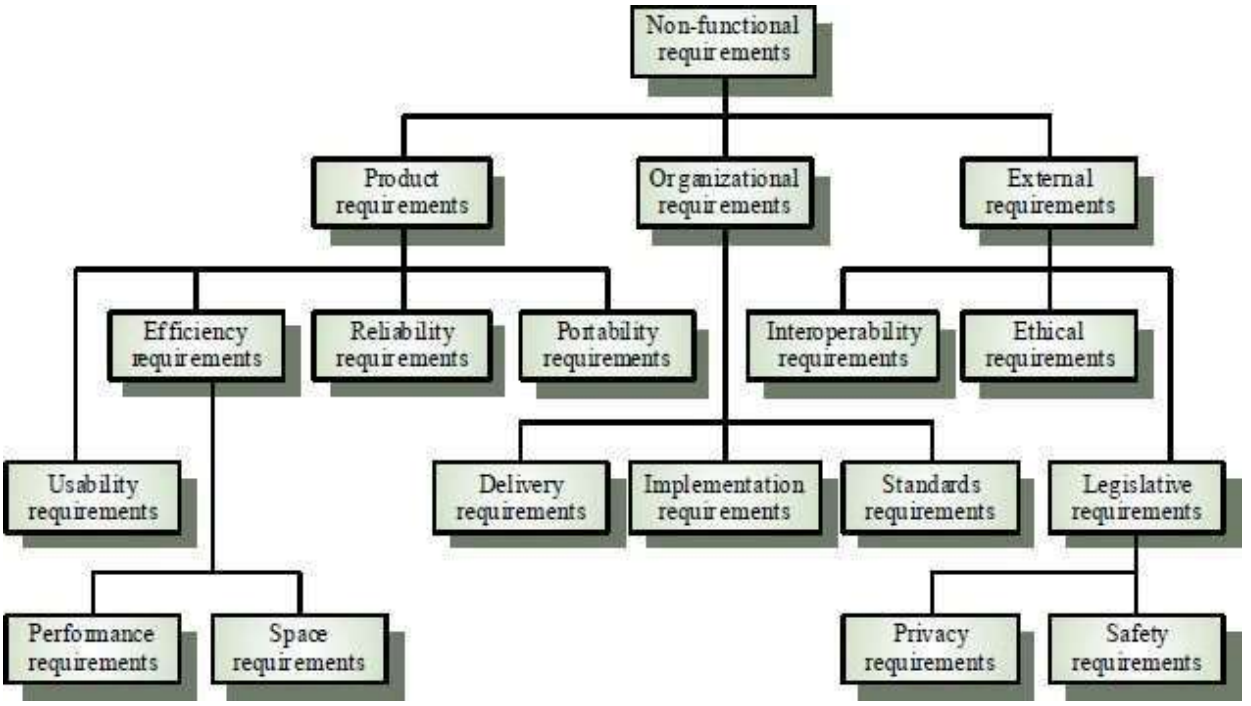


Figure 6 Non functional requirements

Some Non-Functional Requirements are as follows:

• Reliability

The structure must be reliable and strong in giving the functionalities. The movementsmust be made unmistakable by the structure when a customer has revealed a couple of enhancements. The progressions made by the Programmer mustbe Project pioneerand in addition the Test designer.

• Maintainability

The system watching and upkeep should be fundamental and focus in its approach. There should not be an excess of occupations running on diverse machines such that it gets hard to screen whether the employments are running without lapses.

• Performance

The framework will be utilized by numerous representatives all the while. Since the system will be encouraged on a single web server with a lone database server outside

of anyone's ability to see, execution transforms into a significant concern. Thestructureshould not capitulate when various customers would use everything the while. It should allow brisk accessibility to each and every piece of its customers. For instance,if two test specialists are all the while attempting to report the vicinity of a bug, then there ought not to be any irregularity at the same time.

• Portability

The framework should to be effectively versatile to another framework. This is obligedwhen the web server, which s facilitating the framework gets adhered because of a fewissues, which requires the framework to be taken to anotherframework.

• Scalability

The framework should be sufficiently adaptable to include new functionalities at a later stage. There should be a run of the mill channel,

which can oblige the new functionalities.

- **Flexibility**

Flexibility is the capacity of a framework to adjust to changing situations and circumstances, and to adapt to changes to business approaches and rules. An adaptableframework is one that is anything but difficult to reconfigure or adjust because of diverse client and framework prerequisites. The deliberate division of concerns between the trough and motor parts helps adaptability as just a little bit of the framework is influenced when strategies or principles change.

4.2.1Product Requirements

Correctness: It followed a well-defined set of procedures and rules to engage a conversation with the user and a pre-trained classification model to compute also rigorous testing is performed to confirm the correctness of the data.

Modularity: The complete product is broken up into many modules and well definedinterfaces are developed to explore the benefit of flexibility of the product.

Robustness: This software is being developed in such a way that the overall performance is optimized and the user can expect the results within a limited time with utmost relevance andcorrectness. Non functional requirements are also called the qualities of a system.

4.2.2Basic Operational Requirements:

The customers are those that perform the eight primary functions of systems engineering,with special emphasis on the operator as the key customer. Operational requirements will define the basic need and, at a minimum, will be related to these following points:-

Mission profile or scenario : It describes about the procedures used to accomplish missionobjective. It also finds out the effectiveness or efficiency of the system.

Performance and related parameters : It points out the critical system parameters toaccomplish the mission.

Utilization environments : It gives a brief outline of system usage. Finds out appropriateenvironments for effective system operation.

Operational life cycle : It defines the system lifetime.

1.3Hardware system configuration

- Processor : Pentium Processor and Above
- Ram : 4GB
- Hard disk capacity: 500 GB

1.4Software system configuration

- Operating system : Linux, Windows 7,8,10
- Programming Language: Python
- Framework : Anaconda
- IDE: Spyder
- DL Libraries : NumPy, Pandas

Chapter 5

SYSTEM DESIGN

5.1 System Development Methodology

System Development methodology is the development of a system or method for unique situation. Having a proper methodology helps us in bridging the gap between the problem statement and turning it into a feasible solution. It is usually marked by converting the System Requirements Specifications (SRS) into a real world solution.

System design takes the following inputs: Statement of work. Requirement determination plan.Current situation analysis. Proposed system requirements including a conceptual data model andmetadata (data about data).The development method followed in this project is waterfall model.

5.1.1 Model Phases

The waterfall model is a sequential software development process, in which progress is seen asflowing steadily downwards (like a waterfall) through the phases of Requirement initiation, Analysis, Design, Implementation, Testing and maintenance.

Requirement Analysis: This phase is concerned about collection of requirement of thesystem.This process involves generating document and requirement review.

System Design: Keeping the requirements in mind the system specifications are translated in to asoftware representation. In this phase the designer emphasizes on:algorithm, data structure, software architecture etc.

Coding: In this phase programmer starts his coding in order to give a full sketch of product. Inother words system specifications are only converted in to machine readable compute code

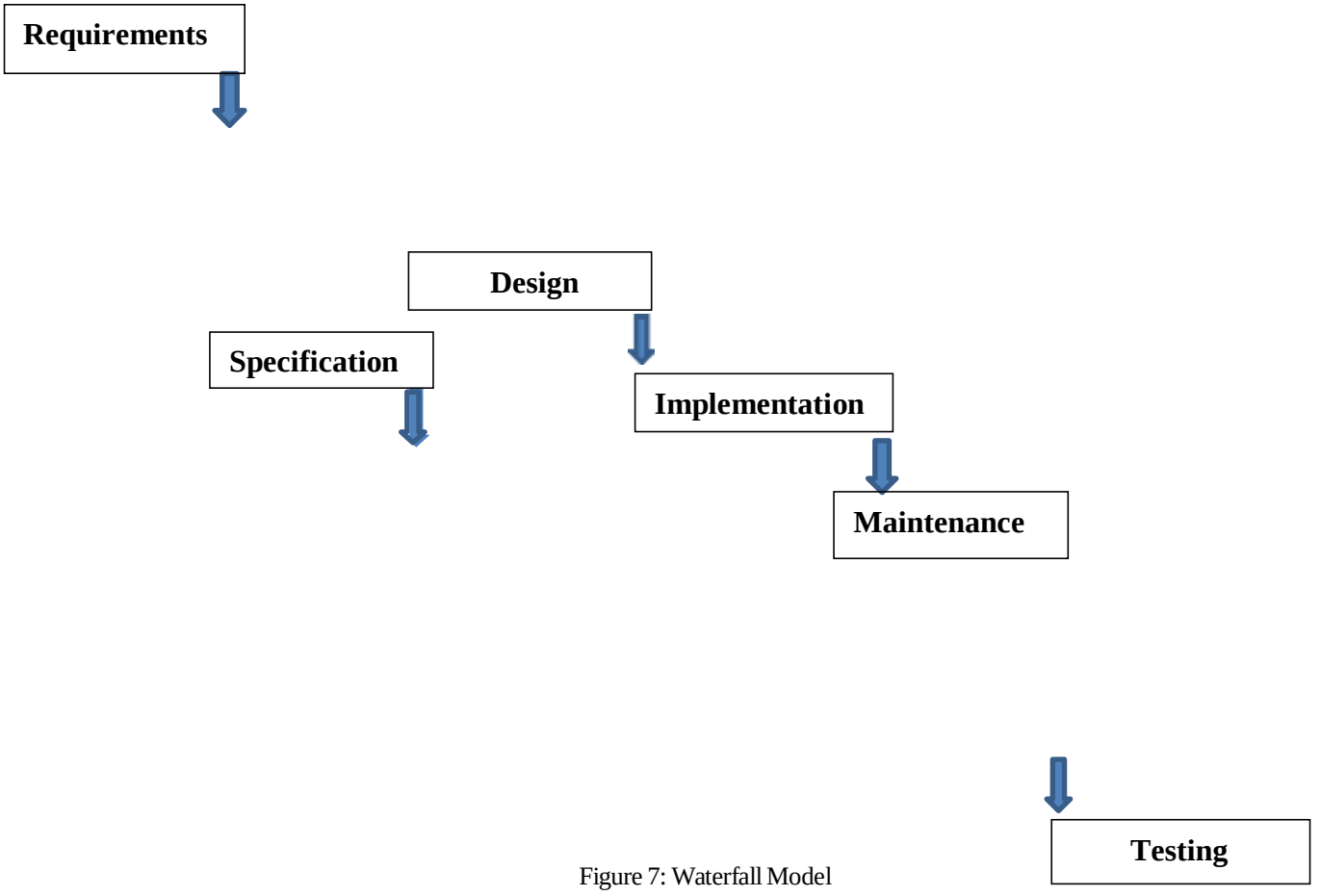


Figure 7: Waterfall Model

Implementation: The implementation phase involves the actual coding or programming of thesoftware. The output of this phase is typically the library, executables, user manuals and additional software documentation

Testing: In this phase all programs (models) are integrated and tested to ensure that the completesystem meets the software requirements.

The testing is concerned with verification and validation.

Maintenance: The maintenance phase is the longest phase in which the software is updated tofulfill the changing customer need, adapt to accommodate change in the external environment,correct errors and oversights previously undetected in the testing phase, enhance the efficiency ofthe software.

5.1.2 Advantages of Waterfall model

- Clear project objective
- Stable project requirements
- Progress of system is measurable.
- Logic of software development is clearly understood.
- Better resource allocation.

5.2 System Architecture

A system architecture is a conceptual model using which we can define the structure and behaviour of that system. It is a formal representation of a system. Depending on the context,system architecture can be used to refer to either a model to describe the system or a method usedto build the system. Building a proper system architecture helps in analysis of the project, especially in the early stages.

An overview of our model is shown in the figure given below. Our model consists of Road accident data that upon data pre-processing will form a dataset. Data pre- processing is one ofthe most important task in data mining. It deals with handling missing values orremoving attributes and makes it a structured form of data in order to perform analysis on it.

The dataset obtained will now be subjected to various data mining techniques. Clustering willbe performed on the given dataset. The main aim of performing clustering is to divide the datainto different clusters or groups such that the objects within a group are similar to each otherwhereas objects in other clusters are different from each other.[19]. There are severalclustering algorithms available: Hierarchical clustering technique like Ward method, single linkage, complete linkage etc, K means and latest class clustering (LCC).Other clustering algorithms like K-modes clustering is an enhanced version of K means clustering.

The clusters are then subjected to other algorithms like Association rule mining and trend analysis. Association rule mining is a very popular data mining technique that extracts interesting and hidden relations between various attributes in a large dataset. Association rule mining produces a set of rules that define underlying patterns in the dataset. The associativityis known by the frequency of their occurrence together in the dataset.

The trend analysis is performed to determine the upcoming trends based on the total countof accidents for each cluster. The trends analysis can also be applied on the entire dataset as well as the clusters. The trends can show a positive or negative trend for the future based on the current and past trends. There can be difference in trends for various clusters as there mightbe different factors dominant in causing accidents for that particular cluster. This trend will help us analyse the extent to which the measures taken across the years to reduce accidents has contributed in reducing the accident rate.

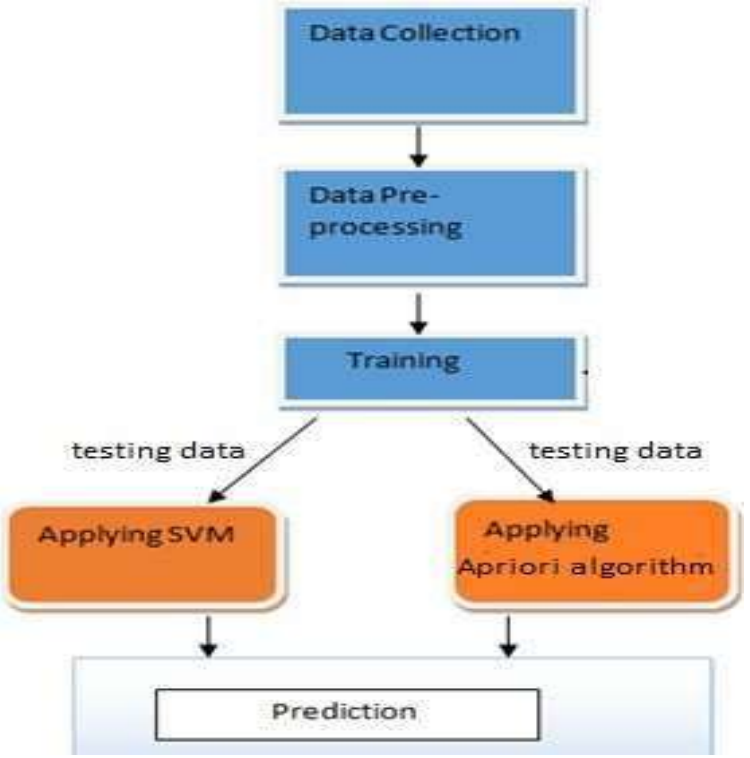


Figure 8 :System architecture

5.3 Use case Diagram:

A use case defines a goal-oriented set of interactions between external entities and the systemunder consideration. The external entities which interact with the system are its actors. A set of use cases describe the complete functionality of the system at a particular level of detail and it can be graphically denoted by the use case diagram.

A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved. A use case diagram can identify the different types of users of a system and the different use cases and will often be accompanied by other types of diagrams as well.

In software and systems engineering, a use case is a list of steps, typically defining interactions between a role (known in Unified Modeling Language (UML) as an "actor") anda system, to achieve a goal. The actor can be a human, an external system, or time.

In systems engineering, use cases are used at a higher level than within software engineering,often representing missions or stakeholder goals. The detailed requirements may then be captured in Systems Modeling Language (SysML) or as contractual statements.

The Seqeunce of activities that are carried out are the same as the other diagrams .Use case for this module indicates the users interaction with the system as a whole rather than individualmodules .All the encryption mechanisms are carried out via the login page that redirects the user to the particular functionality that he or she wishes to implement .

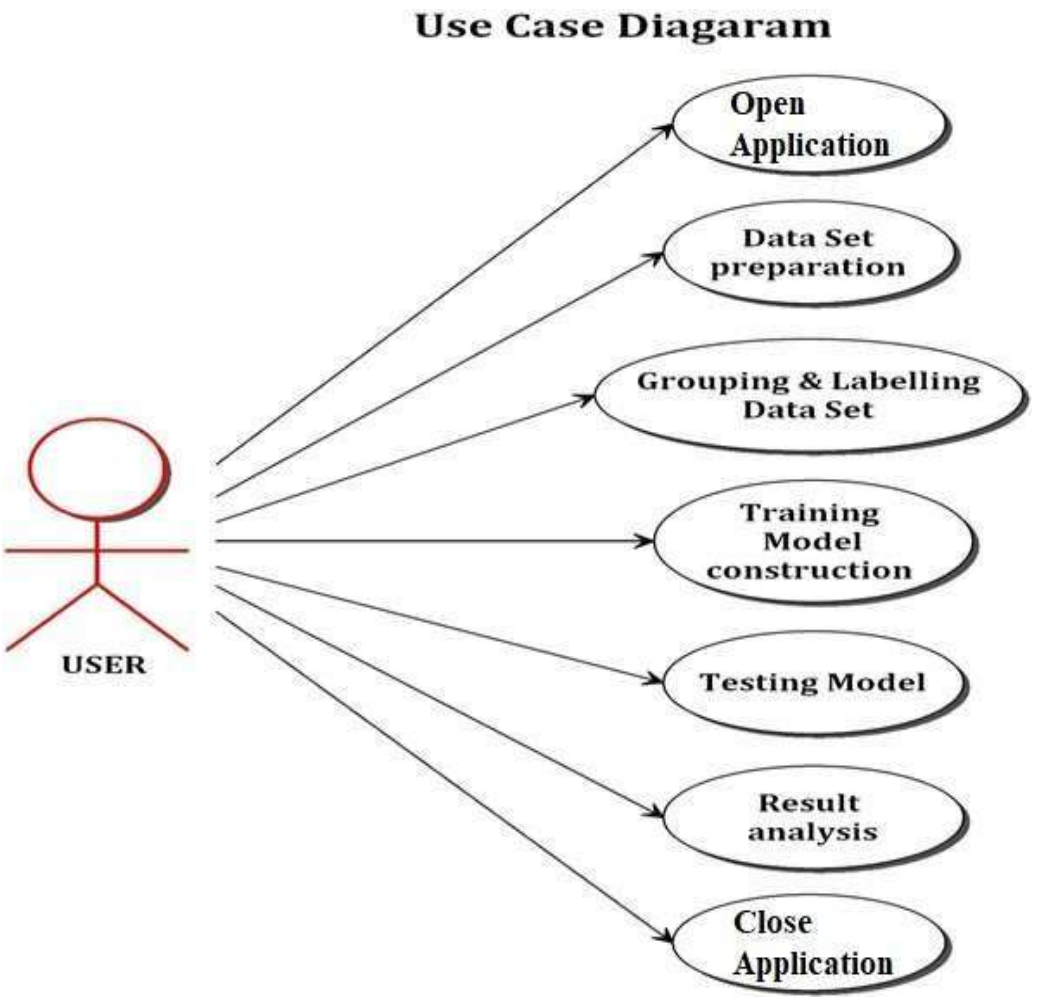


Figure 9: usecase diagramSequence Diagram:

Sequence diagram are an easy and intuitive way of describing the behavior of a system by viewing the interaction between the system and the environment. A sequence diagramshowsan interaction arranged in a time sequence. Sequence diagrams are typically associated with use case realizations in the Logical View of the system under development. Sequencediagrams are sometimes called event diagrams or event scenarios.

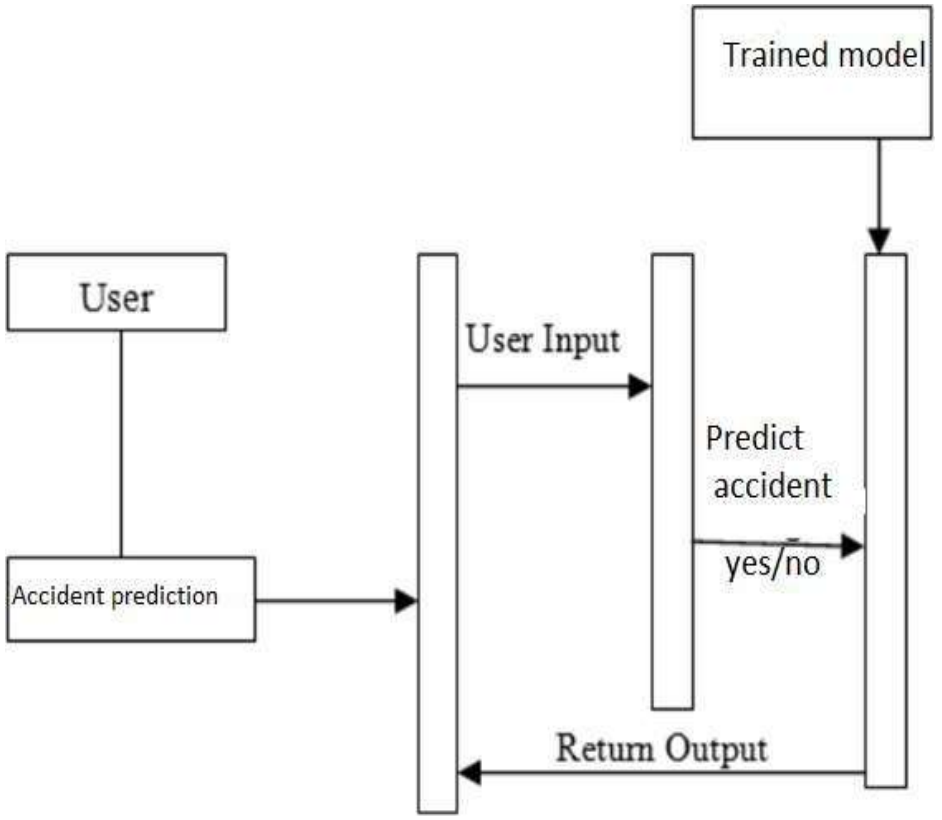


Figure 10: Sequence diagram

CHAPTER -6

TESTING

6.1 Software testing introduction

Software testing is a process used to help identify the correctness, completeness and quality of developed computer software. Software testing is the process used to measure the quality of developed software .Testing is the process of executing a program with the intent of findingerrors. Software testing is often referred to as verification & validation.

6.2 STLC (Software Testing Life Cycle):

Testing itself has many phases i.e. is called as STLC. STLC is part of SDLC

- Test Plan
- Test Development
- Test Execution
- Analyze Result
- Defect Tracking

TEST PLAN

It is a document which describes the testing environment, purpose, scope, objectives, test strategy, schedules, mile stones, testing tool, roles and responsibilities, risks, training, staffingand who is going to test the application, what type of tests should be performed and how it will track the defects.

TEST DEVELOPMENT

Preparing test cases, test data, Preparing test procedure, Preparing test scenario, Writing test script.

TEST EXECUTION

In this phase we execute the documents those are prepared in test development phase .

ANALYZE RESULT:

Once executed documents will get results either pass or fail. We need to analyze the results during this phase.

DEFECT TRACKING:

Whenever we get defect on the application we need to prepare the bug report file and forwardsto Test Team Lead and Dev Team. The Dev Team will fix the bug. Again we have to test theapplication. This cycle repeats till we get the software without defects.

6.3 TYPES OF TESTING:

White Box TestingBlack Box TestingGrey box testing

WHITE BOX TESTING:

White box testing as the name suggests gives the internal view of the software. This type of testing is also known as structural testing or glass box testing as well, as the interest lies in what lies inside the box.

BLACK BOX TESTING:

Its also called as behavioral testing. It focuses on the functional requirements of the software.Testing either functional or non functional without reference to the internal structure of the component or system is called black box testing.

GREY BOX TESTING:

Grey box testing is the combination of black box and white box testing. Intention of this testing is to find out defects related to bad design or bad implementation of the system.

6.4 LEVEL OF TESTING USED IN PROJECTUnit testing

Initialization testing is the first level of dynamic testing and is the first responsibility of developers and then that of the test engineers. Unit testing is performed after the expected testresults are met or differences are explainable/acceptable.

Integration testing

All module which make application are tested . Integration testing is to make sure that the interaction of two or more components produces results that satisfy functional requirement.

System testing

To test the complete system in terms of functionality and non functionality. It is black box testing, performed by the Test Team, and at the start of the system testing the complete systemis configured in a controlled environment.

Functional testing

The outgoing links from all the pages from specific domain under test. Test all internal links.Test links jumping on the same pages. Check for the default values of fields. Wrong inputs tothe fields in the forms.

Alpha testing

Alpha testing is final testing before the software is released to the general public. This testingis conducted at the developer site and in a controlled environment by the end user of the software.

Beta testing

The beta test is conducted at one or more customer sites by the end user of the software. Thebeta test is conducted at one or more customer sites by the end user of the software.

Chapter 7

IMPLEMENTATION

There are 4 major stages in the working of the project where each stage performs animportant aspect of the project. The 4 stages are

7.1 Rules

def Rule_display(): root1=Tk() root1.title("Rules") root1.geometry('800x800')

root1.configure(background="white")

label_6 = ttk.Label(root1, text = 'support',font=("Helvetica", 16),background="white")label_6.grid(row=1,column=0)

Entry_6 = Entry(root1) Entry_6.grid(row=1,column=1)

label_7 = ttk.Label(root1, text = 'confidence',font=("Helvetica", 16),background="white")label_7.grid(row=2,column=0)

Entry_7 = Entry(root1) Entry_7.grid(row=2,column=1)def ruleplot():

import matplotlib.pyplot as pltimport matplotlib.cm as cm from apyori import apriori import pandas as pd import numpy as np

df=pd.read_csv('3g.csv')df.shape

df.columns del df['SrNo']del df['Fatal']

```
del df['Grevious']del df['Minor'] del df['Injured'] del df['Date']

del df['a'] records = []

for i in range(0, 807):

records.append([str(df.values[i,j]) for j in range(0, 10)])

association_rules = apriori(records, min_support=float(Entry_6.get()),
min_confidence=float(Entry_7.get()), min_lift=1.0)

association_results = list(association_rules)for item in association_results

pair = item[0]

items = [x for x in pair]import random

import matplotlib.pyplot as pltsupport=[]

confidence=[]
lift=[] color=[]
for a in association_results: support.append(a[1]) confidence.append(a[2][0][2])

lift.append(a[2][0][3])

color.append(a[2][0][3]*20.0)print(len(support))

rules = []

for item in association_results:pair = item[0]

items = [x for x in pair]

rules.append("Rule: " + str(items)+"->"+ items[0]+"\n")#print("Rule: " + str(items)+"->"+ items[0])

print(rules)
```

7.2 Plot Graph

```
def plot_graph():root10 = Tk()

root10.title('GRAPHS') root10.geometry('400x400') root10.configure(background="white")"""var = StringVar()

label = Label( root, textvariable = var,font=('arial',20,'bold'),bd=20,background="Powderblue")

var.set("Predict STRESS DETECTOR")label.grid(row=0,columnspan=6)

"""

from PIL import ImageTk,Imagedef graph1():

image = Image.open("graph1.png")

image = image.resize((250, 250), Image.ANTIALIAS)img = ImageTk.PhotoImage(image)

panel1 = Label(root, image=img)panel1.image = img panel1.grid(row=3,column=0)

def graph2():

image = Image.open("graph2.png")
```

```
image = image.resize((250, 250), Image.ANTIALIAS)img = ImageTk.PhotoImage(image)

panel2 = Label(root, image=img)panel2.image = img panel2.grid(row=3,column=1)

def graph3():

image = Image.open("graph3.png")

image = image.resize((250, 250), Image.ANTIALIAS)img = ImageTk.PhotoImage(image)
panel3 = Label(root, image=img)panel3.image = img panel3.grid(row=3,column=2)

                                                    def graph4():

import numpy as np # linear algebra

import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)import os

import matplotlib.pyplot as pltimport seaborn as sns

from pandas import DataFrame , read_csv df=pd.read_csv('Details_of_road_accident_deaths_by_situation_state.csv')

ncount=df['STATE/UT'].value_counts() ap=df[df['STATE/UT']=='ANDHRA PRADESH']

ap1=ap[ap['Year']==2001]a=ap1.head(14)

ap=df[df['STATE/UT']=='KARNATAKA']

ap1=ap[ap['Year']==2001] a=ap1.head(14) t=df[df['CAUSE']=='Total Truck/Lorry']t.head()
```

7.3 Risk Prediction

```
def risk_predict():root11 = Tk()root11.title('RISK PREDICTION') root11.geometry('400x400') root11.configure(background="white")

                                                    var = StringVar()

label = Label( root11, textvariable =var,font=('arial',20,'bold'),bd=20,background="white")

var.set("RISK PREDICTION")

label.grid(row=0,columnspan=6)

label_1 = ttk.Label(root11, text ='Area',font=("Helvetica", 16),background="white")label_1.grid(row=1,column=0)

                                                    tkvar = StringVar(root11)

choices = ['A
Narayanapura','Agaram','Banasavadi','Basavanapura','Bellanduru','Benniganahalli','Bharathi Nagar','BTM Layout','C V Raman
Nagar','Chickpete','Devasandra','Dharmaraya Swamy Temple','Dodda Nekkundi','Domlur','Garudachar
Playa','Gurappanapalya','Hagadur','HAL Airport','Halsoor','Hemmigepura','Horamavu','Hoysala Nagar','HSR
Layout','Hudi','J PNagar','Jaraganahalli','Jayanagar East','Jeevanbhima
Nagar','Jogupalya','K RPuram','Kacharkanahalli','Kadugodi','Kammanahalli','Konena
Agrahara','Madivala','Marathahalli','New Tippasandara','Other','other','RamamurthyNagar','Sampangiram
Nagar','Sarakki','Shantala Nagar','Singasandra','SudhamNagara','Varthuru','Vasanthpura','Vijnana
Nagar','Vijnanapura','Yelchenahalli']

popupMenu = OptionMenu(root11, tkvar, choices[1], *choices)

#Label(root, text="select area",background="purple2").grid(row=0,column=0)popupMenu.grid(row=1,column=1)
```

tkvar.set('Select area')def pass1():

import pandas as pd import numpy as np global kmf2label kmf2label = {'Area' : ""}
kmf2label['Area'] = {'Kadugodi': 1, 'Garudachar Playa': 1, 'Hudi': 1, 'Other': 1,
'Devasandra': 1,'Hagadur': 1, 'Bellanduru': 0, 'Marathahalli': 0, 'Dodda Nekkundi': 0 }bdf = pd.read_excel('bangalore-cas-alerts.xlsx')

bdf.info()

bdf = bdf.rename(columns = {'deviceCode_time_recordedTime_$date':'timestamp'})bdf['timestamp'] = pd.to_datetime(bdf['timestamp'])

bdf['eventDate'] = pd.to_datetime(bdf['timestamp']) bdf['eventDate'] = bdf['eventDate'].dt.strftime('%Y%m%d')

bdf['e_hour'] = pd.to_datetime(bdf['timestamp'], format = '%H:%M:%S').dt.hourbdf['ehourCat'] = 0

badf = pd.read_excel('bangalore-accident-zones.xlsx')b = pd.merge(b1, badf, on = ['Area'], how = 'left')

b = b.rename(columns = {'deviceCode_pyld_alarmType':'Alarm_Type'})b = b.rename(columns =

{'deviceCode_pyld_speed':'Plying_Speed'}) b['hasOversped'] = np.where(b.Plying_Speed > 60, 1, 0)

b['hasOversped'] = np.where(b.Alarm_Type == 'Overspeed', 1, b['hasOversped'])for column in ['temperature', 'visibility', 'condition']:

b[column].fillna(b[column].mode()[0], inplace=True)b['visibility'] = np.where(b['visibility'] < 10, 0, 1)

df = b.copy()

df['hasOversped'] = np.where(b.hasOversped == 1, 'Yes', 'No')
df['visibility'] = np.where(b.visibility == 0, 'Low', 'High')

df['ehourCat'] = b['ehourCat'].map({1: 'Early', 2: 'PeakM', 3: 'RegularM'})

b['Accident_Severity'] = b['Accident_Severity'].map({'High': 3, 'Medium': 2, 'Low':
1})

b['Pothole_Severity'] = b['Pothole_Severity'].map({'High': 3, 'Medium': 2, 'Low': 1})

b['Alarm_Type'] = b['Alarm_Type'].map({'PCW': 1, 'FCW': 2, 'Overspeed': 3, 'HMW':
4, 'UFCW': 5, 'LDWL': 6, 'LDWR': 7})

b['condition'] = b['condition'].map({'Clear': 1, 'Sunny': 2, 'Passing clouds': 3,

'Broken clouds': 4, 'Scattered clouds': 5, 'Fog': 6, 'Haze': 7, 'Partly cloudy': 8,

'Mild': 9, 'Drizzle. Broken clouds': 10})

7.4 Entry New Data

data_new = [] def new_data():

root10 = Tk()

root10.title('Predict Heart DETECTOR')root10.geometry('400x400') root10.configure(background="white") """var = StringVar()

label = Label(root, textvariable =
var,font=('arial',20,'bold'),bd=20,background="Powderblue")

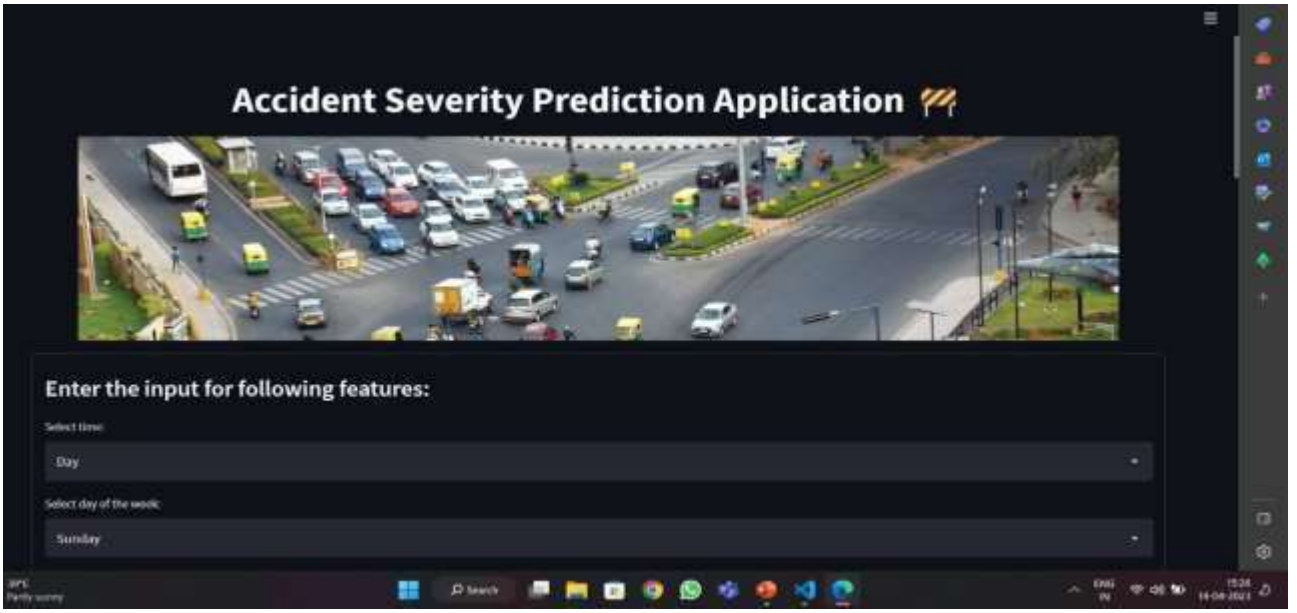
var.set("Predict STRESS DETECTOR")label.grid(row=0,columnspan=6)

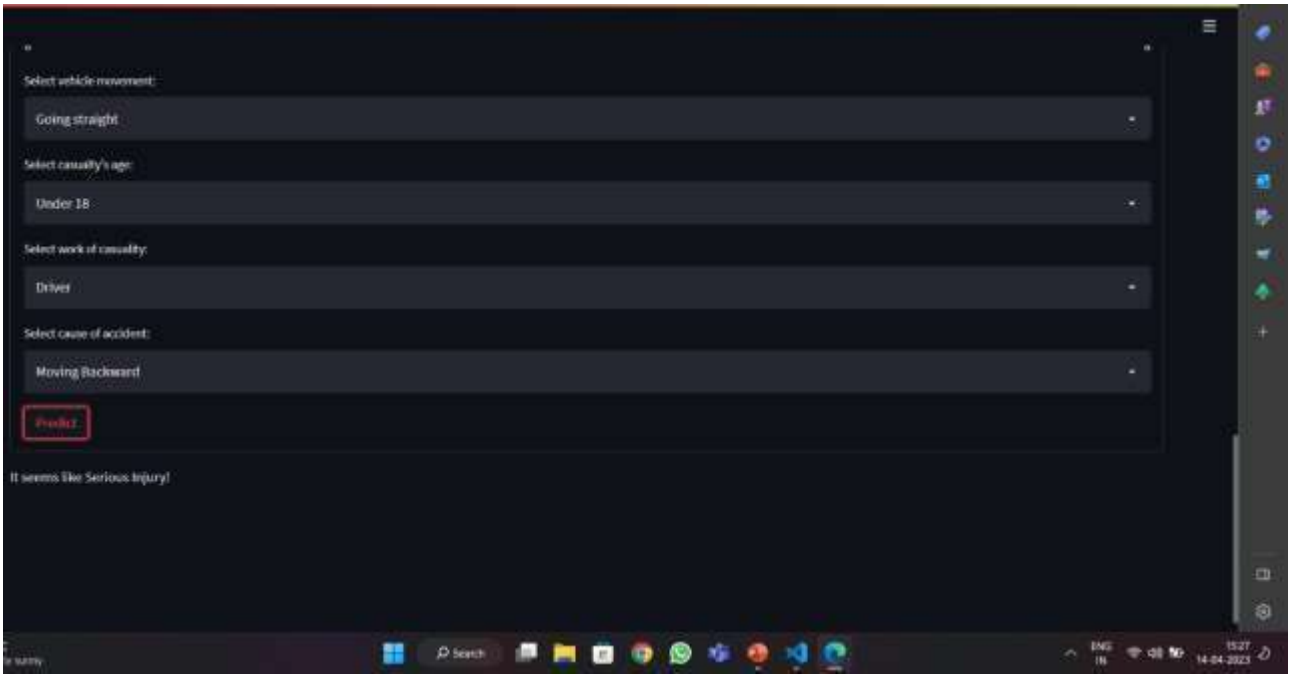
"""

```
label_1 = ttk.Label(root10, text ='Date',font=("Helvetica", 16),background="white")
label_1.grid(row=0,column=0) Entry_1 = Entry(root10) Entry_1.grid(row=0,column=1)

label_2 = ttk.Label(root10, text = 'Time',font=("Helvetica", 16),background="white")label_2.grid(row=1,column=0)

Entry_2 = Entry(root10) Entry_2.grid(row=1,column=1)

label_3 = ttk.Label(root10, text = 'Location',font=("Helvetica", 16,),background="white")

label_3.grid(row=2,column=0) Entry_3 = Entry(root10) Entry_3.grid(row=2,column=1)

label_4 = ttk.Label(root10, text = 'Type of vehicle' ,font=("Helvetica", 16),background="white")

label_4.grid(row=3,column=0) Entry_4 = Entry(root10) Entry_4.grid(row=3,column=1)

label_5 = ttk.Label(root10, text = 'Type of accident',font=("Helvetica", 16),background="white")

label_5.grid(row=4,column=0) Entry_5 = Entry(root10) Entry_5.grid(row=4,column=1)
```

Chapter 8

8.1SCREEN SHORTS





8.2Modules

COLLECTING DATA

PREPARING THE DATA

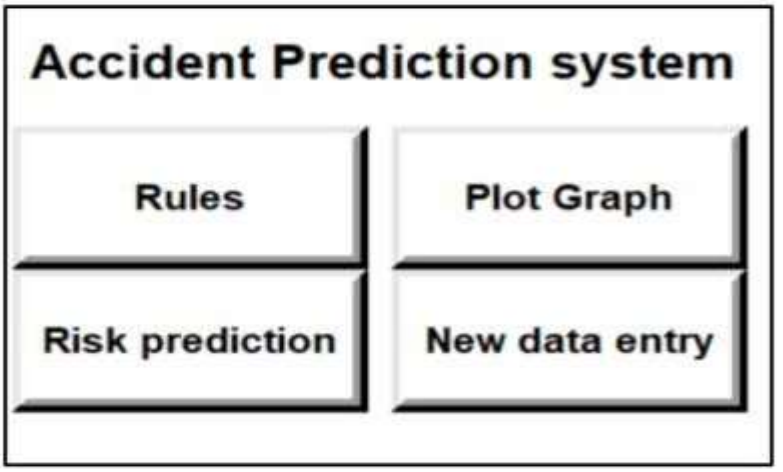
CHOOSING THE MODEL

TRAINING A MODEL

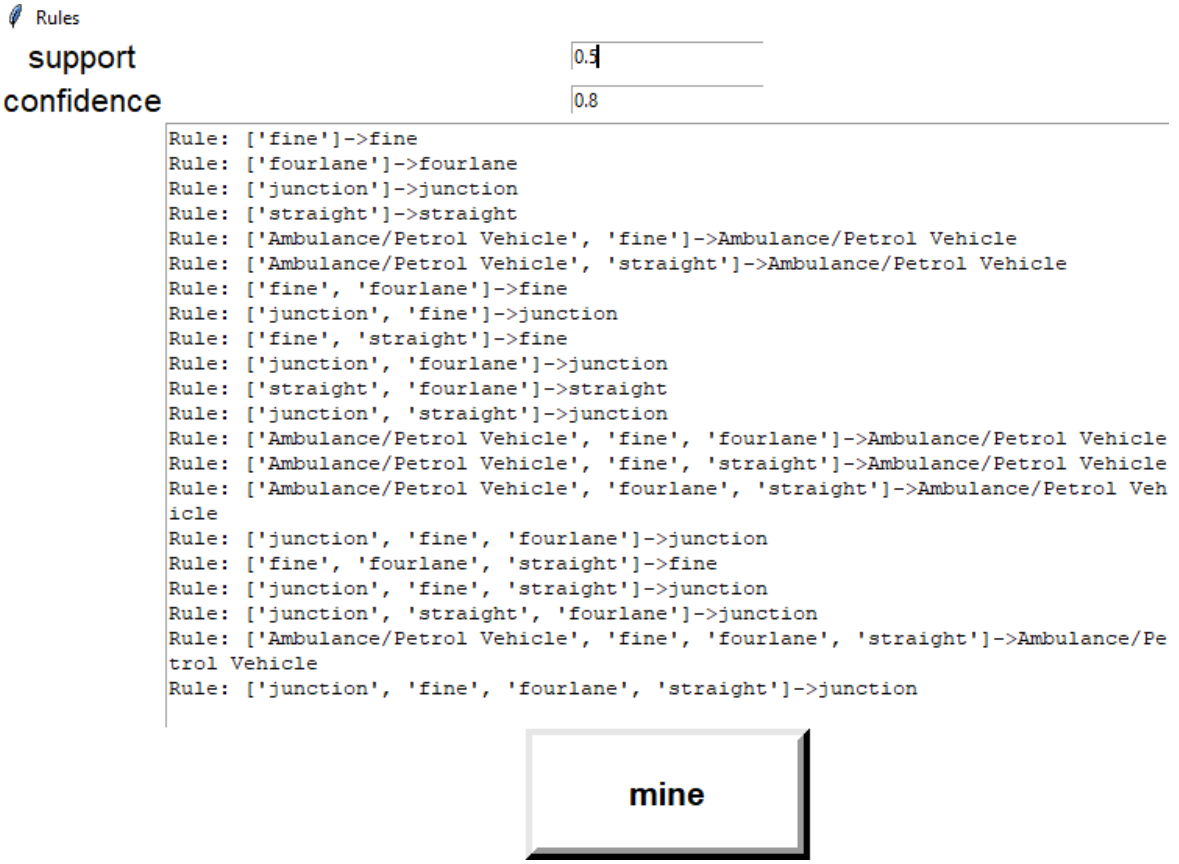
EVALUATING THE MODEL

PARAMETER TUNING

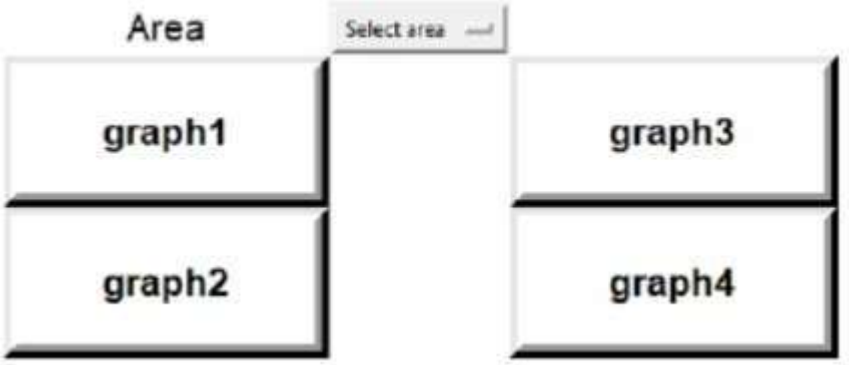
MAKING PREDICTIONS



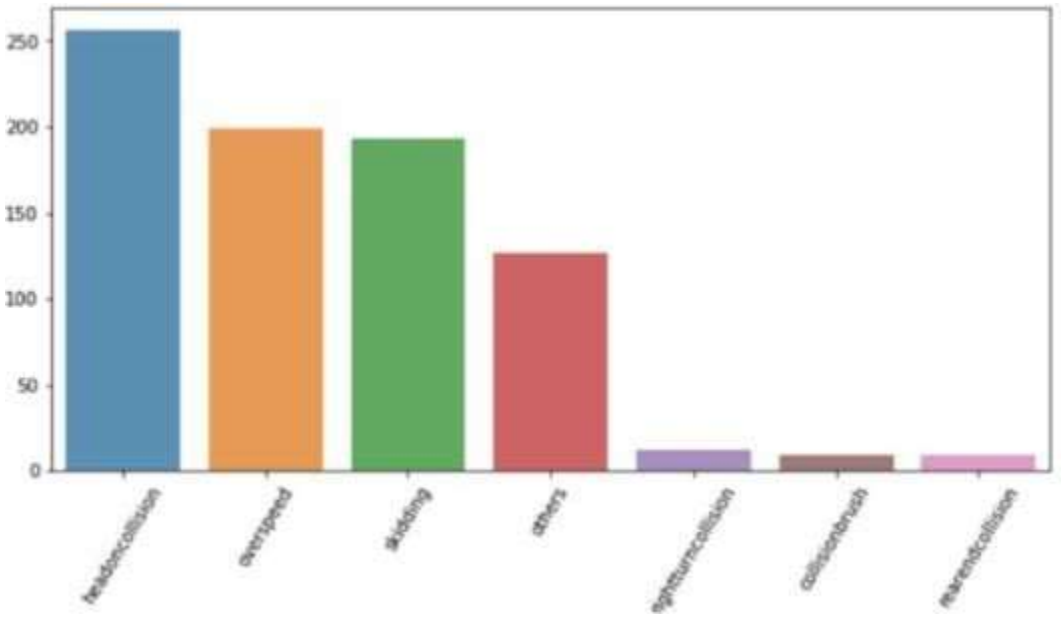
8.3Rule mining



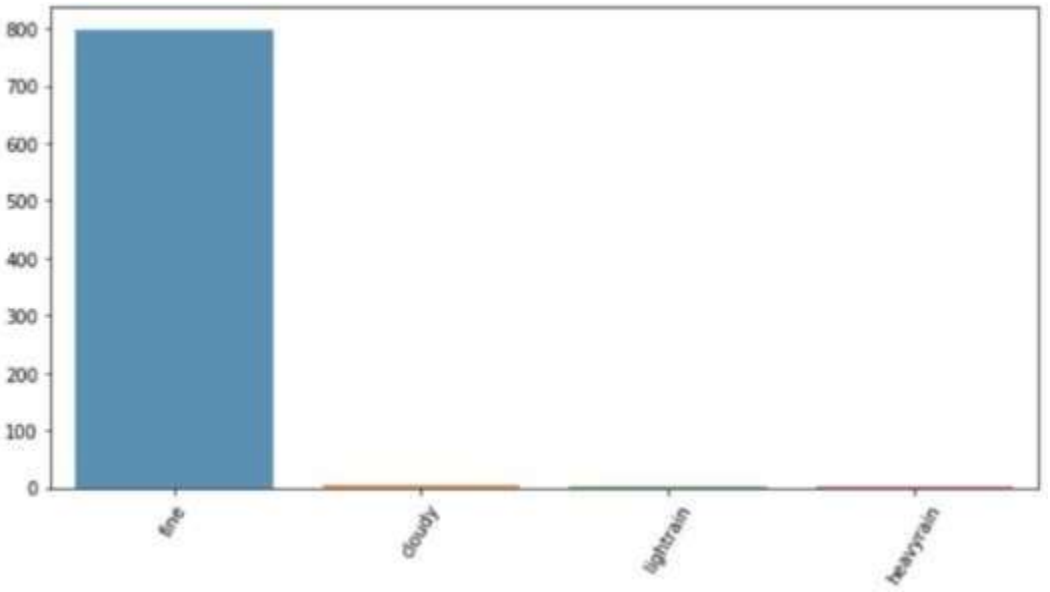
8.4Plot graph



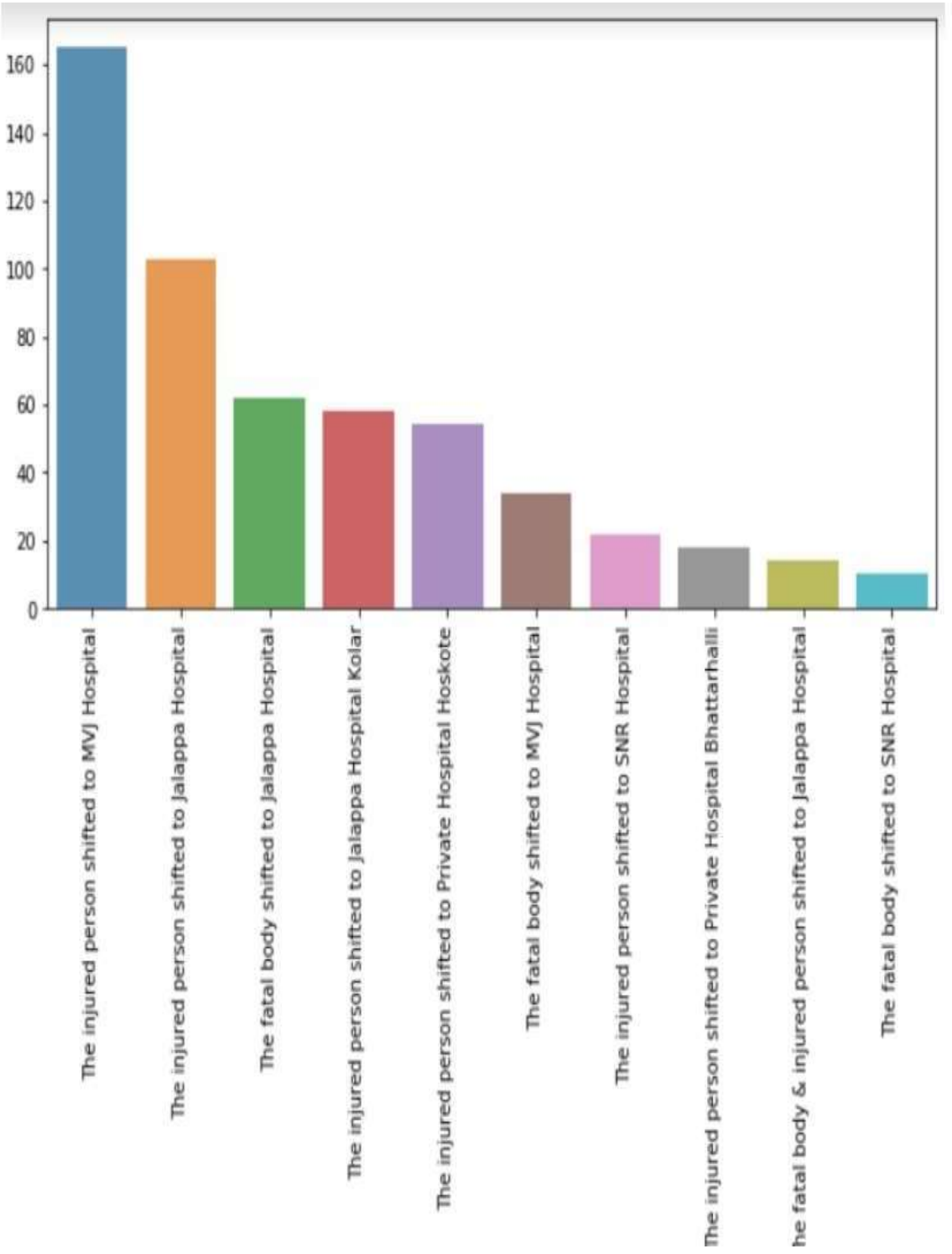
8.4.1 GRAPH 1



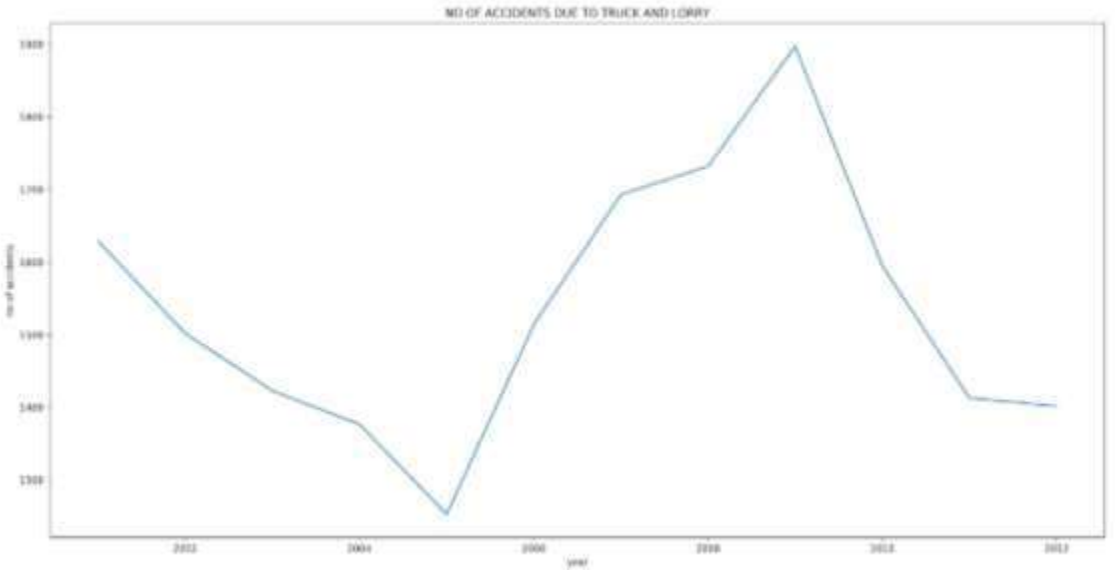
8.4.2 GRAPH 2



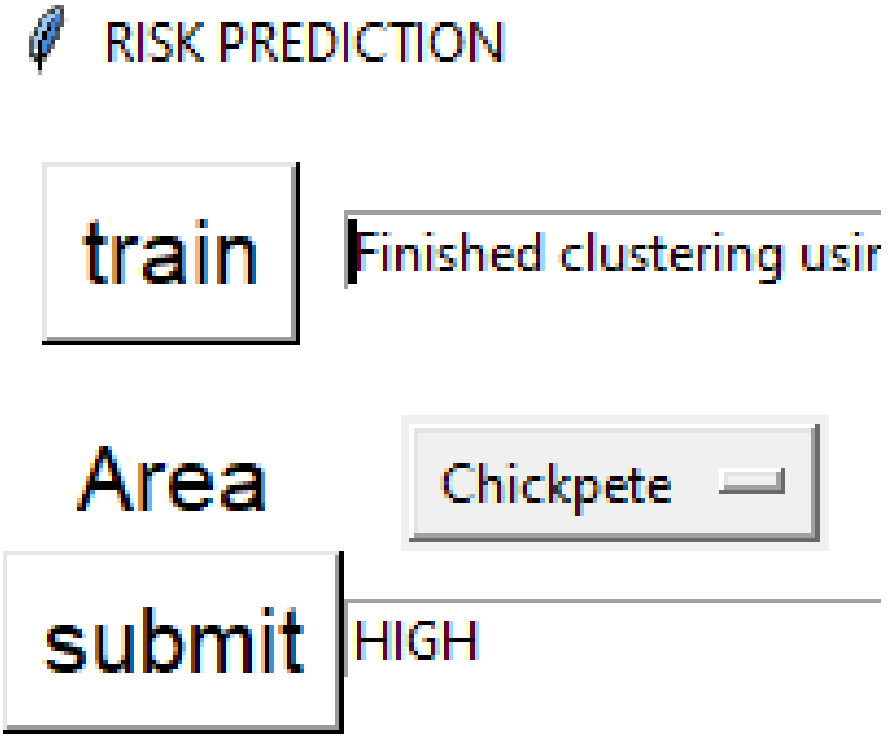
8.4.3 GRAPH 3


160
140
120
100
80
60
40
20
0
The injured person shifted to MVJ Hospital
The injured person shifted to Jalappa Hospital
The fatal body shifted to Jalappa Hospital
The injured person shifted to Jalappa Hospital Kolar
The injured person shifted to Private Hospital Hoskote
The fatal body shifted to MVJ Hospital
The injured person shifted to SNR Hospital
The injured person shifted to Private Hospital Bhattarhalli
he fatal body & injured person shifted to Jalappa Hospital
The fatal body shifted to SNR Hospital

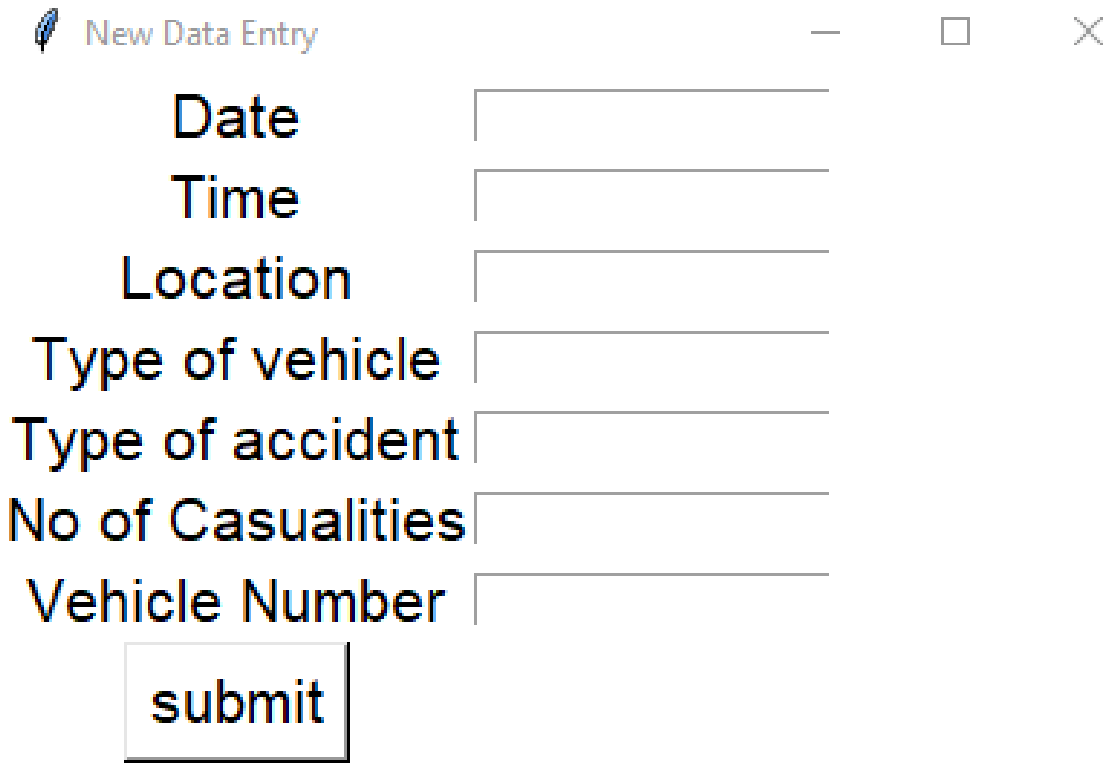
8.4.4 GRAPH 4



8.4Risk prediction



8.5New data entry



CHAPTER 9

CONCLUSION

Road Accidents are caused by various factors. By going through all the research papers itcan be concluded that Road Accident cases are hugely affected by the factors such as types ofvehicles, age of the driver, age of the vehicle, weather condition, road structure and so on. Thus we have build an application which gives efficient prediction of road accidents based onthe above mentioned factors.

The classification algorithm of the entire dataset. In the Road Accident prediction final resultis to find the percentage of accident in particular area. Having lower number of features helpsthe algorithm to converge faster and increases accuracy. In the Road Accidentprediction finalresult is to find the percentage of accident in particular area. Then we apply logistic regressionon these features and obtain the least error.

ACKNOWLEDGEMENT

We express our gratitude and thanks to Our Parents first for giving health and a sound mind for completing this project. We give all the glory and thanks to our almighty GOD for showering upon, the necessary wisdom and grace for accomplishing this project .

It is our pleasant duty to express deep sense of gratitude to our Honorable Chairman Dr.S.DILLI SARAN RAJ, and our Beloved CEO Mam Smt. SUJATHA MARAN, for their kind encouragement.

We have unique pleasure in thanking our Principal Dr.A.DEVARAJU, M.Tech., Ph.D., for his unflinching devotion, which lead us to complete this project.

We express our faithful and sincere gratitude to our Head of the Department & our Project Coordinator Dr.THANGAVEL, M.E., Department of Computer Science and Engineering, for her support & valuable suggestions.

We express our faithful and sincere gratitude to our guide MRS.ELAVARASI ,M.E, Department of Computer Science and Engineering , for his valuable guidance and support throughout the project.

We render our thanks to all the Faculty and Lab Technicians of Department of Computer Science and Engineering for their timely assistanc

CHAPTER 10

REFERENCES

- Ayushi Jain, Garmin Ahuja, Anuran Jana, Deepti Mehrotra,Data mining approach toanalyze the road accidents in India, proc of IEEE, 2016
- Global status report on road safety: supporting a decade of action, Geneva, World healthorganization, 2013.
- Open Government Data (OGD) platform India [online]. available 2016:

https://data.gov.in/catalogs/sector/Transport-9383

- Ms. Gagan deep Kaur, Harpreet Kaur, Prediction of the cause of accident and accidentprone location on roads using data mining techniques, Proc of IEEE,2017.
- Handbook of Research on Computational Science and Engineering: Theory andpractice.
- Irina Makarova, Ksenia Shubenkova, Eduard Mukhametdinov, and Anton Pashkevich,Safety related problems of transport system and their solutions, proc of IEEE, 2018
- WHO, “Global status report on road safety 2015”
- Peden, World Health Organization. Ed. by Margie- 2004. World report on road trafficinjury prevention. Geneva: World Health Organization.
- World report on road traffic injury prevention: summary

http://www.who.int/worldhealthday/2004/infomaterials/world_report/en/summary_en_rev
.pdf, (July 24, 2004). (2004).