

ROAD ACCIDENT ANALYSIS on URBAN ROADS

Submitted in partial fulfillment of the requirements for
the award of
Bachelor of Engineering degree in Computer Science and Engineering by

B.Thamby Jeevan Kumar (17SCS5029)
B. Chandra Shekhar (17SCS0030)



**DEPARTMENT OF COMPUTER SCIENCE AND
ENGINEERING SCHOOL OF COMPUTING
SATHYABAMA INSTITUTE OF SCIENCE AND
TECHNOLOGY JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600119, TAMILNADU**

AUGUST 2021



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC

(Established under Section 3 of UGC Act, 1956)

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI- 600119

www.sathyabamauniversity.ac.in



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **B.THAMBY JEEVAN KUMAR(17SCS5029)**, **B.CHANDRA SHEKAR(17SCS0030)** who carried out the project entitled "**ROAD ACCIDENT ANALYSIS on URBAN ROADS**" under my supervision from December 2020 to march 2021.

Internal Guide
M Maheswari, M.E,Ph.D

Head of the Department
Dr. S.VIGNESHWARI , M.E., Ph.D.,
Dr.L.LAKSHMANAN, M.E., Ph.D.,

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I B. THAMBY JEEVAN KUMAR (17SCS5029) and B. CHANDRA SHEKAR(17SCS0030)

hereby declare that the Project Report entitled **Road accident analysis on urban roads** done by me under the guidance of Dr. /Prof. / Mr. /Ms. **M Maheswari, M.E, Ph.D** (Internal) and(External) at _____(Company name and address) is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E., Ph.D., Dean**, School of Computing , **Dr.S.Vigneshwari M.E., Ph.D., and Dr.L.Lakshmanan M.E., Ph.D.**, Heads of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. M Maheswari M.E., Ph.D.**, for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

This report shows the results from the research study by applying large scale data mining methods into analysis of accidents on the roads. The data sets collected from traffic fatal accidents are huge, multidimensional, and heterogeneous. Moreover, they may contain incomplete and enormous values, which make its exploration and understanding a very demanding task. The target data of this study is collected by the Road Administration Data. The intention is to investigate the usability of robust clustering, association and frequent item sets, and visualization methods to the road traffic accident analysis. While the results show that the selected data mining methods are able to produce understandable patterns from the data, finding more fertilized information could be enhanced with more detailed and comprehensive data sets. *k*-means algorithm takes accident frequency count as a parameter to cluster the locations. Then we used association rule mining to characterize these Surface Condition. The rules revealed different factors associated with road accidents at different drunk and drive with varying accident frequencies. The association rules for high-frequency accident location disclosed that intersections on highways are more dangerous for every type of fatal accidents. This analysis will be helpful for the users to track the dangerous parts of the roads where frequent accidents occurred for a safe journey.

TABLE OF CONTENTS

CHAPTER NAME	TITLE	PAGE NO
	ABSTRACT	
	LIST OF FIGURES	
1	INTRODUCTION	1
	1.1 OBJECTIVE	2
2	AIM AND SCOPE OF THE PROJECT	3
	2.1 AIM OF THE PROJECT	3
	2.2 SCOPE OF THE PROJECT	3
	2.3 EXISTING SYSTEM	4
	2.4 PROPOSED SYSTEM	4
3	LITERATURE SURVEY	6
	3.1 AUTHOR TITLES	6
4	PROJECT ALGORITHM	12
	4.1 ALGORITHM	12
	4.2 SYSTEM ARCHITECTURE	13
	4.3 DATA FLOW DIAGRAM	14
	4.4ROLE OF DFD	14
	4.5 UML DIAGRAM	15
	4.5.1 GOALS	15
	4.6 USECASE	16
	4.7 SYSTEM ANALYSIS	17
	4.7.1 HARDWARE REQUIRMENTS	17
	4.7.2 SOFTWARE REQUIRMENTS	17
5	SOFTWARE ENVIRONMENT	18
	5.1 PYTHON	18
	5.2 HISTORY OF PYTHON	18
	5.3 PYTHON FEATURES	19
	5.4 ECONOMICAL FEASIBILITY	23
	5.5 TECHNICAL FEASIBILITY	23
	5.6 SOCIAL FEASIBILITY	23
6	MODULE IMPLEMENTATIONA ND DISCUSSION	24
	6.1 MODULES	24
	6.2 USER INTERFACE MODULE	24
	6.3 LOGIN MODULE	24
	6.4 REGISTRATION MODULE	25
	6.5 ANALYSIS MODULE	25
	6.6 REPORT ACCIDENT MODULE	25
7	SYSTEM TESTING	26
	7.1 INPUT DESIGN	26
	7.2 OUTPUT DESIGN	26
	7.3 TYPES OF TESTING	27
	7.4 REQUIRMRNT ANALYSIS	30
8	CONCLUSION	32
	REFERENCES	33
	APPENDIX	35
	SCREENSHOTS	35

LIST OF FIGURES

FIGURE NO	FIGURE NAME	PAGE NO
4.2.1	SYSTEM ARCHITECTURE	13
4.6.1	DATA FLOW DIAGRAM	16

CHAPTER 1

INTRODUCTION

There are a lot of vehicles driving on the roadway every day, and traffic accidents could happen at any time anywhere. Some accident involves fatality, means people die in that accident. As human being, we all want to avoid accident and stay safe. To find out how to drive safer, data mining technique could be applied on the traffic accident dataset to find out some valuable information, thus give driving suggestion. Data mining uses many different techniques and algorithms to discover the relationship in large amount of data. It is considered one of the most important tool in information technology in the previous decades. Association rule mining algorithm is a popular methodology to identify the significant relations between the data stored in large database and also plays a very important role in frequent itemset mining. A classical association rule mining method is the Apriori algorithm who main task is to find frequent itemsets, which is the method we use to analyze the roadway traffic data. Classification in data mining methodology aims at constructing a model (classifier) from a training data set that can be used to classify records of unknown class labels. The Naive Bayes technique is one of the very basic probability-based methods for classification that is based on the Bayes' hypothesis with the presumption of independence between each pair of variables. We used the FARS dataset for our study. The Fatal Accidents Dataset contains all fatal accidents on public roads in 2007 reported to the National Highway Transportation Safety Administration. The dataset is downloaded from California Polytechnic State University and all data originally came from FARS. The dataset contains 37,248 records and 55 attributes. The data description can be found in the document FARS.

1.1 OBJECTIVE

The purpose of this study was to conduct an indepth study focusing on the application of event analysis through crash investigation and reconstruction. The objectives of this study were the followings: To identify the contributory factors based on the findings obtained from crash investigation and reconstruction by using a case study. To apply an event analysis in establishing the links between the events to describe the crash scenario based on the available information.

CHAPTER 2

AIM AND SCOPE OF THE PROJECT

2.1 AIM OF THE PROJECT

Road traffic injury is a major global public health problem. Rapid motorization in low and middle-income countries along with the poor safety quality of road traffic systems and the lack of institutional capacity to manage outcomes contribute to a growing crisis.

More than 1.24 million people die each year on the world's roads. Many more suffer permanent disability, and between 20 and 50 million suffer non-fatal injuries. These are mainly in amongst vulnerable road users and involve the most socio-economically active citizens.

2.2 SCOPE OF THE PROJECT

Design is a multi- step that focuses on data structure software architecture, procedural details, algorithm etc... and interface between modules. The design Process also translate the requirements into presentation of software that can be Accessed for quality before coding begins. Computer software design change Continuously as new methods; better analysis and border understanding evolved. Software design is at relatively early stage in its revolution. Therefore, software design methodology lacks the depth, flexibility and quantitative nature that are normally associated with more classical engineering disciplines. However techniques for software designs do exit, criteria for design qualities are available and design notation can be applied.

2.3 EXISTING SYSTEM

The traffic accident using data mining technique that could possibly reduce the fatality rate. Using a road safety database enables to reduce the fatality by implementing road safety programs at local and national levels. Classification models to predict the severity of injury that occurred during traffic accidents. Association rules mining algorithm on a dataset about traffic accidents which was gathered from Government Traffic Office, Apriori and Predictive Apriori association rules algorithms were applied to the dataset to investigate the connection between recorded accidents and factors to accident severity.

2.3.1 Disadvantages

- Here we are using data mining the fault traffic accident injuries and deaths
- In this the cost of maintain and reparing the roads
- This will not useful for short distance level

2.4 PROPOSED SYSTEM

This paper presents our research to model the severity of injury resulting from traffic accidents using artificial neural networks and decision trees. We have applied them to an actual data set obtained from the National Automotive Sampling System (NASS) General Estimates System (GES). Experiment results reveal that in all the

cases the decision tree outperforms the neural network. Our research analysis also shows that the three most important factors in fatal injury are: driver's seat belt usage, light condition of the roadway, and driver's alcohol usage. Our experiments also showed that the model for fatal and non-fatal injury performed better than other classes. The ability of predicting fatal and non-fatal injury is very important since drivers' fatality has the highest cost to society economically and socially.

2.4.1 ADVANTAGES OF PROPOSED SYSTEM

- speed was available

improve the performance of analysis in fatal and non-fatal accidents

CHAPTER 3

3.1 AUTHOR TITLES

Title 1: Execution of Apriori algorithm of data mining directed towards tumultuous crimes concerning women

Author: divay bansal and lekha bhambhu

Apriori Algorithm is the most popular and useful algorithm of Association Rule Mining of Data Mining. As Association rule of data mining is used in all real life applications of business and industry. Objective of taking Apriori is to find frequent itemsets and to uncover the hidden information. This paper elaborates upon the use of association rule mining in extracting patterns that occur frequently within a dataset and showcases the implementation of the Apriori algorithm in mining association rules from a dataset containing crimes data concerning women. As for this WEKA tool is used for extracting results .For this one dataset is taken from UCI repository and other data is collected manually from the session court of sirsa to collect data on heart melting crimes against women. The main motive to use UCI is to first check the proper working of dataset and then apply Apriori on real dataset against crimes on women which extracts hidden information that what age group is responsible for this and to find where the real culprit is hiding. In last the comparison is done between Apriori & PredictiveApriori Algorithm in which Apriori is better and faster than PredictiveApriori Algorithm

Title 2: Applying association rules mining algorithms for traffic accidents in dubai.

Author: mira A El Tayeb, Vikas Pareek, and Abdelaziz Araar

Association rule mining algorithms are widely used to find all rules in the database satisfying some minimum support and minimum confidence constraints. In order to decrease the number of generated rules, the adaptation of the association rule mining algorithm to mine only a particular subset of association rules where the classification class attribute is assigned to the right-hand-side was investigated in past research. In this research, a dataset about traffic accidents was collected from Dubai Traffic Department, UAE. After data preprocessing, Apriori and Predictive Apriori association rules algorithms were applied to the dataset in order to explore the link between recorded accidents' factors to accident severity in Dubai. Two sets of class association rules were generated using the two algorithms and summarized to get the most interesting rules using technical measures. Empirical results showed that the class association rules generated by Apriori algorithm were more effective than those generated by Predictive Apriori algorithm. More associations between accident factors and accident severity level were explored when applying Apriori algorithm.

Title 3: A perspective analysis of traffic accident using data mining techniques.

Author: S. Krishnaveni and M. Hemalatha

Data Mining is taking out of hidden patterns from huge database. It is commonly used in a marketing, surveillance, fraud detection and scientific discovery. In data mining, machine learning is mainly focused as research which is automatically learnt

to recognize complex patterns and make intelligent decisions based on data. Nowadays traffic accidents are the major causes of death and injuries in this world. Roadway patterns are useful in the development of traffic safety control policy. This paper deals with the some of classification models to predict the severity of injury that occurred during traffic accidents. I have compared Naive Bayes Bayesian classifier, AdaBoostM1 Meta classifier, PART Rule classifier, J48 Decision Tree classifier and Random Forest Tree classifier for classifying the type of injury severity of various traffic accidents. The final result shows that the Random Forest outperforms than other four algorithms.

Title 4: Analysing road accident data using association rule mining

Author: Sachin Kumar and Durga Toshniwal

Road accident is one of the crucial areas of research in India. A variety of research has been done on data collected through police records covering a limited portion of highways. The analysis of such data can only reveal information regarding that portion only; but accidents are scattered not only on highways but also on local roads. A different source of road accident data in India is Emergency Management research Institute (EMRI) which serves and keeps track of every accident record on every type of road and cover information of entire State's road accidents. In this paper, we have used data mining techniques to analyze the data provided by EMRI in which we first cluster the accident data and further association rule mining technique is applied to identify circumstances in which an accident may occur for each cluster. The results can be utilized to put some accident prevention efforts in the areas identified for different categories of accidents to overcome the number of accidents.

Title 5: Extracting Hidden Patterns Within Road Accident Data Using Machine Learning Techniques

Author: KMA Solaiman, Md Mustafizur Rahman, and Nashid Shahriar. Avra

Road accident its may no best oppida together, 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 sweather, 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 featuresbetweenaccidents.Thispaperpresents theresultsoftheframeworkfromthe research study on road accident data of major national highways that pass through Krishna district for the year 2013 by applying machine learning techniques into analysis. These datasets collected from police stations are heterogeneous. Incomplete and erroneous values are corrected using data cleaning measures, and relevanceattributesareidentifiedusingattributeselectionmeasures.Clustersthatare formed using K-melodies, and expectation maximization algorithms are then analysed to discover hidden patterns using a priori algorithm. Results showed that the selected machine learning techniques are able to extract hidden patterns from the data. Density histograms are used for accident data visualization.

Title 6: k groups using the k-means clustering algorithm based on road accident frequency counts

Author: Sachin Kumar et al.

used data mining techniques to identify the locations where high frequency accidents are occurred and then analayze them to identify the factors that have an

effect on road accidents at that locations. The first task is to divide the accident location into k groups using the k-means clustering algorithm based on road accident frequency counts. Then, association rule mining algorithm applied in order to find out the relationship between distinct attributes which are in accident data set and according to that know the characteristics of locations

Title 7: Critical Analysis Reporting Environment (CARE) syste

Author: S. Shanthi et al.

proposed data mining classification technology based on gender classification, in which RndTree and C4.S use AdaBoost Meta classifier to provide high-precision results. From the Critical Analysis Reporting Environment (CARE) system provided by the Fatal Analysis Reporting System (FARS) used by the training data set.

Title 8: Kernel density estimation and K-means clustering to profile road accident hotspots.

Author: Tessa K. Anderson et al.

proposed a method of identifying high-density accident hotspots, which creates a clustering technique that determines that stochastic indices are more likely to exist in some clusters, and can therefore be compared in time and space. The kernel density estimation tool enables the visualization and manipulation of

density-based events as a whole, which in turn is used to create the basic spatial unit of the hotspot clustering method

Title 9: A Review of The Effect of Traffic sand Weather Characteristics on Road Safety.

Author: Yannis T.H.

Despite the existence of generally mixed evidence on the Effect of Traffic parameters, a few patterns can be observed. For instance, traffic flow seems to have a nonlinear relationship with accident rates, Even though some studies suggests linear relationship with accidents.

Title 10: An analysis of Accidents on Small portions NH-3 Indore to Dhamnod.

Author: K.Meshram and H.S. Goliya.

The data for analysis is collected for the period of 2009 to september 2011. More accidents occurred in Manpur region by faulty road geometry. The trend of accidents occurring in urban portions is more than 35% to rate total accidents in each year.

CHAPTER 4

PROJECT ALGORITHM

4.1 ALGORITHM

K-means clustering is a simple unsupervised learning algorithm that is used to solve clustering problems. It follows a simple procedure of classifying a given data set into a number of clusters, defined by the letter "k," which is fixed beforehand. The clusters are then positioned as points and all observations or data points are associated with the nearest cluster, computed, adjusted and then the process starts over using the new adjustments until a desired result is reached.

K-means clustering has uses in search engines, market segmentation, statistics and even astronomy. It is used mainly in statistics and can be applied to almost any branch of study. For example, in marketing, it can be used to group different demographics of people into simple groups that make it easier for marketers to target. Astronomers use it to sift through huge amounts of astronomical data; since they cannot analyze each object one by one, they need a way to statistically find points of interest for observation and investigation.

4.2 DATA FLOW DIAGRAM

Data Flow Diagram (DFD) is a two-dimensional diagram that describes how data is processed and transmitted in a system. The graphical depiction recognizes each source of data and how it interacts with other data sources to reach a mutual output. In order to draft a data flow diagram one must

- Identify external inputs and outputs
- Determine how the inputs and outputs relate to each other
- Explain with graphics how these connections relate and what they result in.

4.3 Role of DFD:

- It is a documentation support which is understood by both programmers and nonprogrammers. As DFD postulates only what processes are accomplished not how they are performed.

- A physical DFD postulates where the data flows and who processes the data.

- It permits analyst to isolate areas of interest in the organization and study them by examining the data that enter the process and viewing how they are altered when they leave.

4.4 UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

4.4.1 GOALS:

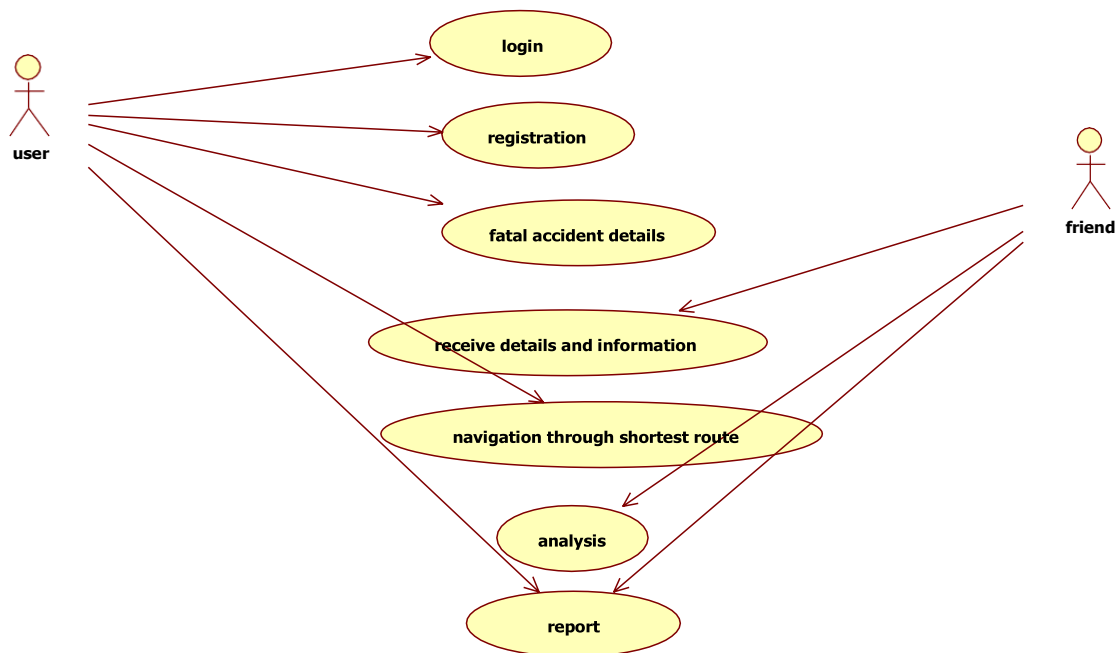
The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.

5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.
7. Integrate best practices.

4.5 USE CASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



4.5.1 DATA FLOW DIAGRAM

4.6 SYSTEM

ANALYSIS SYSTEM

REQUIREMENTS

4.6.1 HARDWARE REQUIREMENTS:

- System - Pentium-IV
- Speed - 2.4GHZ
- Hard disk - 40GB
- Monitor - 15VGA color
- RAM - 512MB

4.6.2 SOFTWARE REQUIREMENTS:

- Operating System - Windows XP
- Coding language - Python

CHAPTER 5

SOFTWARE ENVIRONMENT

5.1 Python:

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

- **Python is Interpreted** – Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- **Python is Interactive** – You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
- **Python is Object-Oriented** – Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
- **Python is a Beginner's Language** – Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

5.2 History of Python

Python was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands.

Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, SmallTalk, and Unix shell and other scripting languages.

Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).

Python is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.

5.3 Python Features

Python's features include –

- **Easy-to-learn** – Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
- **Easy-to-read** – Python code is more clearly defined and visible to the eyes.
- **Easy-to-maintain** – Python's source code is fairly easy-to-maintain.
- **A broad standard library** – Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
- **Interactive Mode** – Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
- **Portable** – Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
- **Extendable** – You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
- **Databases** – Python provides interfaces to all major commercial databases.

- **GUI Programming** – Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
- **Scalable** – Python provides a better structure and support for large programs than shell scripting.

Apart from the above-mentioned features, Python has a big list of good features, few are listed below –

- It supports functional and structured programming methods as well as OOP.
- It can be used as a scripting language or can be compiled to byte-code for building large applications.
- It provides very high-level dynamic data types and supports dynamic type checking.
- It supports automatic garbage collection.
- It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

Python is available on a wide variety of platforms including Linux and Mac OS X. Let's understand how to set up our Python environment.

Getting Python

The most up-to-date and current source code, binaries, documentation, news, etc., is available on the official website of Python <https://www.python.org>.

Windows Installation

Here are the steps to install Python on Windows machine.

- Open a Web browser and go to <https://www.python.org/downloads/>.

- Follow the link for the Windows installer python-XYZ.msifile where XYZ is the version you need to install.
- To use this installer python-XYZ.msi, the Windows system must support Microsoft Installer 2.0. Save the installer file to your local machine and then run it to find out if your machine supports MSI.
- Run the downloaded file. This brings up the Python install wizard, which is really easy to use. Just accept the default settings, wait until the install is finished, and you are done.

The Python language has many similarities to Perl, C, and Java. However, there are some definite differences between the languages.

First Python Program

Let us execute programs in different modes of programming.

Interactive Mode Programming

Invoking the interpreter without passing a script file as a parameter brings up the following prompt –

```
$ python
Python2.4.3(#1,Nov112010,13:34:43)
[GCC 4.1.220080704(RedHat4.1.2-48)] on linux2
Type"help","copyright","credits"or"license"for more information.
>>>
```

Type the following text at the Python prompt and press the Enter –

```
>>>print"Hello, Python!"
```

If you are running new version of Python, then you would need to use print statement with parenthesis as in **print ("Hello, Python!");**. However in Python version 2.4.3, this produces the following result –

```
Hello, Python!
```

Script Mode Programming

Invoking the interpreter with a script parameter begins execution of the script and continues until the script is finished. When the script is finished, the interpreter is no longer active.

Let us write a simple Python program in a script. Python files have extension **.py**. Type the following source code in a test.py file –

```
print"Hello, Python!"
```

We assume that you have Python interpreter set in PATH variable. Now, try to run this program as follows –

```
$ python test.py
```

This produces the following result –

```
Hello, Python!
```

5.4 ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system is well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

5.5 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

5.6 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity.

CHAPTER 6

MODULE IMPLEMENTATION AND DESCRIPTION

6.1 MODULES

- User interface module
- Login module
- Registration module
- Analysis module
- Report Accident module

6.2 User Interface Module

This module provides user an interface to sign up or login into the system. Once the user is logged in he can select Prediction or Analysis or Highway .In each of these modules the user can select certain attributes and make a prediction or analysis. Once the user has completed the analysis or prediction he can logout.

6.3 Login module

When the user of the system encounters the login page, he will be required to enter his username and his password to be able to log into the system. For a login attempt to be successful, the username and password combination input by the user must correspond to values that are available in the database.

6.4 Registration module

On the Report page, the intended user of the system is required to provide his/her username, password, Full name, phone number and email address. Upon entering these details, the data will be sent into the database. All fields are required for successful registration and if one field is left empty, a prompt will come up.

6.5 Analysis module

In this module the user specifies a particular attribute based on which an analysis With respect to its effect on accident can be determined. The analysis will be done using the training data set, Support Vector Machine classifier and the attribute specified by the user. The output will in form of a graph.

6.7 Report Accident module

The report page allows the user to interact with the site administrator by reporting the current state of victims and property involved in an accident as he travels along a particular route. This report is therefore sent to the incidence table of the data base, where it is checked and validated by the administrator and then updated into the various fields of the data base for appropriate report generation. The road's users are allowed to also report the accident occurrences on a particular road at any point in time. The road user can also make reports even if he does not have an account. The user clicks on the "Chat with Officer" link on the homepage and starts a chat with an office; he will be required to enter his name, his phone number, e-mail address and the description of the accident.

CHAPTER 7

SYSTEM TESTING

7.1 INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

7.2 OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the

user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

7.3 Types of testing

There are many types of testing like

- Unit Testing
- Integration Testing
- Functional Testing
- System Testing
- Stress Testing
- Performance Testing
- Usability Testing
- Acceptance Testing
- Regression Testing
- Beta Testing

Unit Testing

Unit testing is the testing of an individual unit or group of related units. It falls under the class of white box testing. It is often done by the programmer to test that the unit he/she has implemented is producing expected output against given input.

Integration Testing

Integration testing is testing in which a group of components are combined to produce output. Also, the interaction between software and hardware is tested in integration testing if software and hardware components have any relation. It may fall under both white box testing and black box testing.

Functional Testing

Functional testing is the testing to ensure that the specified functionality required in the system requirements works. It falls under the class of black box testing.

System Testing

System testing is the testing to ensure that by putting the software in different environments (e.g., Operating Systems) it still works. System testing is done with full system implementation and environment. It falls under the class of black box testing.

Stress Testing

Stress testing is the testing to evaluate how system behaves under unfavorable conditions. Testing is conducted at beyond limits of the specifications. It falls under the class of black box testing.

Performance Testing

Performance testing is the testing to assess the speed and effectiveness of the system and to make sure it is generating results within a specified time as in performance requirements. It falls under the class of black box testing.

Usability Testing

Usability testing is performed to the perspective of the client, to evaluate how the GUI is user-friendly? How easily can the client learn? After learning how to use, how proficiently can the client perform? How pleasing is it to use its design? This falls under the class of black box testing.

Acceptance Testing

Acceptance testing is often done by the customer to ensure that the delivered product meets the requirements and works as the customer expected. It falls under the class of black box testing.

Regression Testing

Regression testing is the testing after modification of a system, component, or a group of related units to ensure that the modification is working correctly and is not damaging or imposing other modules to produce unexpected results. It falls under the class of black box testing.

7.4 REQUIREMENT ANALYSIS

Requirement analysis, also called requirement engineering, is the process of determining user expectations for a new modified product. It encompasses the tasks that determine the need for analysing, documenting, validating and managing software or system requirements. The requirements should be documentable, actionable, measurable, testable and traceable related to identified business needs or opportunities and define to a level of detail, sufficient for system design.

FUNCTIONAL REQUIREMENTS

It is a technical specification requirement for the software products. It is the first step in the requirement analysis process which lists the requirements of particular software systems including functional, performance and security requirements. The function of the system depends mainly on the quality hardware used to run the software with given functionality.

Usability

It specifies how easy the system must be use. It is easy to ask queries in any format which is short or long, porter stemming algorithm stimulates the desired response for user.

Security The state of providing protected access to resource is security. The system provides good security and unauthorized users cannot access the system there by providing high security.

Reliability

It is the probability of how often the software fails. The measurement is often expressed in MTBF (Mean Time Between Failures). The requirement is needed in order to ensure that the processes work correctly and completely without being aborted. It can handle any load and survive and survive and even capable of working around any failure.

Safety

Safety is a measure taken to prevent trouble. Every query is processed in a secured manner without letting others to know one's personal information.

NON- FUNCTIONAL REQUIREMENTS

Portability

It is the usability of the same software in different environments. The project can be run in any operating system.

Performance

These requirements determine the resources required, time interval, throughput and everything that deals with the performance of the system.

Accuracy

The result of the requesting query is very accurate and high speed of retrieving information. The degree of security provided by the system is high and effective.

CHAPTER 8

CONCLUSION

As seen in statistics, association rule mining, and the classification, the environmental factors like roadway surface, weather, and light condition do not strongly affect the fatal rate, while the human factors like being drunk or not, and the collision type, have stronger affect on the fatal rate. From the clustering result we could see that some states/regions have higher fatal rate, while some others lower. We may pay more attention when driving within those risky states/regions. Through the task performed, we realized that data seems never to be enough to make a strong decision. If more data, like non-fatal accident data, weather data, mileage data, and so on, are available, more test could be performed thus more suggestion could be made from the data.

REFERENCE

[1] Divya Bansal and Lekha Bhambhu. Execution of Apriori algorithm of data mining directed towards tumultuous crimes concerning women. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(9), September 2013.

Amira A El Tayeb, Vikas Pareek, and Abdelaziz Araar. Applying association rules mining algorithms for traffic accidents in dubai. *International Journal of Soft Computing and Engineering*, September 2015.

[3] William M Evanco. The potential impact of rural mayday systems on vehicular crash fatalities. *Accident Analysis & Prevention*, 31(5):455–462, September 1999.

[4] K Jayasudha and C Chandrasekar. An overview of data mining in road traffic and accident analysis. *Journal of Computer Applications*, 2(4):32–37, 2009.

[5] S. Krishnaveni and M. Hemalatha. A perspective analysis of traffic accident using data mining techniques. *International Journal of Computer Applications*, 23(7):40–48, June 2011.

[6] Sachin Kumar and Durga Toshniwal. Analysing road accident data using association rule mining. In *Proceedings of International Conference on Computing, Communication and Security*, pages 1–6, 2015.

[7] Eric M Ossiander and Peter Cummings. Freeway speed limits and traffic fatalities in washington state. *Accident Analysis & Prevention*, 34(1):13–18, 2002.

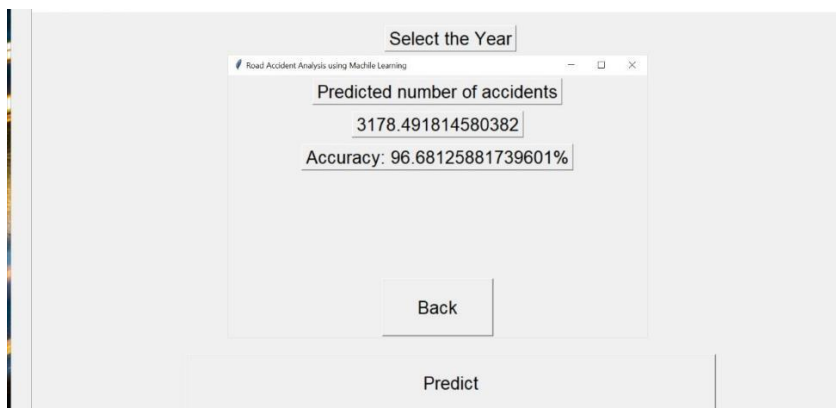
[8] KMA Solaiman, Md Mustafizur Rahman, and Nashid Shahriar. Avra Bangladesh collection, analysis & visualization of road accident data in Bangladesh. In *Proceedings of International Conference on Informatics, Electronics & Vision*, pages 1–6. IEEE, 2013.

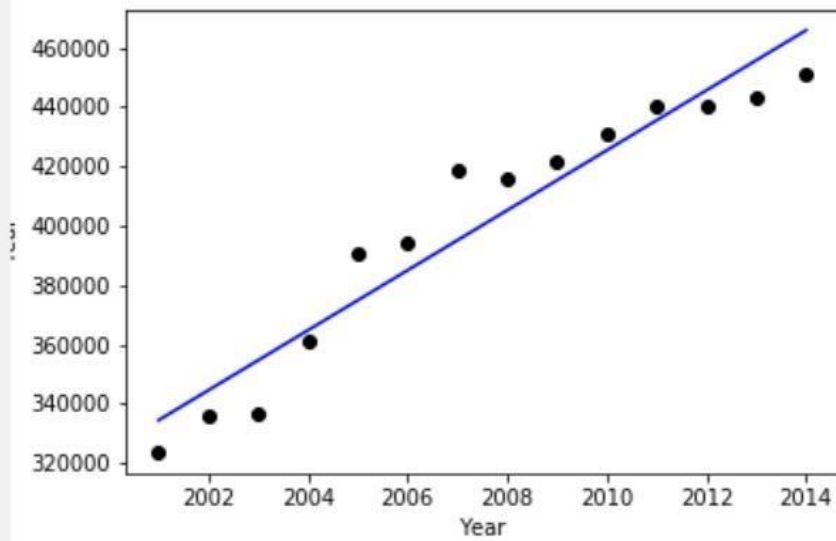
[9] Trac Integrated SCM & Project Management. Fatal Accidents Dataset. [https://wiki.csc.calpoly.edu/datasets/wiki/ HighwayAccidents](https://wiki.csc.calpoly.edu/datasets/wiki/HighwayAccidents)

[10] U.S. Census Bureau. Population Estimates. <http://www.census.gov/popest/data/historical/2000s/vintage 2007/>, 2007.

APPENDIX

SCREENSHOTS





Analysis

Road Accident Analysis using Machine Learning

Timespan of maximum accidents :15-18 hrs (Day)

Number of Accidents: 906639

Back

Year of maximum accidents

Analysis

Timespan of maximum accidents

Timespan of minimum accidents

Month of minimum accidents

Month of maximum accidents

Year of minimum accidents

Year of maximum accidents

Road Accident Analysis using Machine Learning

Predicted number of accidents

0.0

Accuracy: 87.08339596199393%

Back

Select the Year

2002

Select the State

Arunachal Pradesh

Select the Timespan

3-6 hrs. (Night)