ANALYSIS OF ROAD TRAFFIC FATAL ACCIDENTS USING APRIORI ALGORITHM

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

by

VIGHRAHALA AVINASH (36110110)

APPALA BRAHMANANDA GUPTA (36110092)

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

APRIL - 2020

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)
Accredited with "A" grade by NAAC
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119
www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **VIGHRAHALA AVINASH (36110110), APPALA BRAHMANANDA GUPTA (36110092)** who carried out the project entitled "**ANALYSIS OF ROAD TRAFFIC FATAL ACCIDENTS USING APRIORI ALGORITHM**" under my supervision from November 2019 to April 2020.

Internal Guide

Dr. S. PRINCE MARY M.E., Ph.D

Head of the Department

Dr. S. VIGNESWARI M.E., Ph.D

Submitted for Viva voce Examination held on________________________________

Internal Examiner **External Examiner**

DECLARATION

I **VIGHRAHALA AVINASH (36110110)** hereby declare that the Project Report entitled **"ANALYSIS OF ROAD TRAFFIC FATAL ACCIDENTS USING APRIORI ALGORITHM"** done by me under the guidance of **Dr. S. PRINCE MARY M.E., Ph.D.** 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 & 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. S. Prince Mary 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 & Engineering** who were helpful in many ways for the completion of the project

ABSTRACT

The results from the research study on applying large scale 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 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.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ALG	Algorithm
API	Application Program Interface Decades
DECA	Decades
EMRI	Emergency Management Research Institute
FARS	Fatality Analysis Reporting System
GES	General Estimates System
GNP	Gross National Product
MORTH	Ministry of Road Transport and Highways National
NASS	National Automotive Sampling System
RTA	Road Traffic Accidents
SQL	Structured Query Language

CHAPTER 1

INTRODUCTION

India has second largest road network in the world. Road accidents happen quite frequently and they claim too many lives every year. It is necessary to find the root cause for road accidents in order to avoid them. Suitable data mining approach has to be applied on collected datasets representing occurred road accidents to identify possible hidden relationships and connections between various factors affecting road accidents with fatal consequences. The results obtained from data mining approach can help understand the most significant factors or often repeating patterns. The generated pattern identifies the most dangerous roads in terms of road accidents and necessary measures can be taken to avoid accidents in those roads.

Accidents happened due to the negligence of driving vehicle on the roads. There are various reasons responsible for the accident like abandon of traffic rules but road conditions and the traffic are considered the one of prime cause of fatality and causality across the globe. These accidents occur due to dynamic design and development of automobile industries. A traffic crash happens due certain reasons like smashes of two vehicles on road, walking person, animal, or any other natural obstacles. It could result in injury, property damage, and death. Traffic accident analysis required study of the various factor affecting behind them. In survey it’s seen that approximate 1.2 million death and 50 million injuries estimated worldwide every year. The approximate estimation of causality and injuries due to poor road infrastructure is a big challenge before the living beings. The order to deal with the problem, in computational science, we can adopt data mining model for different scenario. In any vehicle accident, it studies about the driver’s behaviour, road infrastructure and possibilities of weather forecast that could be somewhere connected with different accident incidents. The main problem in the study and analysis of accident data is its mix heterogeneous environment and data segmentation which is used widely to overcome accident problem.

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 ALGs to discover the relationship in large amount of data. It is considered one of the most important tool in information technology in the previous DECA. Association rule mining ALG 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 ALG who main task is to find frequent item sets, 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 NHTSA. 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. Information about injuries from medical reports by passengers can be checked based on the passage of passengers who are in vehicles participating at the accident site. The data description can be found in the document FARS.

Road safety becomes a major public health concern when the statistics show that more than 3,000 people around the world happening death daily due to road traffic injury. In addition, road crashes lead to the global economic losses as estimated in road traffic injury costs of US$518 billion per year. The huge economic losses are an economic burden for developing countries. It is reflected that the road crash costs are estimated to be US$100 billion in developing countries which is twice the annual amount of development aid to such countries. Considering within South East Asian countries, the economic growth rate of Thailand continues to move upward with an aggravating road traffic situation due to the heavy negative impact of a higher level of motorization. Over 130,000 fatalities and nearly 500,000 people were permanently disabled due to road crashes over past decades. The economic losses due to the road crashes are; therefore, considerably high, costing approximately US$2,500 million per year (about US$0.3 million per hour), or 3.4 percent of the Gross National Product (GNP). An Asian Development Bank country report focused on the seriousness of the road accident problem upward trend of injuries per accident whereas fatalities per accident remained constant with small fluctuations from 1993-2002. However, the fatality index declined to 16 percent in 2002 from 27 percent in 1993 over this period of time. The collection and use of accurate and comprehensive data related to road accidents is very important to road safety management. It is considered one of the most important tool in information technology in the previous decades. Association rule mining ALG 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 ALG who main task is to find frequent item sets, which is the method we use to analyze the roadway traffic data The road accident data are necessary not only for statistical analysis in setting priority targets but also for in-depth study in identifying the contributory factors to have a better understanding of the chain-of-events.

Sometimes a lack of proper knowledge of crash and proper training of the police officers in charge on systematic data collection procedures from a crash scene adds to the diverging nature of the role of the police and the road safety professionals. These problems have become a burning issue for developing countries addressing road safety without completed crash data due to the negligence of the concerned authorities. A study clearly indicates this limitation - "the reactions are mostly on major accidents, but the interests would fade away rapidly and the problem still remains". The identification of factors affecting road crashes obtained from the crash investigation and reconstruction has not been conducted in practice in the Asian countries. The goal of this study was to initiate this road safety practice in by addressing the timely need for an in-depth study for road accidents. The accident investigation involves the inspection of crash scenes and the documentation of all necessary and available information of each component (i.e. human, vehicle, and road-environment). Accident reconstruction is defined by Baker and Fricke as "…the efforts to determine from whatever information is available, how the accident occurred". Accident reconstruction approach works backward from the evidence of the crash investigation and the remains of the crash to look into the scenario of before (pre crash), during (crash) and after the crash (post crash). The sequential analysis of end results to the initial condition of the events can establish "how" and "why" a particular type of crash occurs. Mathematics and Newtonian physics are applied in this analysis. It can be stated that crash reconstruction goes back to investigate the contributory factors and/or causes behind the crash event based on major and minor physical clues left behind at the crash scene. The techniques of crash reconstruction, trajectory and damage based analysis by using physics simplifies the determination of many important parameters of crash events. Moreover, to obtain a reliable conclusion, detailed information encompassing the system components needs to be thoroughly investigated. The information necessary for reconstruction starts with the crash scene. The answers to the questions of 'why', 'what', 'when' and 'how' should lead the reconstruction process to build up the real scenario of the pre-crash, crash, and post crash. Photographing of important clues and videotaping of the crash scene plays a vital role for the reconstruction. Injury information from occupant medical reports can be verified with the trajectory of the occupants found inside the involved vehicles at the scene. Therefore, an "open mind" investigative attitude is very crucial to search for all the detailed information from the scene. 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.

The growth of the population volume and the number of vehicles on the road cause congestion (jam) in cities that is one of the main transportation issues. Congestion can lead to negative effects such as increasing accident risks due to the expansion in transportation systems. The smart city concept provides opportunities to handle urban problems, and also to improve the citizens' living environment. In recent years, road traffic accidents (RTAs) have become one of the largest national health issues in the world

1.1 OUTLINE 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.

Different parameters such as junction type, collision type, location, month, time of occurrence, vehicle type could be visualized in a certain time sharp to see the how those parameters change and behave with respect to time. Based on those attributes one could also classify the type of accident. Using map API the system could be made more flexible such that it could find the safe stand dangerous roads

1.1.1 ADVANTAGES

- Takes less time
- Efficient
- Improve the performance of analysis in fatal and non-fatal accidents

1.1.2 DISADVANTAGES

- In this the cost of maintain and repairing the roads
- This will not useful for short distance level

CHAPTER 2

LITERATURE REVIEW

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. This support can be obtained from senior programmers, from book or from websites. 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: Execution of Apriori ALG of data mining directed towards tumultuous crimes concerning women

AUTHOR: divya bansal and lekha bhambhu

DESCRIPTION :

Apriori ALG is the most popular and useful ALG 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 item sets and to uncover the hidden information. Empirical results showed that the class association rules generated by Apriori ALG were more effective than those generated by Predictive Apriori ALG. More associations between accident factors and accident severity level were explored when applying Apriori ALG. 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 ALG in mining association rules from a dataset containing crimes data concerning women. As for this WEKA tool is used for extracting result. 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 ALG in which Apriori is better and faster than Predictive Apriori ALG

TITLE : Applying association rules mining ALGs for traffic accidents in dubai.

AUTHOR: Mira A El Tayeb, Vikas Pareek, and Abdelaziz Araar

DESCRIPTION:

Association rule mining ALGs 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 ALG 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 ALGs were applied to the dataset in order to explore the link between recorded accidents' factors to accident severity. Two sets of class association rules were generated using the two ALGs and summarized to get the most interesting rules using technical measures. Empirical results showed that the class association rules generated by Apriori ALG were more effective than those generated by Predictive Apriori ALG. More associations between accident factors and accident severity level were explored when applying Apriori ALG.

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

AUTHOR: S. Krishnaveni and M. Hemalatha

DESCRIPTION:

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 ALGs.

TITLE: Analysing road accident data using association rule mining

AUTHOR: Sachin Kumar and Durga Toshniwal

DESCRIPTION:

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. Expectation maximization ALGs are then analysed to discover hidden patterns using a priori ALG. 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. This paper 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 : Extracting Hidden Patterns Within Road Accident Data Using Machine Learning Techniques

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

DESCRIPTION:

Road accident 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 frame work 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 attributes election measures. Clusters that are formed using K-melodies, and expectation maximization ALGs are then analysed to discover hidden patterns using a priori ALG. 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.

2.1 OBJECTIVE

The purpose of this study was to conduct an in depth 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.

Road and traffic accidents (RTA) are one of the important problems in India. MORTH mentioned in its report that every year there are 0.4 million accidents reported in India, which makes India a country with large accident rate. This report shows that there is a negative trend of accidents from 2012 to 2013; however, as accidents are unpredictable and can occur in any type of situation, there is no guarantee that this trend will sustain in future also. Therefore, the identification of different geographical locations where most of the accidents have occurred and determining the various characteristics related to road accidents at these locations will help to understand the different circumstances of accident occurrence. Kannov and Janson stated that systematic relationship between accident frequency and other variables such as geometry of road, road side features, traffic information and vehicle information can help to develop effective accident prevention measures.

2.2 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 ALG on a dataset about traffic accidents which was gathered from Government Traffic Office, Apriori and Predictive Apriori association rules ALGs were applied to the dataset to investigate the connection between recorded accidents and factors to accident severity.

Accident cases in India are usually recorded by police officer of the region in which the accident has occurred. Also, the area covered by a police station is limited and they keep record of accidents that have occurred in their regions only. Ponnaluri discussed that the report prepared by police only contains the basic information that are not much useful for the research purpose. He suggests that data collection method used by police needs a lot of improvement. However, Indian researchers used these data and analyzed it for some highway portions using statistical methods. Data mining can be described as a novel technique to extract hidden and previously unknown information from the large amount of data. Several data mining techniques such as clustering, classification and association rule mining are widely used in the road accident analysis by researchers of other countries. Geurts et al. used association rule mining technique to understand the various circumstances that occur at high-frequency accident locations on Belgium road networks. Used adaptive regression tree model to build a decision support system for the road accidents. Developed various decision trees to extract different decision rules for different trees to analyze two-lane rural highway data of Spain. They found that bad light conditions and safety barriers badly affect the crash severity. Depaire et al. used clustering technique to analyze road accident data of Belgium and suggest that cluster-based analysis of road accident data can extract better information rather analyzing data without clustering. Used classification and regression tree

(CART) to analyze road accidents data of Iran and found that not using seat belt, improper overtaking and speeding badly affect the severity of accidents. Used Naive Bayes and decision tree classification ALG to analyze factor dependencies related to road safety. Severity of accident is directly concerned with the victim involved in accidents, and its analysis only targets the type of severity and shows the circumstances that affects the injury severity of accidents. Sometime accidents are also concerned with certain locations characteristics, which makes them to occur frequently at these locations. Hence identification of these locations where accident frequencies are high and further analyzing them is very much beneficial to identify the factors that affect the accident frequency at these locations.

2.3 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.

Here data mining techniques used to identify high-frequency accident locations and further analyzing them to identify various factors that affect road accidents at those locations. We first divide the accident locations into *k* groups based on their accident frequency counts using *k*-means clustering ALG. Then association rule mining ALG is applied on these to reveal the correlation between different attributes in the accident data and understand the characteristics of these locations. Hence, our main emphasis will be the interpretation of the outcomes.

CHAPTER 3

METHODOLOGY

Descriptive or predictive mining applied on previous road accidents data in combination with other important information as weather, speed limit or road conditions creates an interesting alternative with potentially useful and helpful outcome for all involved stakeholders. Association rule mining is used to analyze the previous data and obtain the patterns between road accidents. The two criterion used for association rule mining are support and confidence. Apriori ALG is one of the techniques to implement association rule mining. In the proposed system, we use apriori ALG to predict the patterns of road accidents by analyzing previous road accidents data.

K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. The ALG has a loose relationship to the k-nearest neighbor classifier, a popular machine learning technique for classification that is often confused with k-means because of the k in the name. One can apply the 1-nearest neighbor classifier on the cluster centers obtained by k-means to classify new data into the existing clusters.

Designed a multi- step that focuses on data structure software architecture, procedural details, ALG 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.

Road victims are a major global public health problem. Rapid motorization in low and middle income countries, poor quality of traffic safety systems and lack of institutional capacity to manage results contribute to the developing crisis.

In the event of an incident, police, ambulance or highway patrol will arrive at the location immediately and take security measures to prevent death. You will send the accident location to the traffic management system record and check how the accident is understood. All data is generated and stored.

This study is a step forward towards understanding the importance of applying machine learning ALGs for analyzing traffic accident data. A detailed data about accidents must be available for the findings to provide accurate and original facts about the accident occurred in any region. Data mining produces very understandable and useful results, but it can only be more effective if all needed details of data are available. The machine learning approach is more valuable when more attributes and the needed information about the accident is available. Information such as, seasonal limits, no passing zones, status/type of driving license, the number of years with license, apparent and sleepiness were missing in the data. It is important that accident data should contain all needed data or information in order to perfect the modeling of the framework. However, roads must have all needed traffic signs and regulation to prevent road traffic accidents, but they may not be effective enough to reduce accident if knowledge about the cause of the accident is not yet discovered. Therefore, the use of data mining is very important to discover this knowledge. In conclusion, most accidents seem to involve more male drivers than female drivers and it has been also observed that fatal accidents happen during sunny conditions in daylight. However, most accidents occur on Tuesdays than other days of the week. Many of the accidents take place in areas where there are no obstructions at all. While young people are mostly involved in accidents on Saturdays than older drivers. However, there are few cases of accidents involving animals, hence, the risk of death is not very high in such accidents.

Proportionally, old aged drivers are those who are likely involved in fatal accidents than in non-fatal accidents.

3.1 System Architecture Diagram

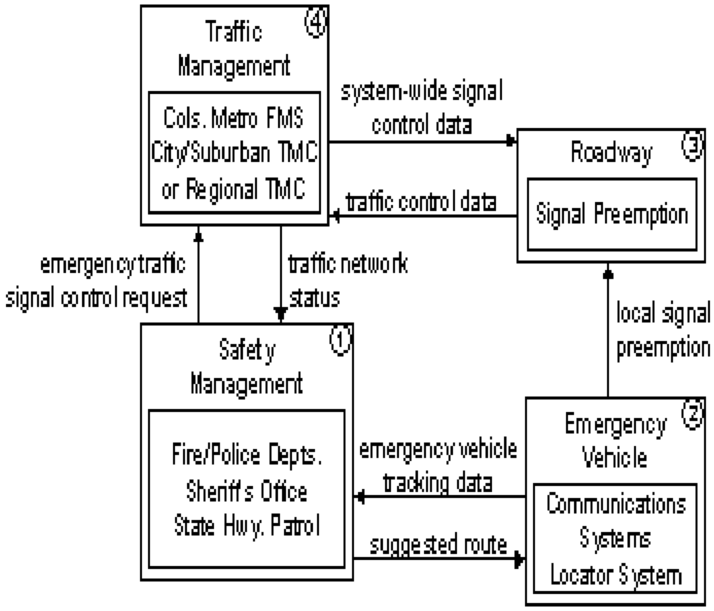


Fig 3.1 System Architecture

In the above Fig 3.1, the system architecture contains the different factors. The main factors are Traffic Management, Safety Management and Emergency Vehicle. The different factors are connected to provide the data. Safety Management consists of fire/police departments is connected to the Emergency Vehicle consists of the communication systems such as locator system and the Emergency Vehicle is connected to Roadway which consists of signal pre-emption and the Roadway is connected to Traffic Management consists of the different types of traffic management control such as metro traffic management control suburban management control.

When an accident occurs, immediately the police, ambulance or highway patrol officers will arrive at the spot and take the immediate safety measures to prevent death and sent the location of the accident to the traffic management system datasets and check the ways to know how the accident occurs and store all the data. The data is used to prevent the accident in that area and to maintain the safety measures. The data contains the location, time, type of vehicle, season etc..,

3.2 Flow Diagrams

3.2.1 ER Diagrams

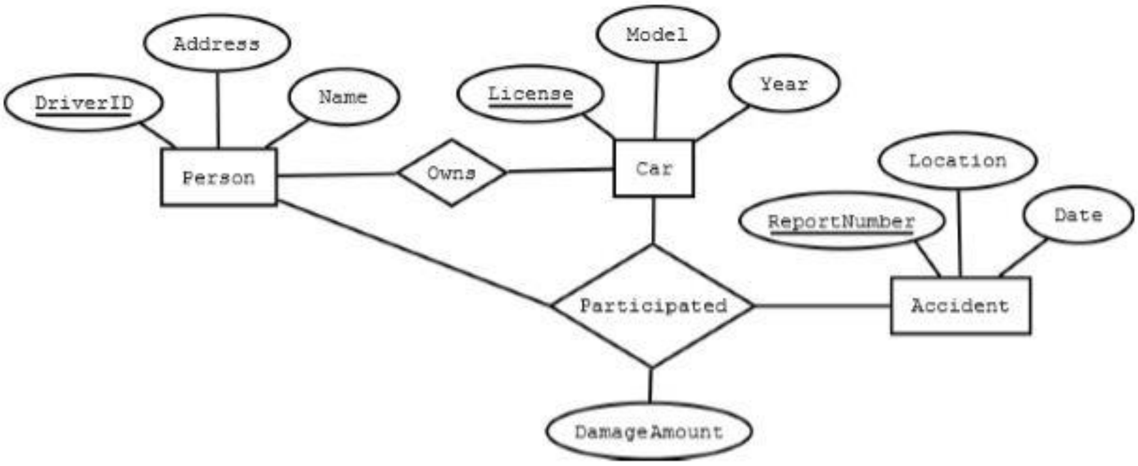


Fig 3.2 ER Diagram

In Fig 3.2, when a person met with an accident, The count of the vehicle which damage vehicle is involved in accident is increased. The data of the vehicle is taken such as date of the accident, location of the accident and give the report number. Also checks the vehicle details such as the vehicle number, year of buying, model of the vehicle and driver license and also enquire about the vehicle owner and upload all the details. After accident, the time of the accident, location of the accident, season of the accident such as summer, autumn etc., and also vehicle of the accident. It is easy to the user who is using this application to know about the details of the percentage of the accident of the region/location. The output will come as the bar graph and pie chart. It will shows the detail information of the accident region.

The other use case is the user and the user case is the one who checks the data. To check the data, the user need to login and can check the data. The user can check the region or location which he wants to go and then he will get of fatal accidents. The user will get the data in the percentage of the accidents and we can look into it. In that percentage bar, we can go and check the different modes such as the vehicle type, season etc., after the login of the user and search the details of the location, it will analyze and give the report and suggest the best route and prevents the accidents. It will very helpful to prevent the accidents.

3.3 ALGORITHMS

3.3.1 k-Means clustering

Clustering is an unsupervised data mining technique whose main task is to group the data objects into different clusters such that objects within a group are more similar than the objects in other clusters. *k*-means ALG is very popular clustering technique for numerical data. It groups the data objects into *k* clusters. There are various clustering ALGs existing but selection of suitable clustering ALG depends on type and nature of data. Our prime motive of this paper is to discriminate the accident location based on their frequency count. We have two choices to do this: First, we can decide a threshold level for each of the category of accident locations and group them in some categories. The problem with this approach is that it is very difficult to identify the number of categories of accident locations and decide a threshold level for each category. The other way is to use *k*-means ALG which can divide the accident locations into different groups. The number of groups can be identified using some cluster selection criteria such as gap statistic.

Initially, we have frequency counts of 87 locations with 7,027 road accidents. In order to divide the location into different groups we used *k*-means ALG. A brief formal description of laput: (D, AJ7D Output: k clusters Method:

The major problem with a clustering ALG is to identify the number of clusters to be made. The weakness of *k*-means clustering is that the user has to provide the value for *k*. An inappropriate value for *k* may lead to wrong clustering results. In this paper, we have used the gap statistics in order to find the value of *k* that can be supplied to divide the accident locations into different groups based on their frequency counts. Gap statistics can be used with any type of clustering technique, but they have been scarcely used to determine the number of clusters in road accident analysis.

Consider a data set D_{ij}, $i = 1,2,\ldots,m$, $j = 1,2,\ldots,n$, consisting of m data objects with values of n attributes. Assuming d_{xy} is the squared Euclidean distance between objects X and Y given by $d_{xy} = \sum (X_j - Y_j)^2$. If the data set has been clustered into k clusters, $c_1, c_2,\ldots,c_k$, where c_i indicates the ith cluster, then $n_i = |c_i|$.

Let $D_i = \sum d_{xy}$, (where $x, y \in c_i$) is the sum of pair-wise distances for all points in cluster i and W_k is the collective within cluster sum of squares around the cluster means and is given by Eq. (1). Gap$_n(k)$ can be defined as the difference between expected and observed values of $\log(W_k)$ and given in Eq. (2). K can be taken for the value maximizing Gap$_n(k)$.

Wk=∑i=1k(12ni)Di,Wk=∑i=1k(12ni)Di, (1)

Gapn(K)=E∗n{log(Wk)}−log(Wk),Gapn(K)=En∗{log(Wk)}−log(Wk), (2)

where E∗nEn∗ denotes the expectation under a sample size n from the reference distribution.

3.3.2 Association rule mining

Association rule mining is a very popular data mining technique based on market basket analysis that extracts interesting rules between various attributes in a large data set. Association rule mining produces a set of rules that define the underlying patterns in the data set. Given a data set *D* of *n* transactions where each transaction $T \in D$. Let $I = \{I_1, I_2, \ldots, I_n\}$ be a set of items. An item set *A* will occur in *T* if and only if $A \subseteq T$. $A \rightarrow B$ is an association rule, provided that $A \subset I$, $B \subset I$ and $A \cap B = \emptyset$. In case of road accident data, an association rule can identify the various attribute values which are responsible for an accident occurrence.

In association rule mining, various interesting measures are there to assess the quality of a rule. These interesting measures for a rule $A \rightarrow B$ are discussed as follows:

Support (S_p)

The support of a rule $A \rightarrow B$ defines the percentage how often *A* and *B* occur together in a data set and can be calculated using Eq. (3). Support is also known as frequency constraint. A set of items satisfying certain support threshold is known as frequent item set. These frequent item sets are further used to generate association rules based on other measures.

Confidence (C_f)

Confidence of a rule $A \rightarrow B$ defines the ratio of the occurrence of *A* and *B* together to the occurrence of *A* only and can be calculated using Eq. (4). Higher the confidence values of a rule $A \rightarrow B$, higher the chances of occurrence of *B* with the occurrence of *A*. Sometimes, only confidence values are not sufficient enough to evaluate the descriptive interest of a rule.

Lift (L_t)

Lift for a rule $A \rightarrow B$ measures the occurrence of *A* and *B* together than expected. In other words, lift is the ratio of the confidence and the expected confidence of a rule. Expected confidence can be defined as the occurrence of *A* and *B* together with the occurrence of *B*. A lift value ranges from 0 to ∞. Lift values greater than 1 make a rule potentially useful for predicting the consequent in future data sets. Lift determines how far from independence are *A* and *B*. Lift measures co-occurrence only and is also symmetric with respect to *A* and *B*. Lift can be calculated using Eq. (5).

Leverage (L_v)

Leverage for a rule $A \rightarrow B$ measures the difference of *A* and B appearing together in the data set and the expectation if *A* and *B* are statistically dependent [19]. It can be calculated using Eq. (6). The values for leverage range from [−0.25 to +0.25]. A leverage value 0 indicates that the variables are statistically independent. It will increase towards +1 if the variables occur more often together and will decrease towards −1 if one of the variables alone occurs more often.

Conviction (C_v)

Conviction is another measure that undertakes some of the weaknesses of confidence and lift. Conviction of a rule $A \rightarrow B$ compares the probability that *A* occurs without *B* if they are dependent with the actual frequency of the appearance of *A* without *B*. Conviction is not symmetric i.e. conviction ($A \rightarrow B$) ≠ conviction ($B \rightarrow A$). Conviction is rather inspired in the logical definition of implication and attempts to calculate the degree of implication of any rule. The value for conviction ranges within [0.5, ∞]. The values which are distant from 1 indicate interesting rules. In conviction, the supports of both antecedent and consequent are taken into account. It can be calculated using Eq. (7):

$$Sp=P(A\cap B)N,Sp=P(A\cap B)N, \qquad (3)$$

where *N* is the total number of accident records.

$$Cf=P(A\cap B)P(A),Cf=P(A\cap B)P(A), \qquad (4)$$

$$Lt=P(A\cap B)P(A)\times P(B),Lt=P(A\cap B)P(A)\times P(B), \qquad (5)$$

$$Lv=P(A\cap B)-P(A)\times P(B),Lv=P(A\cap B)-P(A)\times P(B), \qquad (6)$$

$$Cv=P(A)\times P(B)P(A\cap B), \qquad (7)$$

For a better understanding of the above concept, consider the following short example in Table 1, from road accident domain in which the set of items is *I* = {Fog, High Traffic, Speed > 100, Low Traffic, Fatal Accident}. In Table 1, 1 shows the presence of the item and 0 indicates the absence of the item. The major problem with a clustering ALG is to identify the number of clusters to be made. The weakness of *k*-means clustering is that the user has to provide the value for *k*. An inappropriate value for *k* may lead to wrong clustering results. In this paper, we have used the gap statistics in order to find the value of *k* that can be supplied to divide the accident locations into different groups based on their frequency counts. Gap statistics can be used with any type of clustering technique, but they have been scarcely used to determine the number of clusters in road accident analysis.

Determining The Number of Clusters

One problem of clustering techniques is how to determine the best number of expected clusters. K-means algorithm requires the user to enter the number of clusters k. In our framework we have used k-means algorithm to divide the accident frequency times associated with locations into different groups depending on the Elbow method [30] to determine the number of clusters K. The Elbow method is one of the optimal methods that depends on both the measure of similarities within a cluster and the parameters that used for partitioning. The idea of partitioning is to create clusters where the variation within a cluster is minimized. The quality of cluster can be measured by summing the squared distances between each object within the cluster

3.4 System Requirements

3.4.1 Hardware requirements:

- System - Pentium-IV
- Speed - 2.4GHZ

- Hard disk - 40GB
- Monitor - 15VGA colour
- RAM - 512MB

3.4.2 Software requirements

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

3.5 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.

3.6 ANALYSIS OF TRAFFIC ACCIDENTS

As stated earlier, the academic participation to develop an improved road-safety structure to reduce crashes has been enormous. Universities have conducted a plethora of research on various subjects that are linked to the road's traffic accidents and safety. Every state's department of transportation has provided universities with research funds to conduct studies to help improve the overall structure of the traffic system. In an accident, there are multiple contributing factors which account for a crash's risk and unforeseen mishaps. These factors include, but are not limited to, vehicle design, road design, road environment, the driver's speed, weather conditions, light conditions, and the driver's demeanor. Road crashes are the consequence of multiple factors and scenarios that cannot be avoided at the time of performing analysis and cannot be incorporated to yield an improved product to reduce accidents.

CHAPTER 4

RESULTS AND DISCUSSION

Even though, descriptive data mining methods have clearly been able to uncover reasonable information from Indian Region accident data sets. However, the results remain at a very general level and do not yet provide a detailed knowledge for traffic accident experts. This study is a step forward towards understanding the importance of applying machine learning ALGs for analyzing traffic accident data. A detailed data about accidents must be available for the findings to provide accurate and original facts about the accident occurred in any region. Data mining produces very understandable and useful results, but it can only be more effective if all needed details of data are available. The machine learning approach is more valuable when more attributes and the needed information about the accident is available. Information such as, seasonal limits, no passing zones, status/type of driving license, the number of years with license, apparent and sleepiness were missing in the data. It is important that accident data should contain all needed data or information in order to perfect the modeling of the framework. However, roads must have all needed traffic signs and regulation to prevent road traffic accidents, but they may not be effective enough to reduce accident if knowledge about the cause of the accident is not yet discovered. Therefore, the use of data mining is very important to discover this knowledge. In conclusion, most accidents seem to involve more male drivers than female drivers and it has been also observed that fatal accidents happen during sunny conditions in daylight. However, most accidents occur on Tuesdays than other days of the week. Many of the accidents take place in areas where there are no obstructions at all. While young people are mostly involved in accidents on Saturdays than older drivers. However, there are few cases of accidents involving animals, hence, the risk of death is not very high in such accidents. Proportionally, old aged drivers are those who are likely involved in fatal accidents than in non-fatal accidents.

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 ALGs 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 ALG 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 ALG who main task is to find frequent item sets, 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 reported to the NHTSA (NHTSA). The dataset provided is originally came from FARS (Fatal Accident Reporting System).

Data Mining is a computational technique to deal with large and complex data set and these data sets can be of normal, nominal and mixed. It is quite easy to use in variety of domain belong to science and management; also, it could be used in fraud identification and many more scientific cases as well as in accident severity problem. Partition of objects in a group of clusters or in a homogeneous set is a fundamental operation of data mining.

Clustering is a method to partition objects in a similar group. The k-means ALG having a good efficiency for clustering large data sets but restricted in forming clusters for real word data while working only on numerical data because it helps in reducing the cost function by altering the meaning of the clusters [1,3]. Data mining technique is recognized as reliable technique for analysis of traffic accident severity problem and finding factors behind them. Damage like property, people due to road accident is undesirable. Happened that road accident incidents are more common at certain places that can help in identifying factors behind them. 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 of US$518 billion per year.

4.1 SCREENSHOTS

We have used data sets of accidents occur in all areas, the data set includes type of vehicle, the type of season, reason, etc., and it analyze the dataset and produce the graphical format. These graphs are used to show the output in screenshots. Each of the graph shows the data in bar graph or pie chart. The pie chart graph contains the data like location wise, vehicle wise. The bar graph contains the data like year wise data. These graph contains the different modes such as the accident prone zone, accident vehicle, year of accident etc., it also consists of prevention methods to avoid the accidents.

4.1.1 *ACCIDENT PRONE ZONE*

Accident prone zone graph represents the number accidents occurred in the particular location, for each location it shows unique colour. Analyze the datasets and it shows in pie chart.

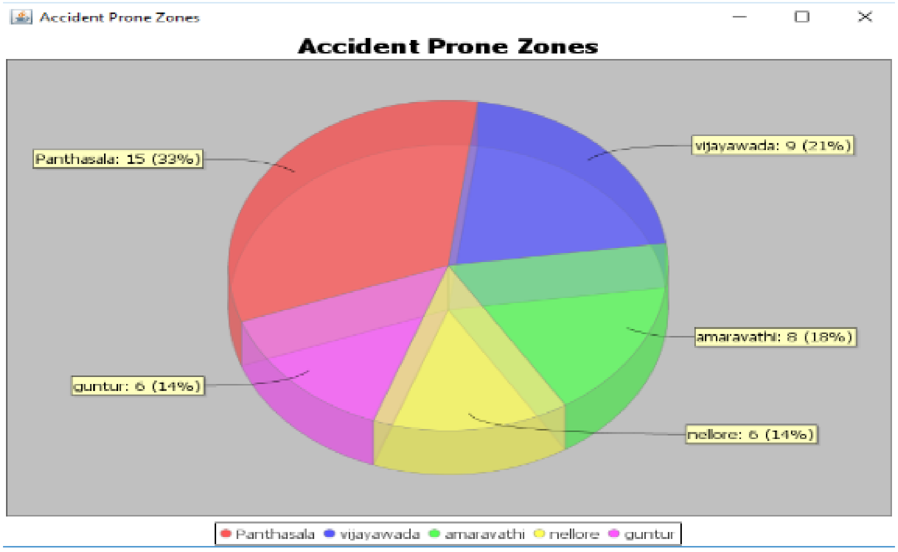


Fig 4.1. Accident prone zone

In the above Fig 4.1, the accident prone zone graph represents the number accidents occurred in the particular location, for each location it shows unique colour. Analyze the datasets and it shows in pie chart. This is the first page we got when we started to see the accident data. In this graph, the location wise data is available. Each location is indicated as different colour. Each colour box contains the name of the region or location along with percentage of the accident and average accidents per year.

4.1.2 *STATISTICAL INFORMATION*

Statistical information graph Analyses the data sets and it represent the Number of accidents occurred in the particular year, for each year it shows unique colour. X-axis represents the number of accidents and the Y-axis represents the particular year.

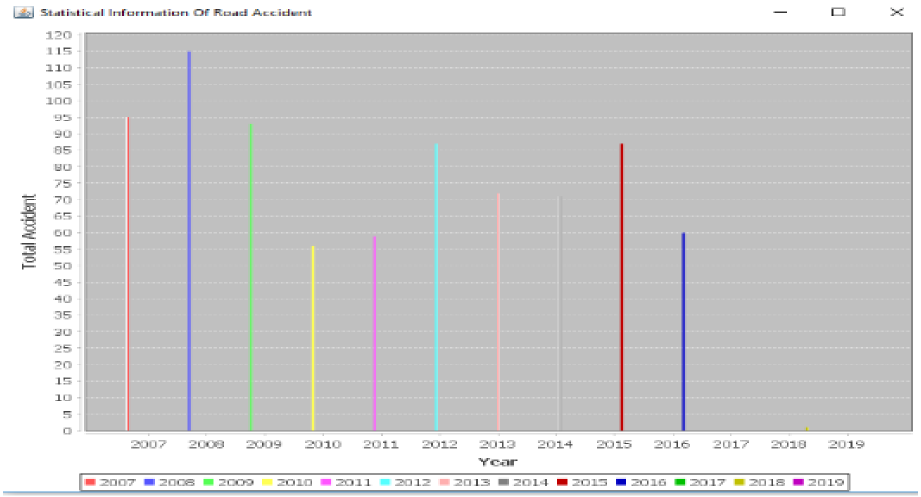


Fig 4.2. Statistical Information

In the above Fig 4.2, the statistical information graph Analyze the data sets and it represent the Number of accidents occurred in the particular year, for each year it shows unique colour. X-axis represents the number of accidents and the Y-axis represents the particular year. Each bar indicates the year and the graph bar differentiate by the separate colour.

4.1.3 *ACCIDENT PRONE VEHICLE*

Accident prone zone vehicle graph analyses the data and shows the highest vehicle got accident in the particular location, it represents the data in pie chart.

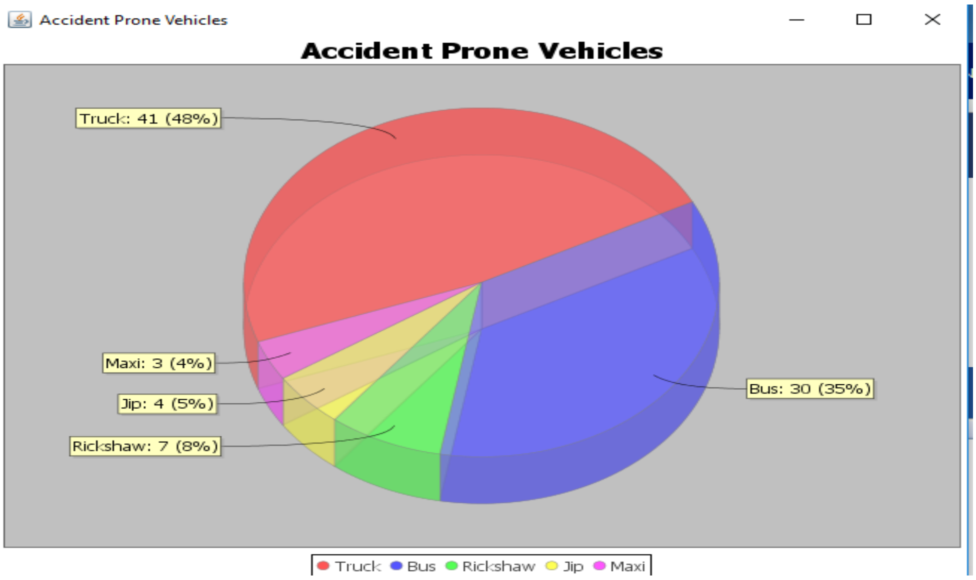


Fig 4.3. Accident Prone Vehicle

In the above Fig 4.3, the accident prone zone vehicle graph analyze the data and shows the highest vehicle got accident in the particular location which is shown in the above graph, here the above is drawn based on accident prone vehicles. Each colour indicates the different types of vehicles. The each vehicle type shown along with percentage and average vehicle accidents occurs.

4.1.4 *PREVENTION METHODS*

Prevention method shows the safety measures to avoid the accident. It shows the safety measures for two wheelers separately and four wheelers separately. Finally it analyses the vehicle type and shows the prevention methods in table format.

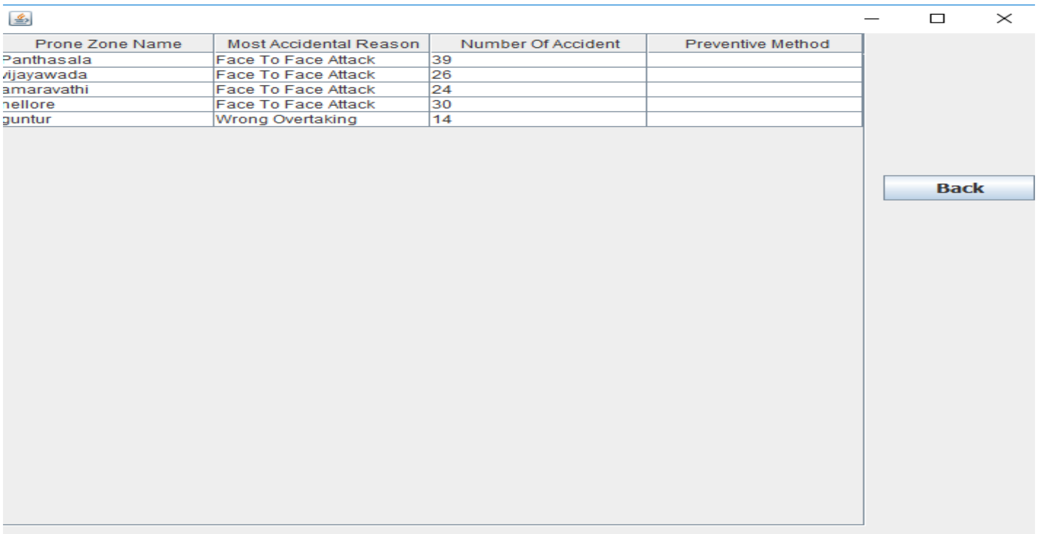


Fig 4.4. Prevention Method

In the above fig 4.4, Prevention method shows the safety measures to avoid the accident. It shows the safety measures for two wheelers separately and four wheelers separately. Finally it analyze the vehicle type and shows the prevention methods in table format.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 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 effect 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.

5.2 FUTURE WORK

- Message alert system if the vehicle reaches the accident prone zone
- Accident data automatically storage system using Internet Of Things
- Automatically alerts the emergency vehicles

REFERENCES

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

[2] Amira A El Tayeb, Vikas Pareek, and Abdelaziz Araar. Applying association rules mining ALGs 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 accidentdata 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

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

APPENDIX

Source Code

```
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */
package roadaccident;

import java.awt.HeadlessException;
import java.sql.Connection;
import java.sql.DriverManager;
import java.sql.PreparedStatement;
import java.sql.ResultSet;
import java.sql.SQLException;
import javax.swing.table.DefaultTableModel;



public class Accident_Table extends javax.swing.JFrame {

    /**
     * Creates new form Accident_Table
     */
    public Accident_Table() {
        initComponents();

            try {
             String url = "jdbc:mysql://localhost:3306/road_accident";
```

```
            String user = "root";

            String password = "root";

            Connection conn = (Connection) DriverManager.getConnection(url,
user, password);


            //DefaultTableModel model = (DefaultTableModel)
Reason_Table.getModel();

            DefaultTableModel
model=(DefaultTableModel)accident_table.getModel();

            String query2 = " SELECT * from accident ";


            PreparedStatement preparedStmt2 =
conn.prepareStatement(query2);

            preparedStmt2.execute();

            ResultSet rs2 = preparedStmt2.getResultSet();




            while (rs2.next()) {


              // model.addRow(new
Object[]{rs2.getInt("reason_id"),rs2.getString("reason_name")});

             model.addRow(new
Object[]{rs2.getInt("accident_no"),rs2.getString("date"),rs2.getInt("time_id"),r
s2.getInt("location_id"),rs2.getInt("type_id"),rs2.getInt("reason_id"),rs2.getIn
t("season_id"),rs2.getInt("injured"),rs2.getInt("killed")});


            }

            conn.close();

        } catch (SQLException | HeadlessException ex) {
```

```
            //System.out.println("Welcome");
            System.out.println(ex);

        }
    }

    /**
     * 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() {

        jScrollPane1 = new javax.swing.JScrollPane();
        accident_table = new javax.swing.JTable();
        act_bt1 = new javax.swing.JButton();
        act_bt2 = new javax.swing.JButton();


setDefaultCloseOperation(javax.swing.WindowConstants.EXIT_ON_CLOS
E);

        accident_table.setFont(new java.awt.Font("Tahoma", 1, 10)); //
NOI18N
        accident_table.setModel(new javax.swing.table.DefaultTableModel(
```

```
new Object [][] {

},
new String [] {
    "accident no", "date", "time_id", "location_id", "type_id",
"reason_id", "season_id", "injured", "killed"
}
));
jScrollPane1.setViewportView(accident_table);

act_bt1.setFont(new java.awt.Font("Tahoma", 1, 10)); // NOI18N
act_bt1.setText("Back");
act_bt1.addActionListener(new java.awt.event.ActionListener() {
    public void actionPerformed(java.awt.event.ActionEvent evt) {
        act_bt1ActionPerformed(evt);
    }
});

act_bt2.setFont(new java.awt.Font("Tahoma", 1, 10)); // NOI18N
act_bt2.setText("Back To Main Page");
act_bt2.addActionListener(new java.awt.event.ActionListener() {
    public void actionPerformed(java.awt.event.ActionEvent evt) {
        act_bt2ActionPerformed(evt);
    }
});

javax.swing.GroupLayout layout = new
javax.swing.GroupLayout(getContentPane());
```

```
getContentPane().setLayout(layout);
layout.setHorizontalGroup(
layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
    .addGroup(layout.createSequentialGroup()
        .addComponent(jScrollPane1, javax.swing.GroupLayout.PREFERRED_SIZE, 639, javax.swing.GroupLayout.PREFERRED_SIZE)
        .addGap(18, 18, 18)
.addGroup(layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING, false)
            .addComponent(act_bt2, javax.swing.GroupLayout.PREFERRED_SIZE, 133, javax.swing.GroupLayout.PREFERRED_SIZE)
            .addComponent(act_bt1, javax.swing.GroupLayout.DEFAULT_SIZE, javax.swing.GroupLayout.DEFAULT_SIZE, Short.MAX_VALUE))
        .addContainerGap())
);
layout.setVerticalGroup(
layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
    .addGroup(layout.createSequentialGroup()
.addGroup(layout.createParallelGroup(javax.swing.GroupLayout.Alignment.LEADING)
            .addGroup(layout.createSequentialGroup()
                .addGap(116, 116, 116)
                .addComponent(act_bt1)
                .addGap(37, 37, 37)
                .addComponent(act_bt2))
```

```
                .addGroup(layout.createSequentialGroup()
                    .addContainerGap()
                    .addComponent(jScrollPane1,
javax.swing.GroupLayout.PREFERRED_SIZE, 458,
javax.swing.GroupLayout.PREFERRED_SIZE)))
                .addContainerGap(javax.swing.GroupLayout.DEFAULT_SIZE,
Short.MAX_VALUE))
        );

        pack();
    }// </editor-fold>

    private void act_bt2ActionPerformed(java.awt.event.ActionEvent evt) {



         setVisible(false);
      Login l=new Login();
        l.setVisible(true);

        // TODO add your handling code here:
    }

    private void act_bt1ActionPerformed(java.awt.event.ActionEvent evt) {
        // TODO add your handling code here:

         setVisible(false);
      Accident ac=new Accident();
        ac.setVisible(true);
```

```
    }

    /**
     * @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(Accident_Table.class.getName()).log(jav
a.util.logging.Level.SEVERE, null, ex);
        } catch (InstantiationException ex) {

java.util.logging.Logger.getLogger(Accident_Table.class.getName()).log(jav
a.util.logging.Level.SEVERE, null, ex);
```

```
        } catch (IllegalAccessException ex) {

java.util.logging.Logger.getLogger(Accident_Table.class.getName()).log(java.util.logging.Level.SEVERE, null, ex);
        } catch (javax.swing.UnsupportedLookAndFeelException ex) {

java.util.logging.Logger.getLogger(Accident_Table.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 Accident_Table().setVisible(true);
            }
        });
    }
    // Variables declaration - do not modify
    private javax.swing.JTable accident_table;
    private javax.swing.JButton act_bt1;
    private javax.swing.JButton act_bt2;
    Private javax.swing.JScrollPane jScrollPane1;
    // End of variables declaration
}
```

Turnitin Originality Report

Processed on: 03-Mar-2020 9:09 PM CST
ID: 1268862970
Word Count: 1615
Submitted: 1

Similarity Index	Similarity by Source	
4%	Internet Sources:	3%
	Publications:	3%
	Student Papers:	2%

18 By 18 18

2% match (Internet from 24-Feb-2019)
https://www.ijrter.com/published_special_issues/01-12-2017/investigation-of-road-traffic-fatal-accidents-using-data-mining-techniques.pdf

1% match (Internet from 25-Feb-2020)
https://www.tandfonline.com/doi/full/10.1080/24751839.2018.1448205

1% match (Internet from 29-Oct-2019)
https://ijsiet.org/vol32018/11.pdf

1% match (student papers from 04-May-2018)
Submitted to Institute of Technology, Tralee on 2018-05-04

< 1% match (student papers from 07-Apr-2018)
Submitted to Indian Institute of Technology, Madras on 2018-04-07

Ananlysis of Road Traffic Fatal Accidents using Apriori Algorithm Appala Brahmananda Gupta, Vighrahala Avinash. Student Department of Computer Science and Engineering, Sathyabama Institute of Science And Technology, Chennai, India. Appala Brahmananda Gupta:- ab.gupta1548@gmail.com Vighrahala Avinash:- avinashvighrahala123@gmail.com Abstract The consequences of research on the utilization of broad information assortment techniques in auto collision investigation on Finnish streets. The informational indexes are gathered from traffic lethal mishaps that are enormous, multidimensional, and heterogeneous. Additionally, they can contain fragmented and wrong evaluations, make look into and comprehend troublesome assignments. The objective information of this examination was gathered by the Finnish Road Administration Datasets. The point is to explore the utilization of stable groupings, affiliations and components and strategies that are regularly used to envision auto collision investigation. While the outcomes show that the picked information extraction technique can make a justifiable information model, discovering increasingly point by point data can be improved with progressively nitty gritty data and complete informational collections. K implies that the calculation utilizes the measure of harm as a parameter to gather areas. At that point we utilized affiliation rule mining to describe these Surface Condition. The principles uncovered various variables related with street mishaps at various alcoholic and drive with changing mishap frequencies. The rules of affiliation of high-recurrence mishap area shows that crossing points on interstates are increasingly hazardous for each sort of deadly mishap. Keywords: Road

Analysis of Road Fatal Accidents Using Apriori Algorithm

Vighrahala Avinash[1*], Appala Brahmananda Gupta[2],
Dr. S.Prince Mary[3]

[1*,2] UG Student, Department of Computer Science and Engineering,
Sathyabama Institute of Science and Technology, Chennai, India
[3]Associate Professor, Department of Computer Science and Engineering,
Sathyabama Institute of Science and Technology, Chennai, India

[1]avinashvighrahala123@gmail.com, [2]ab.gupta1548@gmail.com,
[3]princemary.cse@sathyabama.ac.in

Abstract. The consequences of research on the utilization of road information assortment techniques in auto collision investigation on Finnish streets. The informational indexes are gathered from traffic lethal mishaps that are enormous, multidimensional, and heterogeneous. Additionally, they can contain fragmented and wrong evaluations, make look into and comprehend troublesome assignments. The objective information of this examination was gathered by the Finnish Road Administration Datasets. The point is to explore the utilization of stable groupings, affiliations and components and strategies that are regularly used to envision auto collision investigation. While the outcomes show that the picked information extraction technique can make a justifiable information model, discovering increasingly point by point data can be improved with progressively nitty gritty data and complete informational collections. K implies that the calculation utilizes the measure of harm as a parameter to gather areas. At that point we utilized affiliation rule mining to describe these Surface Condition. The principles uncovered various variables related with street mishaps at various alcoholic and drive with changing mishap frequencies. The rules of affiliation of high-recurrence mishap area shows that crossing points on interstates are increasingly hazardous for each sort of deadly mishap.
.

Keywords: Road Traffic Accidents, Apriori Algorithm, Safety Measurements.

1 Introduction

There is a great deal of vehicles driving on road consistently and traffic mishaps can occur whenever in anyplace. A few mishaps include the demise, which implies that the individuals kicked the bucket in the mishap[1,2]. As people, we as a whole need to evade a mishap and remain safe. To get some answers concerning more secure methods for driving, information mining innovation can be applied to an assortment of traffic information to discover significant data and in this manner give driving counsel[3,4]. Information mining utilizes a wide range of methods and calculations to find the relationship in a lot of information. It is viewed as one of the most significant apparatuses in data innovation in earlier decades. Rule mining of affiliation calculation is famous philosophy to distinguish the significant of relations between the information put away in huge databases and furthermore assumes a significant job in inconsistent itemset mining [5,6]. A traditional affiliation rule mining technique is the Apriori calculation whose fundamental errand is to find visit the datasets, which are the strategy we can use to investigate the road traffic information. The characterization in the information extraction technique expects to make a model (classifier) of preparing information records, by which information records with obscure class names can be grouped [7,8]. The Naive Baye's strategy is one of the exceptionally essential likelihood-based strategies for classification that depends on the Bayes' theory with the assumption of autonomy between each pair of factors.

We use FARS information for our exploration. The lethal mishap record contains every single deadly mishap on open streets in 2007 that have been accounted for to the National Road Traffic Safety Administration [9,10,11]. Informational collections were downloaded from the college of California Polytechnic State University and all the information initially originated from FARS. The informational collection contains 37,248 informational indexes and 55 characteristics. Information portrayals can be found in the FARS archive.

2 Related Work

Car crashes including information mining systems can possibly lesser mortality. Utilizing a street wellbeing database permits passing to be decreased by executing street security programs at the nearby and national level[12,13]. Grouping models to anticipate the seriousness of wounds brought about by auto collisions. Rules mining of affiliation calculation on a dataset about the auto collisions which are assembled from Government Traffic Department, Apriori affiliation rules calculations are applied to dataset to research the association between the recorded mishaps and the elements to mishap seriousness.

3 Proposed System

The seriousness of wounds because of car crashes with counterfeit neural systems and choice trees. We apply it to substantial informational collections acquired from the National General Sampling System (NES) from the National Sampling System (GASS). The test results show that choice trees are better than neural systems in all cases. Our examination investigation likewise shows that the three most significant factors in deadly wounds are: the utilization of a driver's safety belt, gentle path conditions and the utilization of liquor by the driver. Our trials likewise show that deadly and non-lethal injury models perform superior to different classes. The capacity to foresee lethal and non-deadly wounds is significant because the demise of the driver causes the most noteworthy monetary and social expenses for the network. And furthermore, the deadly injury relies upon different variables like Season, perceivability of the street in morning or night, Vehicle type, the purpose behind a mishap, area of the mishap. The motivation behind this investigation is as per the following: Identify the components that added to the contextual analysis dependent on the aftereffects of the PTP assessment and reproduction. Apply occasion investigation to the occasion connect to represent mishap situations dependent on accessible data.

Fig 1. Architecture Diagram

- **Flow Diagram**

The Data Flow Diagram is a two-dimensional graph that depicts how the information is handled and transmitted in a framework. Charts perceive every datum source and how it connects with other information sources to accomplish commonly advantageous outcomes. To draft an information stream chart, one must

- Identify outside sources of info and yields.
- Determine how the information sources and yields identify with one another.
- Use the chart to clarify how this connection is associated and what causes it.

Job of Data Flow Diagram:

It is documentation bolster that is comprehended by the two developers and nonprogrammers. DFD proposes just what procedures are practiced not how they are performed. A physical DFD hypothesizes where the information streams and who forms the information. It allows an expert to segregate the zones of enthusiasm for the association rule mining and to study them by looking at information provided that enter the procedure and the review how they are adjusted when they leave.

4 Results and Discussion

We have used data sets of accidents occur in all areas, the data set includes a type of vehicle, the type of season, reason, etc., and it analyses the dataset and produces the graphical format.

Accident Prone Zone Graph

Accident-prone zone graph represents the number of accidents occurs in a location, for each location it shows unique colour. Analyse the datasets and it shows in the pie chart figure 2.

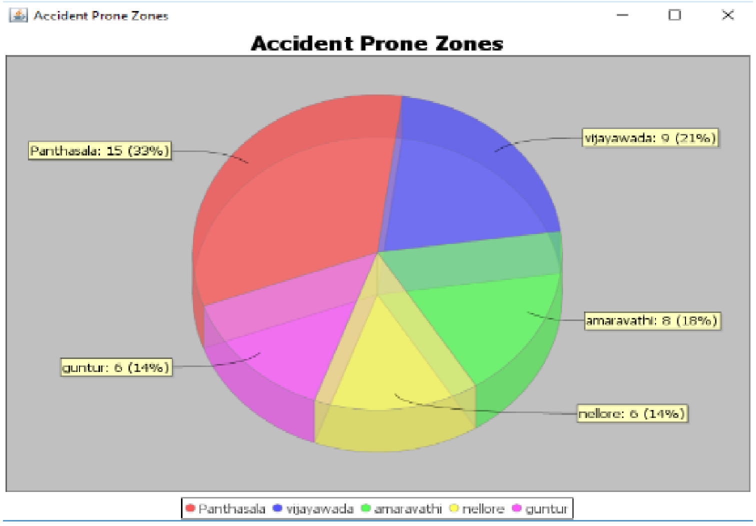


Fig.2. Accident-Prone Zone

Statistical Information

Factual data chart dissect the datasets and it speaks to the quantity of mishaps happens in the specific year, for every year it shows exceptional shading. The X-hub of the diagram speaks to the quantity of the mishaps and the Y-hub of chart speaks to the specific year shown in figure 3.

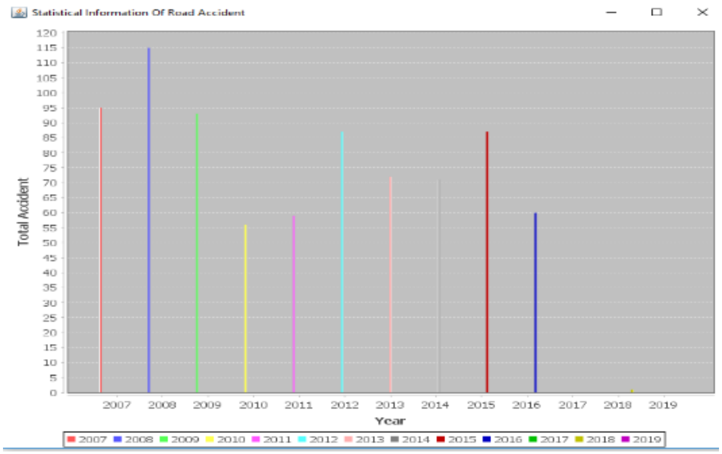


Fig.3. Statistical Information

Accident Prone Vehicle Graph

Accident-prone zone vehicle graph analyses the data and shows the highest vehicle got an accident in a location; it represents the data in the pie chart figure 4.

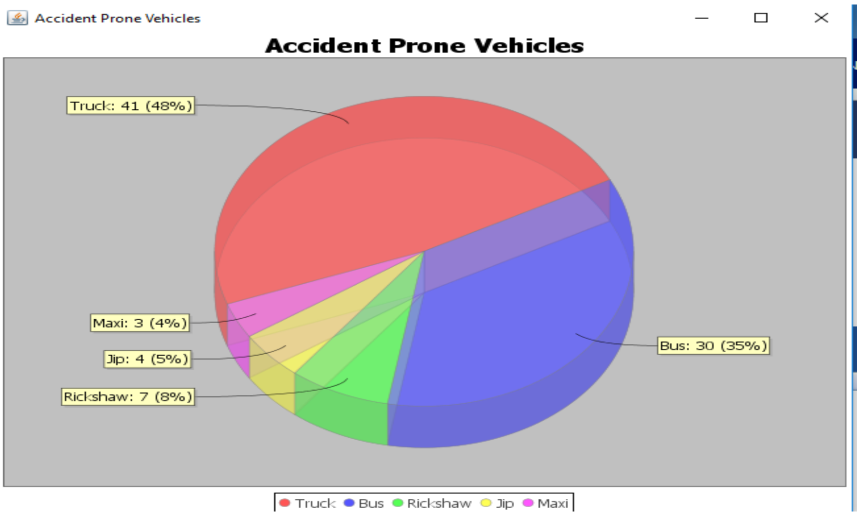
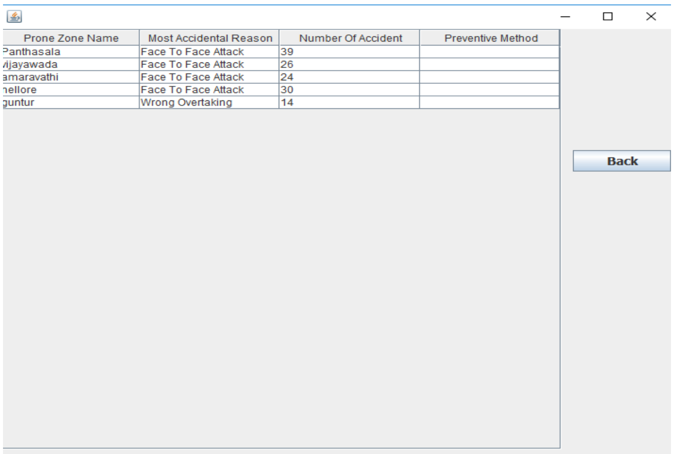


Fig. 5. Accident Prone Vehicle

Prevention Methods

The prevention method shows the safety measures to avoid the accident. It shows the safety measures for two-wheelers separately and four-wheelers separately. Finally, it analyses the vehicle type and shows the prevention methods in table format shown in figure 6.



Prone Zone Name	Most Accidental Reason	Number Of Accident	Preventive Method
Panthasala	Face To Face Attack	39	
vijayawada	Face To Face Attack	26	
amaravathi	Face To Face Attack	24	
nellore	Face To Face Attack	30	
guntur	Wrong Overtaking	14	

Fig.6. Prevention Methods

5 Conclusion

As the measurements appear, the extraction of affiliation and characterization rules, natural segments, for example, street surface, timing, light and climate condition status don't affect large death rates, though human factors, for example, movement infection or movement affliction and kind of impact have more impact on fatal rates. From the consequences of the gathering, we can see some nation's areas have the higher death rates while the others have lower death rates. We can drive cautiously on that specific district or state when we realize that the zone is threat inclined region or clumsy zone. One can think about that locale by observing the deadly mishap information provided from the datasets given. It is easy to use but difficult to know the information from datasets when somebody like police, NGO's erase or delete the information. In theories, the information like climate information, area information, time zone information and vehicle information and so on.

References

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.
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 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. Sheela, AC Santha, and C. Jayakumar. "Comparative Study of Syntactic Search Engine and Semantic Search Engine: A Survey." In 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), vol. 1, pp. 1-4. IEEE, 2019.
10. S. Prince Mary, Dr. Baburaj (2016), “A novel framework for an efficient online recommendation system using constraint-based web usage mining techniques”, Biomedical Research, Special Issue: S12345-S6789, Indexed in Web of Science.
11. Lakshmi. M, Refonaa .J, Vivek, J(2015) ,“Tracking of bio medical waste using global positioning system”, IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2015, IEEE, PP. 1-6.
12. S.L.Jany Shabu, C.Jayakumar (2018), “Multimodal Image Fusion and Bee Colony Optimization for Brain Tumor Detection”, Journal of Engineering and Applied Sciences, Asian Research Publishing Network (ARPN), Vol.13, Issue.8, pp.no.2899-2904.
13. M Selvi, P M Joe Prathap, “Analysis & Classification of Secure Data Aggregation in Wireless Sensor Networks”, International Journal of Engineering and Advanced Technology, Volume-8 Issue-4, April 2019, pp.1404-1407.