

EXTRACTING HIDDEN PATTERNS WITHIN ROAD ACCIDENT DATA

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

by

POTLURI SRI TARUN (REG.NO. 37110578)
PYDI AKHIL (REG.NO. 37110610)



**DEPARTMENT OF COMPUTER SCIENCE
AND ENGINEERING
SCHOOL OF COMPUTING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade “A” by NAAC
JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119**

MARCH - 2021



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)



Accredited with “A” grade by NAAC
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600119
www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **POTLURI SRI TARUN (REG.NO. 37110578)**, **PYDI AKHIL (REG.NO. 37110610)** who have done the Project work as a team who carried out the project entitled “**EXTRACTING HIDDEN PATTERNS WITHIN ROAD ACCIDENT DATA**” under my supervision from **NOVEMBER 2020 to APRIL 2021**.

Internal Guide

Dr. P. SARDAR MARAN, M.E., Ph.D.

Head of the Department

Submitted for Viva voce Examination held on

Internal Examiner

External Examiner

DECLARATION

I, **POTLURI SRI TARUN (REG.NO. 37110578)** and **PYDI AKHIL (REG.NO. 37110610)** hereby declare that the Project Report entitled “**EXTRACTING HIDDEN PATTERNS WITHIN ROAD ACCIDENT DATA**” done by us under the guidance of **Dr. P. SARDAR MARAN, M.E., Ph.D.,** (Internal) is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE:

PLACE: Chennai-119.

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

We are pleased to acknowledge our sincere thanks to **Board of Management of SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project and for completing it successfully. We are grateful to them.

We convey our 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 & Engineering for providing us necessary support and details at the right time during the progressive reviews.

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

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

ABSTRACT

This report presents the results from the research study on applying largescale data mining methods into analysis of traffic accidents on the Finnish roads. The data sets collected from traffic fatal accidents are huge, multidimensional, and heterogeneous. Moreover, they may contain incomplete and erroneous values, which make its exploration and understanding a very demanding task. The target data of this study was collected by the Finnish Road Administration Datasets. The intention is to investigate the usability of robust clustering, association and frequent itemsets, 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.

TABLE OF CONTENTS

CHAPTER No.	TITLE	PAGE No.
	ABSTRACT	
	LIST OF FIGURES	
	LIST OF ABBREVIATIONS	
1	INTRODUCTION	1
	1.1 INTRODUCTION	1
	1.2 SCOPE OF PROJECT	2
	1.3 OBJECTIVE	2
2	LITERATURE SURVEY	3
	2.1 INTRODUCTION	3
3	SYSTEM DESIGN AND METHODOLOGY	7
	3.1 EXISTING SYSTEM	7
	3.2 PROPOSED SYSTEM	7
	3.3 SYSTEM ANALYSIS	8
	3.3.1 SYSTEM REQUIREMENT	8
	3.3.1.1HARDWARE REQUIREMENT	8
	3.3.1.2 SOFTWARE REQUIREMENT	
	3.4 APPLICATION OF JAVA	8
	3.5FEATURES OF JAVA	9
	3.6 COLLECTION FRAMEWORK	11
	3.7 MY SQL	14

	3.7.1SQL SERVER MANAGEMENT STUDIO	14
	3.7.2 CREATE NEW DATABASE	15
	3.7.3 ALGORITHM	16
	3.8 DATAFLOW DAIGRAM	
	3.8.1 ER DAIGRAM	18
	3.8.2 USE CASE DIAGRAM	18
	3.8.3 CLASS DIAGRAM	19
	3.8.4 ACTIVITY DIAGRAM	19
	3.8.5 SEQUENCE DIAGRAM	20
	3.9 MODULE	21
4	RESULT AND ANALYSIS	23
	4.1 STATISTICAL RESULT	23
	4.2 PIE CHART RESULT	24
	4.3ANALYSIS	25
5	CONCLUSION AND FUTURE WORK	26
	REFERENCES	27
	SCREENSHOTS	29
	SOURCECODE	32

LIST OF FIGURES

FIGURE No.	FIGURE NAME	PAGE No.
3.5.4	JAVA ARCHITECTURE	10
3.6	COLLECTION FRAMEWORK	12
3.7	MY SQL	13
3.7.1	SQL SERVER MANAGEMENT STUDIO	14
3.7.2	CREATE NEW DATABASE	15
3.7.3	SYSTEM ARCHITECTURE	17
3.8	ER DIAGRAM	18
3.8.2	USE CASE DIAGRAM	19
3.8.5	SEQUENCE DIAGRAM	20
3.8.5.1	COLLOBRATION DIAGRAM	20
4.1	BAR GRAPH	23
4.2.1	PIE CHART REPRESENTATION OF ACCIDENT PRONE ZONES	24
4.2.2	PIE CHART REPRESENTATION OF ACCIDENT PRONE VEHICLES.	25

LIST OF ABBREVIATIONS

FARS	FATAL ACCIDENT REPORTING SYSTEM
WEKA	WAIKATO ENVIRONMENT FOR KNOWLEDGE ANALYSIS
UCI	UNIQUE CLIENT IDENTIFIER
EMRI	EMERGENCY MANAGEMENT RESEARCH INSTITUTE
NASS	NATIONAL AUTOMOTIVE SAMPLING SYSTEM
GES	GENERAL ESTIMATE SYSTEM
GUI	GRAPHICAL USER INTERFACE
RDBMS	REALTIONAL DATABASE MANAGEMENT SYSTEM
UML	UNIFIED MODELING LANGUAGE

CHAPTER 1

INTRODUCTION

1.1 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.

1.2 SCOPE 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.

1.3 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

LITERATURE SURVEY

2.1 INTRODUCTION:

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

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

Author: divaybansal and lekhabhambhu

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 itemsets frequent and touncover the hidden information. This paper elaborates upon the use of associationrule 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 & Predictive Apriori Algorithm in which Apriori is better and faster than Predictive Apriori Algorithm.

Title2: 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.

Title3: 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 accidentdata 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 it may not be stoppable 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 weather, traffic on the road, load in the vehicle, type of road, health condition of driver, and speed can also be the reasons for accidents. Hidden patterns in accidents can be extracted so as to find the common features between accidents. This paper presents the results of the framework from the 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 relevance attributes are identified using attribute selection measures. Clusters that are formed using K-means, 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.

CHAPTER 3

SYSTEM DESIGN AND METHODOLOGY

3.1 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.

Drawbacks:

- 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

3.2 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.

ADVANTAGES OF PROPOSED SYSTEM:

- speed was available
- improve the performance of analysis in fatal and non-fatal accidents

3.3 SYSTEM ANALYSIS:

3.3.1 SYSTEM REQUIREMENTS:

3.3.1.1 HARDWARE REQUIREMENTS:

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

3.3.1.2 SOFTWARE REQUIREMENTS:

- Operating System - Windows XP
- Coding language - Java
- IDE - Net beans
- Database -MYSQL

3.4 APPLICATION OF JAVA:

Java is one of the world's most important and widely used computer languages, and it has held this distinction for many years. Unlike some other computer languages whose influence has weared with passage of time, while Java's has grown.

Java is widely used in every corner of world and of human life. Java is not only used in softwares but is also widely used in designing hardware controlling software components. There are more than 930 million JRE downloads each year and 3 billion mobile phones run java.

Following are some other usage of Java:

1. Developing Desktop Applications
2. Web Applications like Linkedin.com, Snapdeal.com etc
3. Mobile Operating System like Android
4. Embedded Systems
5. Robotics and games etc.

3.5 FEATURES OF JAVA:

The prime reason behind creation of Java was to bring portability and security feature into a computer language. Beside these two major features, there were many other features that played an important role in moulding out the final form of this outstanding language. Those features are;

1) Simple

Java is easy to learn and its syntax is quite simple, clean and easy to understand. The confusing and ambiguous concepts of C++ are either left out in Java or they have been re-implemented in a cleaner way.

Eg: Pointers and Operator Overloading are not there in java but were an important part of C++.

2) Object Oriented

In java everything is Object which has some data and behaviour. Java can be easily extended as it is based on Object Model.

3) Robust

Java makes an effort to eliminate error prone codes by emphasizing mainly on compile time error checking and runtime checking. But the main areas which Java improved were Memory Management and mishandled Exceptions by introducing automatic Garbage Collector and Exception Handling.

4) Platform Independent

Unlike other programming languages such as C, C++ etc. which are compiled into platform specific machines. Java is guaranteed to be write-once, run-anywhere language.

On compilation Java program is compiled into byte code. This byte code is platform independent and can be run on any machine, plus this byte code format also provide security. Any machine with Java Runtime Environment can run Java Programs.

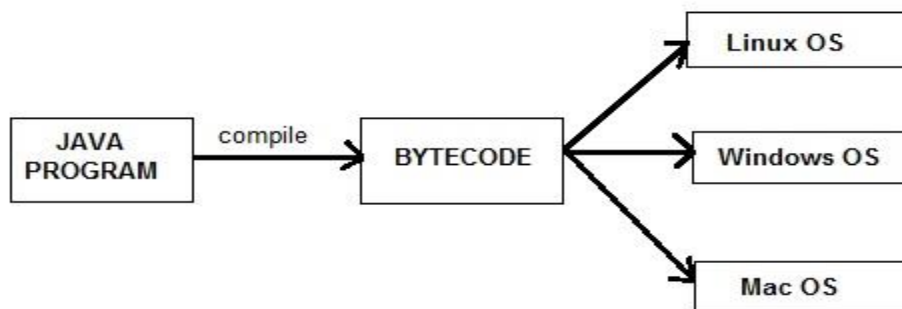


Fig:3.5.4: java architecture

5) Secure

When it comes to security, Java is always the first choice. With java secure features it enable us to develop virus free, temper free system. Java program always

runs in Java runtime environment with almost null interaction with system OS, hence it is more secure.

6) Multi-Threading

Java multithreading feature makes it possible to write program that can do many tasks simultaneously. Benefit of multithreading is that it utilizes same memory and other resources to execute multiple threads at the same time, like While typing, grammatical errors are checked along.

7) Architectural Neutral

Compiler generates byte codes, which have nothing to do with a particular computer architecture, hence a Java program is easy to interpret on any machine.

Portable

Java Byte code can be carried to any platform. No implementation dependent features. Everything related to storage is predefined, example: size of primitive data types

High Performance

Java is an interpreted language, so it will never be as fast as a compiled language like C or C++. But, Java enables high performance with the use of just-in-time compiler.

3.6 COLLECTION FRAMEWORK:

Collection framework was not part of original Java release. Collections was added to J2SE 1.2. Prior to Java 2, Java provided adhoc classes such as Dictionary, Vector, Stack and Properties to store and manipulate groups of objects. Collection framework provides many important classes and interfaces to collect and organize group of a objects



Fig:3.6:collection framework

3.7 MYSQL:

MySQL, officially, but also called "My Sequel" is the world's most widely used open-source relational database management system (RDBMS) that runs as a server providing multi-user access to a number of databases, though SQLite probably has more total embedded deployments. The SQL phrase stands for Structured Query Language.

The MySQL development project has made its source code available under the terms of the GNU General Public License, as well as under a variety of proprietary agreements. MySQL was owned and sponsored by a single for-profit firm, the Swedish company MySQL AB, now owned by Oracle Corporation. MySQL is a popular choice of database for use in web applications, and is a central component of the widely used LAMP open source web application software stack (and other 'AMP' stacks).

LAMP is an acronym for "Linux, Apache, MySQL, Perl/PHP/Python." Freesoftware-open source projects that require a full-featured database management system often use MySQL.

For commercial use, several paid editions are available, and offer additional functionality. Applications which use MySQL databases include: TYPO3, MODx, Joomla, WordPress, phpBB, MyBB, Drupal and other software. MySQL is also used in many high-profile, large-scale websites, including Wikipedia, Google (though not for searches), Facebook, Twitter, Flickr and YouTube. Microsoft is the vendor of SQL Server. We have different editions of SQL Server, where SQL Server Express is free to download and use. SQL Server uses T-SQL (Transact-SQL). T-SQL is Microsoft's

proprietary extension to SQL. TSQL is very similar to standard SQL, but in addition it supports some extra functionality, built in functions, etc.

T-SQL expands on the SQL standard to include procedural programming, local variables, and various support functions for string processing, date processing, mathematics, etc. SQL Server consists of a Database Engine and a Management Studio (and lots of other stuff which we will not mention here). The Database engine has no graphical interface - it is just a service running in the background of your computer (preferable on the server). The Management Studio is graphical tool for configuring and viewing the information in the database. It can be installed on the server or on the client (or both).

SQL Server Management Studio can also be used to create a new database, alter any existing database schema by adding or modifying tables and indexes, or analyze performance. It includes the query windows which provide a GUI based interface to write and execute queries.

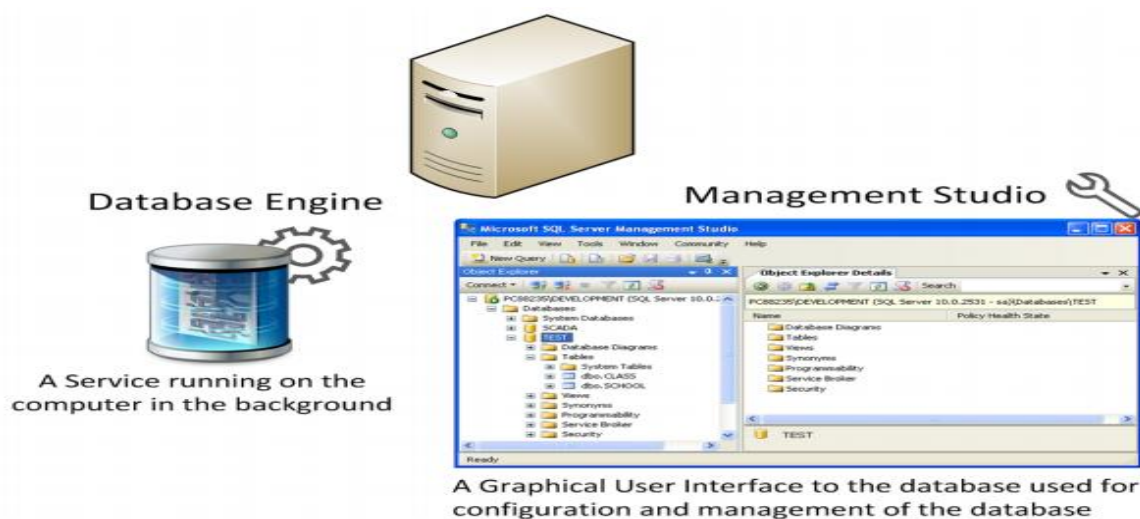


Fig:3.7:my sql

3.7.1 SQL SERVER MANAGEMENT STUDIO:

SQL Server Management Studio is a GUI tool included with SQL Server for configuring, managing, and administering all components within Microsoft SQL Server. The tool includes both script editors and graphical tools that work with objects and features of the server. As mentioned earlier, version of SQL Server Management Studio is also available for SQL Server Express Edition, for which it is known as SQL Server Management Studio Express.

A central feature of SQL Server Management Studio is the Object Explorer, which allows the user to browse, select, and act upon any of the objects within the server. It can be used to visually observe and analyze query plans and optimize the database performance, among others. SQL Server Management Studio can also be used to create a new database, alter any existing database schema by adding or modifying tables and indexes, or analyze performance. It includes the query windows which provide a GUI based interface to write and execute queries.

When creating SQL commands and queries, the “Query Editor” (select “New Query” from the Toolbar) is used (shown in the figure above). With SQL and the “Query Editor” we can do almost everything with code, but sometimes it is also a good idea to use the different Designer tools in SQL to help us do the work without coding (so much).

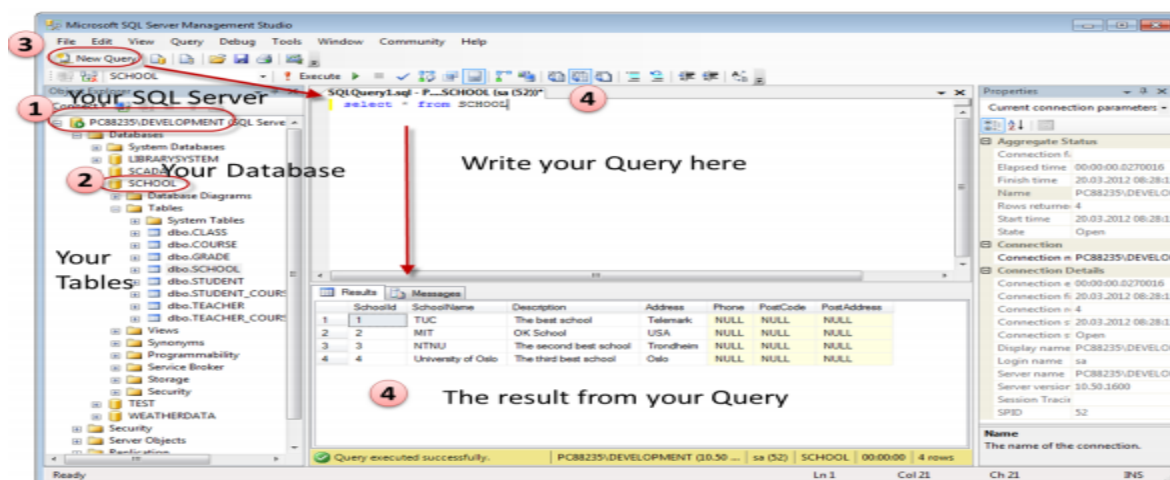


Fig:3.7.1:sql server management studio

3.7.2 CREATE A NEW DATABASE:

It is quite simple to create a new database in Microsoft SQL Server. Just right-click on the “Databases” node and select “New Database...”

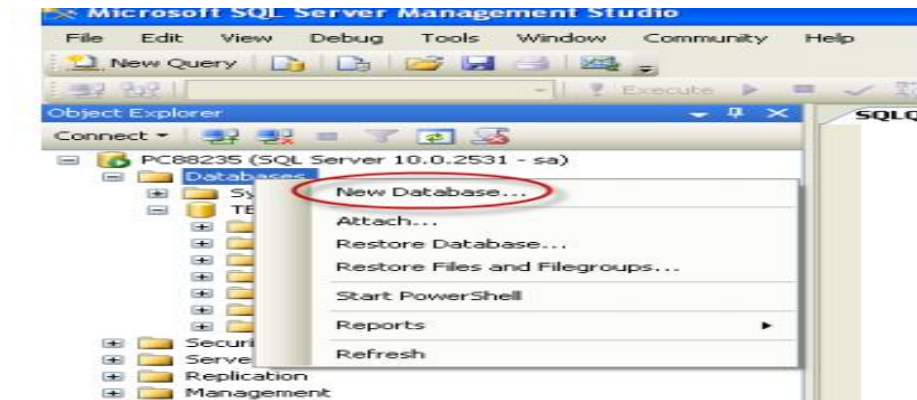


Fig:3.7.2:create new database

The OLAP Services feature available in SQL Server version 7.0 is now called SQL Server 2000 Analysis Services. The term OLAP Services has been replaced with the term Analysis Services. Analysis Services also includes a new data mining component. The Repository component available in SQL Server version 7.0 is now called Microsoft SQL Server 2000 Meta Data Services. References to the component now use the term Meta Data Services. The term repository is used only in reference to the repository engine within Meta Data Services

SQL-SERVER database consist of six type of objects,

They are,

1. TABLE
2. QUERY
3. FORM
4. REPORT
5. MACRO

Table

A database is a collection of data about a specific topic.

Views of Table

We can work with a table in two types,

1. Design View
2. Datasheet View

Design View

To build or modify the structure of a table we work in the table design view. We can specify what kind of data will be hold.

Datasheet View

To add, edit or analyses the data itself we work in tables datasheet view mode.

Query

A query is a question that has to be asked the data. Access gathers data that answers the question from one or more table. The data that make up the answer is either dynaset (if you edit it) or a snapshot (it cannot be edited). Each time we run query, we get latest information in the dynaset. Access either displays the dynaset or snapshot for us to view or perform an action on it, such as deleting or updating.

3.7.3ALGORITHM:

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.

SYSTEM ARCHITECTURE:

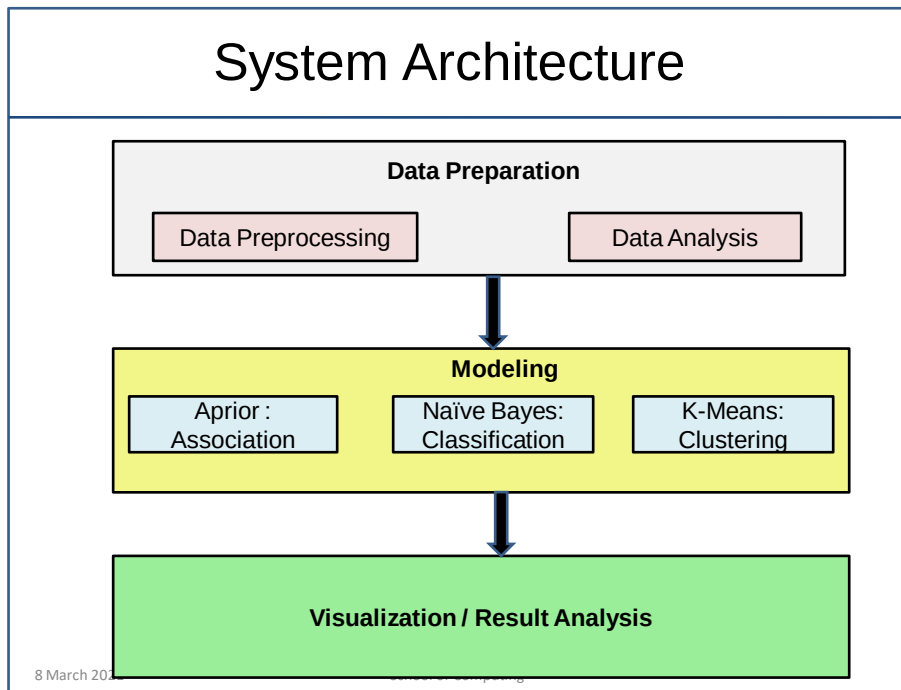


Fig:3.7.3:system architecture

3.8 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.

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.

3.8.1 ER diagrams

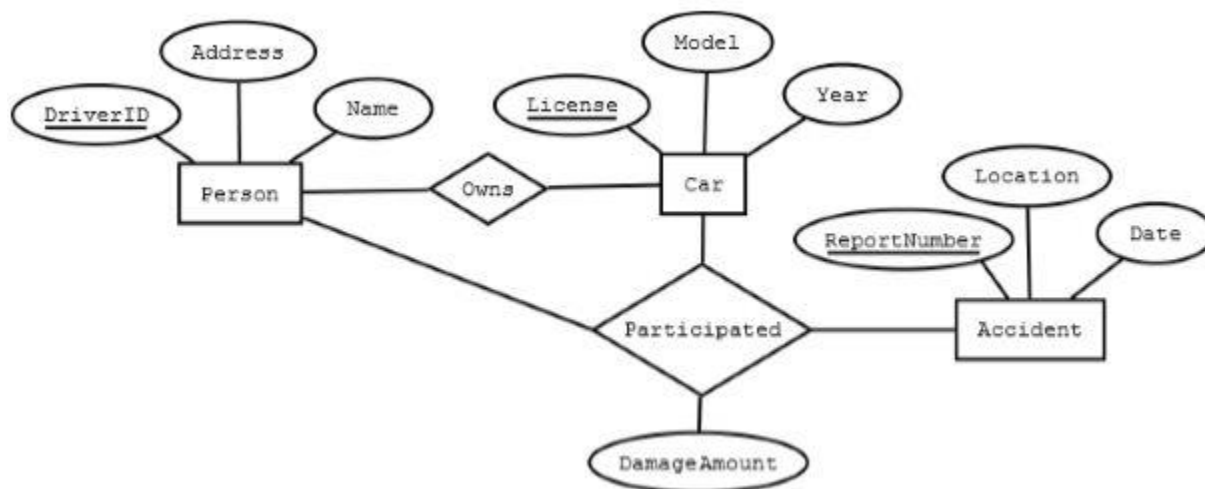


Fig:3.8:er daigram

3.8.2 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.

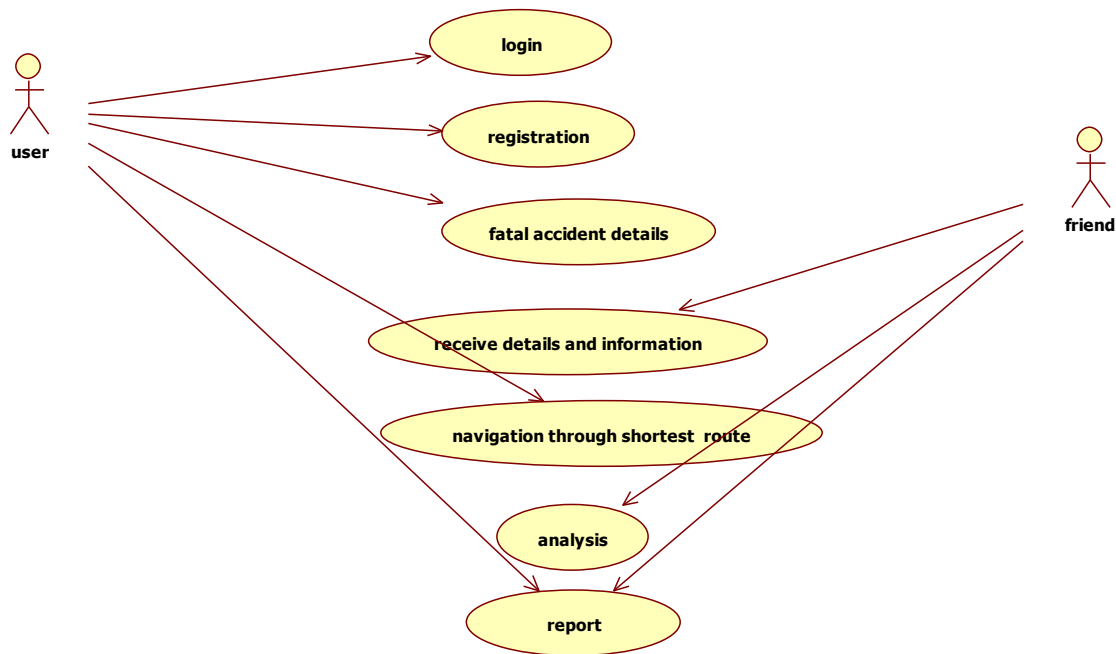


Fig:3.8.2: use case daigram

3.8.3 CLASS DIAGRAM:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

3.8.4 ACTIVITY DIAGRAM:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

3.8.5 SEQUENCE DIAGRAM:

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

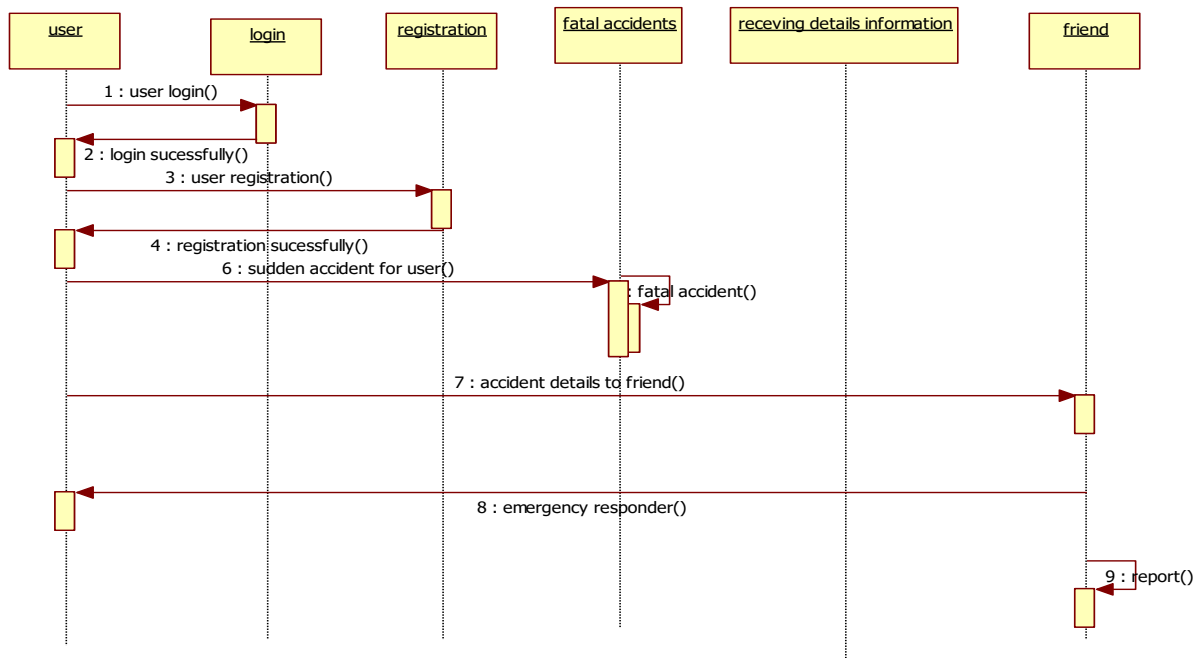


Fig:3.8.5:sequence diagram

COLLOBRATION DAIGRAM:

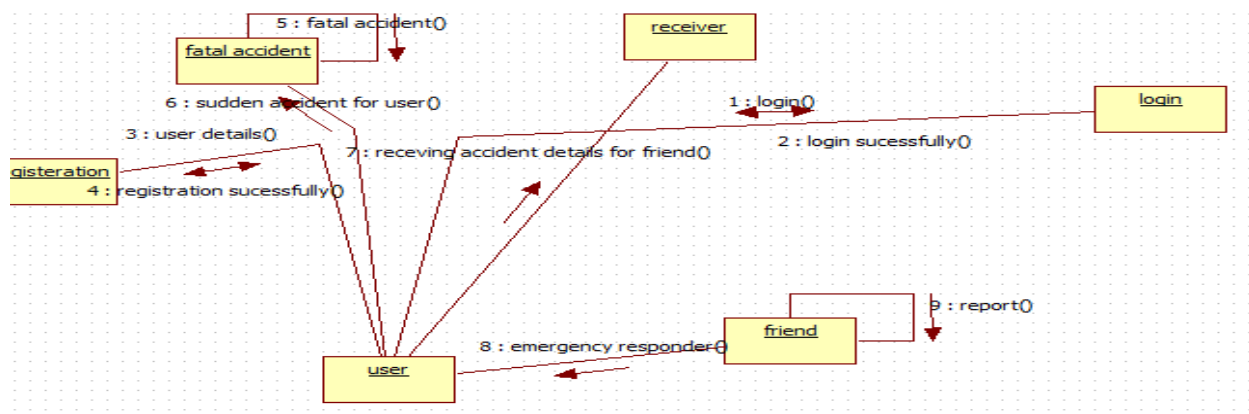


Fig:3.8.5.1: collaboration daigram

3.9 MODULE DESCRIPTION:

MODULES:

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

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.

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.

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.

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.

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. This information is sent to the database where it will be validated by the administrator and finally updated. Below are screenshots of the Road Traffic Accident Monitoring System

CHAPTER 4

RESULT AND ANALYSIS

4.1 Statistics results

The number of fatal accident in each year are shown in Fig.4.1 . The most fatal accidents happened in 2008 and the least in 2010.

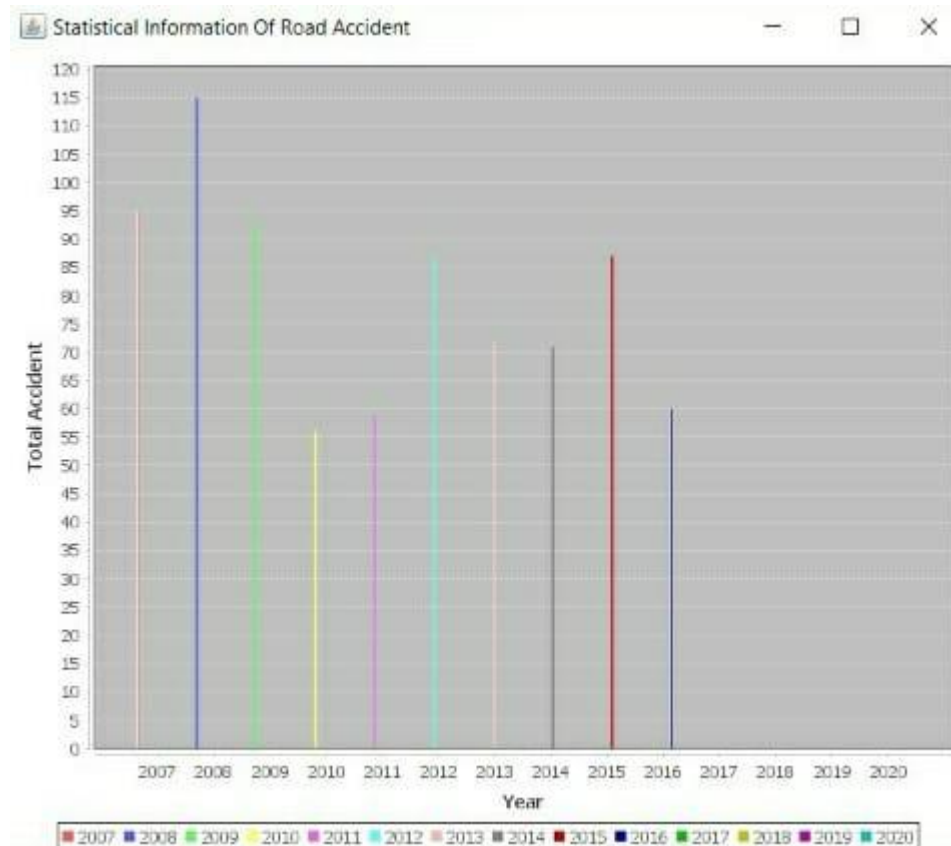
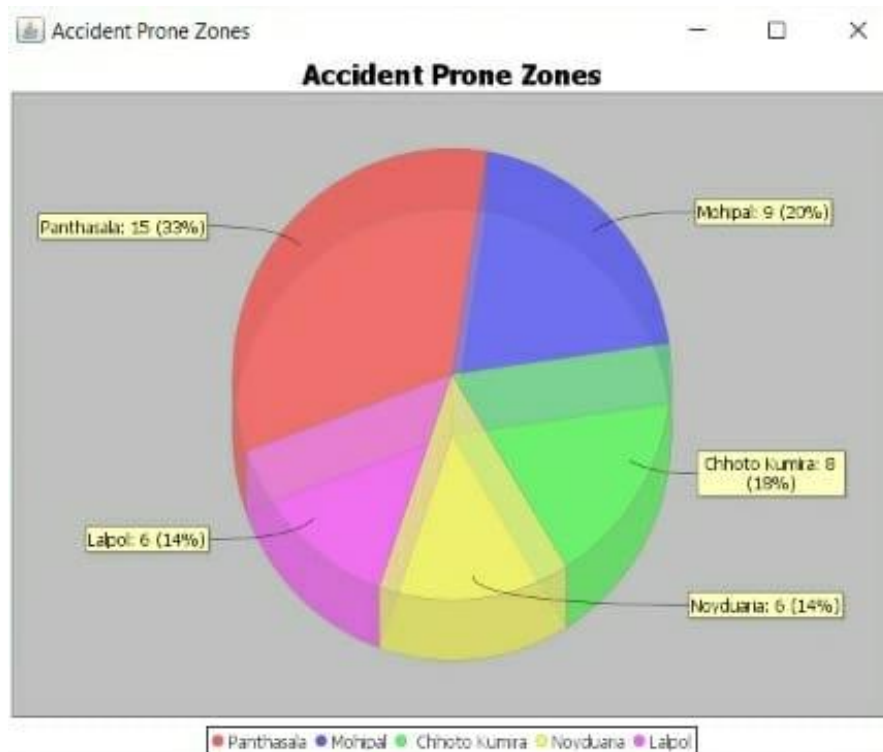


Fig 4.1 Bar Graph

4.2.1 Pie Chart Representation of accident prone zones

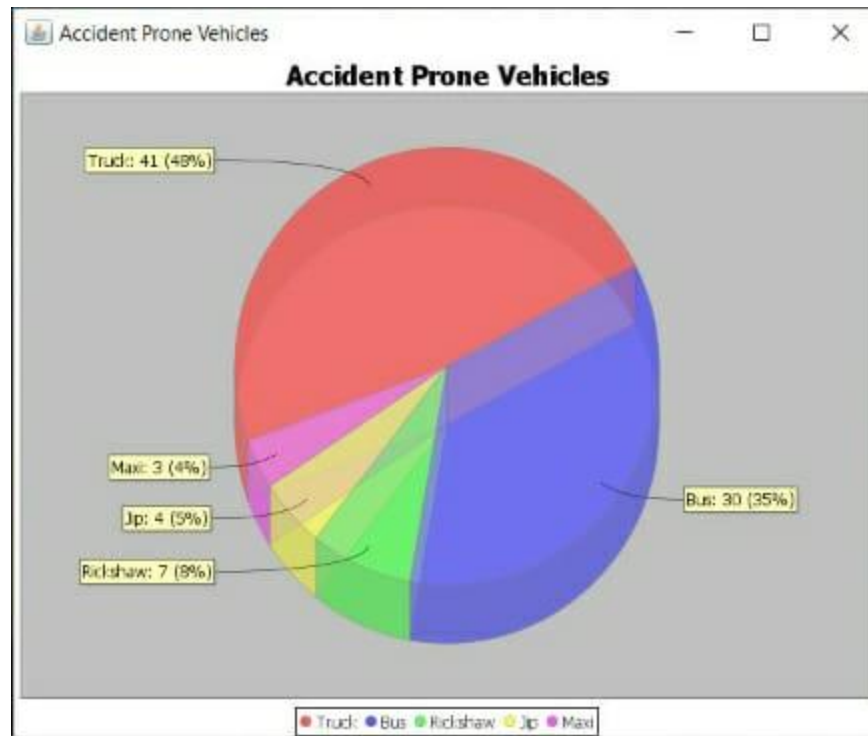
The percentage of accidents that took place in a particular place are represented in the Fig. 4.2.1 This pie chart gives the percentage of accidents taken place in particular locations.



4.2.1 Pie chart representation of Accident prone zones

4.2.2 Pie Chart Representation of accident prone vehicles

The percentage of accidents that took place with respect to a particular vehicle are represented in the Fig. 4.2.2 This pie chart gives the percentage of accidents taken place with respect to particular vehicles.



4.2.2 Pie chart representation of Accident prone vehicles

4.3 Analysis

From the clustering result we could see that some states/regions or vehicles have higher fatal rate, while some others lower. We may pay more attention when driving within those risky states/regions or vehicles.

CHAPTER 5

CONCLUSION

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.

REFERENCES

- [1] Divya Bansal and LekhaBhambhu. Execution of Apriori algorithm of data mining directed towards tumultuous crimes concerning women. International Journal of AdvancedResearch in Computer Science and Software Engineering, 3(9), September 2013.
- [2] Amira A El Tayeb, Vikas Pareek, and Abdelaziz Araar. Applying association rules mining algorithms for traffic accidents in dubai. International Journal of Soft Computingand 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 ComputerApplications, 2(4):32–37, 2009.
- [5] S. Krishnaveni and M. Hemalatha. A perspective analysis of traffic accident using data mining techniques. InternationalJournal of Computer Applications, 23(7):40–48, June 2011.
- [6] Sachin Kumar and Durga Toshniwal. Analysing road accidentdata using association rule mining. In Proceedings ofInternational Conference on Computing, Communication andSecurity, 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 InternationalConference 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>

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

APPENDIX

A. SCREEN SHOTS



TABLE NAME : LOCATION

For Information Storing

Location Id:

Location Name:

Lat:

Lon:

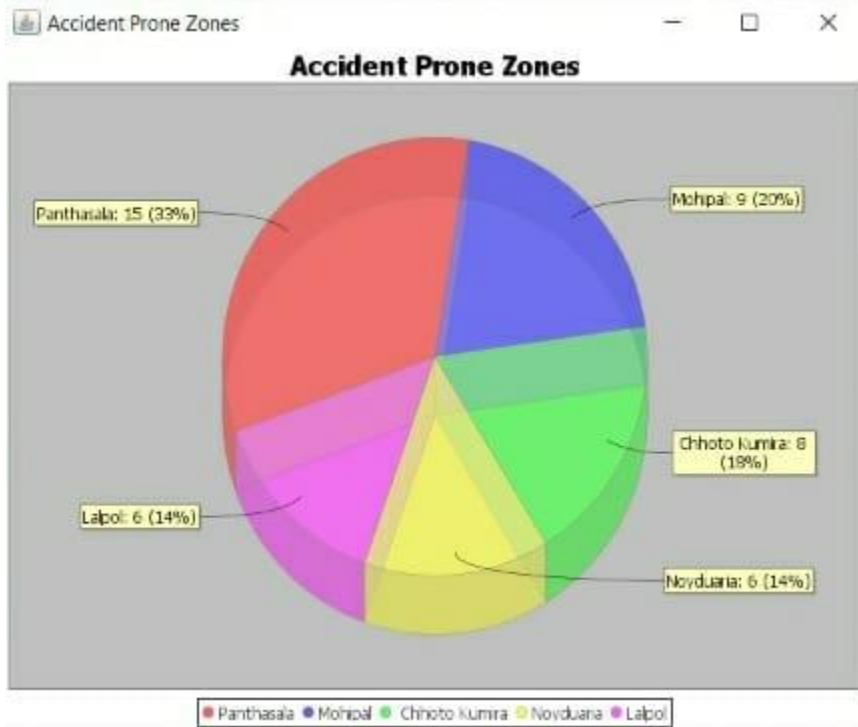
Information Removing

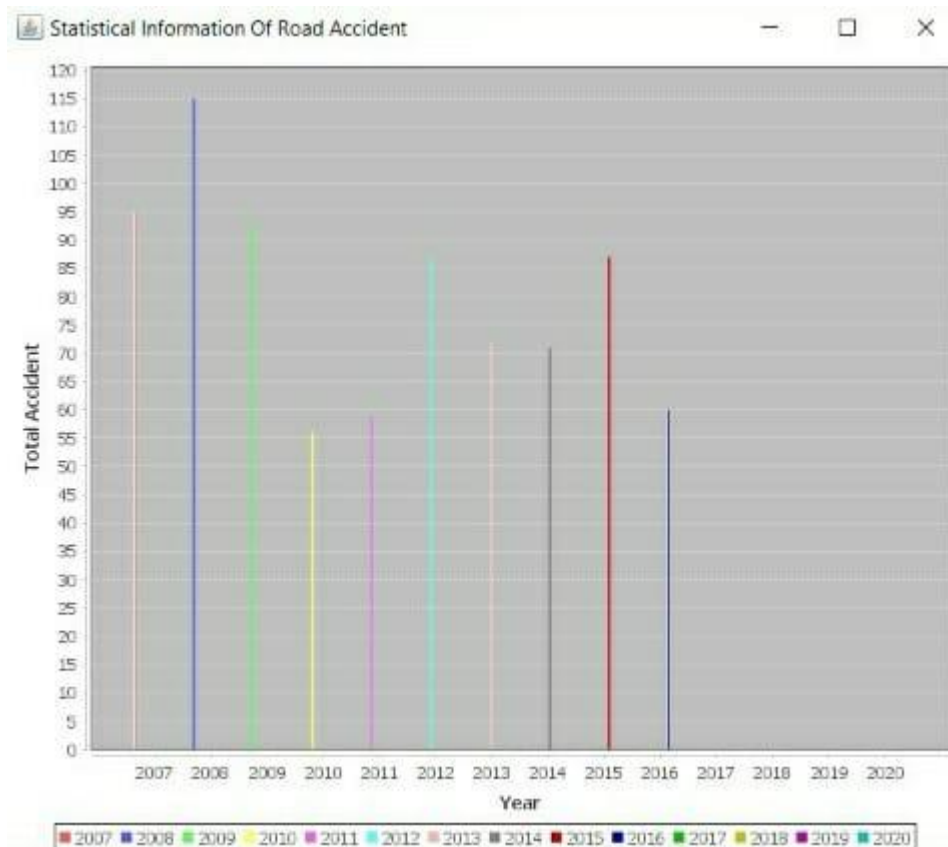
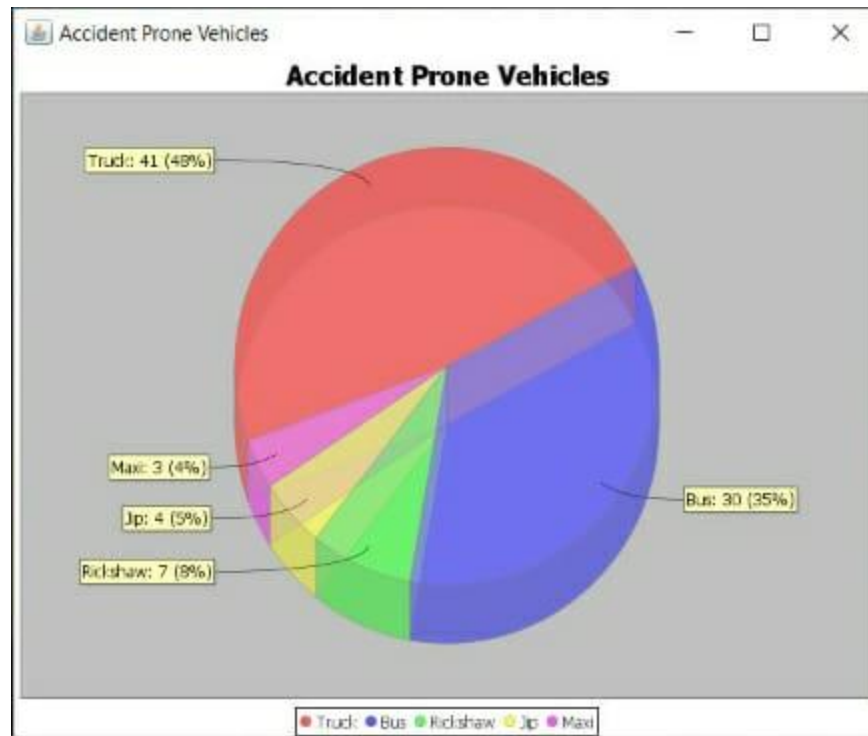
Select Location Name:

For Information Updating

Select Location Name:

New Location Name:





B. SOURCE CODE

```
package roadaccident;

import javax.swing.JOptionPane;

public class Login extends javax.swing.JFrame {

    /**
     * Creates new form Login
     */

    public Login() {
        initComponents();

        setDefaultCloseOperation(EXIT_ON_CLOSE);
    }

    /**
     * This method is called from within the constructor to initialize the form.
     * WARNING: Do NOT modify this code. The content of this method is always
     * regenerated by the Form Editor.
     */
    @SuppressWarnings("unchecked")
    // <editor-fold defaultstate="collapsed" desc="Generated Code">
    private void initComponents() {
```

```

jPanel1 = new javax.swing.JPanel();
tx1 = new javax.swing.JTextField();
l_bt1 = new javax.swing.JButton();
tx2 = new javax.swing.JTextField();
jLabel1 = new javax.swing.JLabel();
jLabel2 = new javax.swing.JLabel();
jLabel3 = new javax.swing.JLabel();

setDefaultCloseOperation(javax.swing.WindowConstants.EXIT_ON_CLOSE);
setPreferredSize(new java.awt.Dimension(1400, 1800));
getContentPane().setLayout(null);

jPanel1.setBackground(new java.awt.Color(0, 204, 153));

tx1.addActionListener(new java.awt.event.ActionListener() {
    public void actionPerformed(java.awt.event.ActionEvent evt) {
        tx1ActionPerformed(evt);
    }
});

l_bt1.setFont(new java.awt.Font("Meiryo UI", 1, 14)); // NOI18N
l_bt1.setText("Log In");

l_bt1.setBorder(javax.swing.BorderFactory.createTitledBorder(null, "",
    javax.swing.border.TitledBorder.DEFAULT_JUSTIFICATION,
    javax.swing.border.TitledBorder.DEFAULT_POSITION, new java.awt.Font("Tahoma", 0,
    13), new java.awt.Color(240, 240, 240))); // NOI18N

l_bt1.addActionListener(new java.awt.event.ActionListener() {
    public void actionPerformed(java.awt.event.ActionEvent evt) {
        l_bt1ActionPerformed(evt);
    }
});

```



```

.addContainerGap()

    );

    JPanel1Layout.setVerticalGroup(

JPanel1Layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)

.addGroup(JPanel1Layout.createSequentialGroup())

.addContainerGap()

.addGroup(JPanel1Layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)

.addComponent(tx1, javax.swing.GroupLayout.Alignment.TRAILING,
javax.swing.GroupLayout.PREFERRED_SIZE,
javax.swing.GroupLayout.DEFAULT_SIZE,
javax.swing.GroupLayout.PREFERRED_SIZE)

.addComponent(jLabel1, javax.swing.GroupLayout.PREFERRED_SIZE, 20,
javax.swing.GroupLayout.PREFERRED_SIZE))

.addGap(26, 26, 26)

.addGroup(JPanel1Layout.createParallelGroup(javax.swing.GroupLayout.Alignment.BASELINE)

.addComponent(tx2, javax.swing.GroupLayout.PREFERRED_SIZE,
javax.swing.GroupLayout.DEFAULT_SIZE,
javax.swing.GroupLayout.PREFERRED_SIZE)

.addComponent(jLabel2, javax.swing.GroupLayout.PREFERRED_SIZE, 20,
javax.swing.GroupLayout.PREFERRED_SIZE))

.addGap(17, 17, 17)

.addComponent(l_bt1)

.addContainerGap()

    );

getContentPane().add(JPanel1);

JPanel1.setBounds(410, 240, 380, 130);

```

```

        jLabel3.setIcon(new
javax.swing.ImageIcon(getClass().getResource("/roadaccident/roadc.jpg"))); // NOI18N
        jLabel3.setText("jLabel3");
        getContentPane().add(jLabel3);
        jLabel3.setBounds(0, -200, 1340, 1610);

```

```

pack();

```

```

} // </editor-fold>

```

```

private void l_bt1ActionPerformed(java.awt.event.ActionEvent evt) {

```

```

    String str1, str2;

```

```

    str1=tx1.getText();

```

```

    str2=tx2.getText();

```

```

    /*if(str1=="noman"&&str2=="min")

```

```

    {

```

```

    }*/

```

```

    if(str1.equals("admin")&&str2.equals("admin"))

```

```

    {

```

```

        setVisible(false);

```

```

        Admin1 a=new Admin1();

```

```

        a.setVisible(true);

```

```

    }

```

```

    else if(str1.equals("user")&&str2.equals("123"))

```

```

    {

```

```

setVisible(false);
Home_Page h=new Home_Page();
h.setVisible(true);
    }
    else
    {
OptionPane.showMessageDialog(null,"Incorrect User Name or Password");

    }

    // TODO add your handling code here:
}

private void tx1ActionPerformed(java.awt.event.ActionEvent evt) {
    // TODO add your handling code here:
}

/**
 * @param args the command line arguments
 */
public static void main(String args[]) {
    /* Set the Nimbus look and feel */
    //<editor-fold defaultstate="collapsed" desc=" Look and feel setting code (optional)
">

    /* If Nimbus (introduced in Java SE 6) is not available, stay with the default look
and feel.

    * For details see
http://download.oracle.com/javase/tutorial/uiswing/lookandfeel/plaf.html

```

```

        */
    try {
        for (javax.swing.UIManager.LookAndFeelInfo info :
            javax.swing.UIManager.getInstalledLookAndFeels()) {
            if ("Nimbus".equals(info.getName())) {
                javax.swing.UIManager.setLookAndFeel(info.getClassName());
                break;
            }
        }
    } catch (ClassNotFoundException ex) {
        java.util.logging.Logger.getLogger(Login.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
    } catch (InstantiationException ex) {
        java.util.logging.Logger.getLogger(Login.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
    } catch (IllegalAccessException ex) {
        java.util.logging.Logger.getLogger(Login.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
    } catch (javax.swing.UnsupportedLookAndFeelException ex) {
        java.util.logging.Logger.getLogger(Login.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
    }
}
//</editor-fold>

```

```

/* Create and display the form */
java.awt.EventQueue.invokeLater(new Runnable() {
    public void run() {
        new Login().setVisible(true);
    }
});

```

```
}  
// Variables declaration - do not modify  
private javax.swing.JLabel jLabel1;  
private javax.swing.JLabel jLabel2;  
private javax.swing.JLabel jLabel3;  
private javax.swing.JPanel jPanel1;  
private javax.swing.JButton l_bt1;  
private javax.swing.JTextField tx1;  
private javax.swing.JTextField tx2;  
// End of variables declaration  
}
```