

# Research on Road Traffic Fatal Accidents using Data Mining Techniques

S.P. Godlin Jasil, K. Sai Preeti, Shaik Arifa Banu

*Abstract— Road safety plays a major role in our day-to-day life and also transportation system, due to its priority, it has become the major concern for everyone. In order to increase the road safety, traffic rules are included in education, clear and careful predictive analysis and study is done on factors effecting fatal accidents. We apply predictive analysis, statistical analysis and some algorithms related to data mining which includes FARS such as Apriori algorithm, associative rule techniques are used. These methods help in encountering the road fatal accidents that cause due to mentioned factors. These factors may include climatic and surface conditions and also drunken drivers or may be condition of vehicles also. Clusters are formed using simple k-means clustering algorithms. Finally road safety driving rules are made based on the factors effecting, clusters formed and predictive analysis and prior information.*

## I. INTRODUCTION

Due to increase in the population rate, there has been a major growth in number of vehicles used. Though public is requested to use the common public ransport, this has been a major increase in the vehicles being used. In proportion to this, road accidents have also being increased due to lack of road safety precaution and other related factors like climatic conditions and surface conditions, drunken drivers and also condition of vehicles [8]. This leads to increase in the fatality rate and need to corrected.

Since there is a increase in vehicules every day ,accidents can happen at any time and in any place. Some accident involves fatality, which means that people die in accidents. To stay safely we should reduce accidents. To Know about the safety , data mining technique can be applied to accident dataset to know about the information and it will be useful to the drivers.

Data mining generally utilizes many distinct kinds of methods and algorithms to find the connection in big amounts of information linked to big information sets [7]. It is regarded one of the most important tools in information technology in the last centuries, the Association rule mining algorithm is a common methodology for identifying the important relationships between data stored in big databases and also plays a very important part in frequently set item mining. A classic association rule mining technique in the Apriori algorithm where the primary job is to discover frequent item sets from big information sets, which is the technique we normally use to evaluate the information sets for road traffic. The data mining methodology classification method primarily seeks at building a model (classifier) from

a training data set that can be used to classify documents of unknown class labels. The Naïve Bayes method is one of the very fundamental probability-based classification techniques based on the hypothesis of the Bayes with the presumption that each pair of variables is independent. The algorithmic chase program is used to determine the moving cluster position.[6,9 ].

In our day-to-day life multiple number of accidents occurs due to various reasons and the accident count increases and the transportation system and the road safety is the important factor to reduce the accident count. In accidents, many people get injured and also some may die. Accidents can occur in different ways depending upon different situations. There are many factors and reasons for road accidents which may include road conditions, lighting conditions, whether conditions, weight of the vehicle, no. of people travelling in the vehicle, speed of the vehicle at the time of accident, etc.,.

## II. RELATED WORK

[1]. Prajakta S.Kasbe, Apeksha V sakhare,” Review on road accidents data analysis using data minig techniques”, 2017 IEEE International conference on innovations in information embedded and communication systems (ICIIECS).

In this paper ,Deharadun distinct of Uttarakhand state road accident data set to analyze accident patterns, the accident data has been collected from ambulance services which contain 17 features .From the dataset ,13 features has been taken for analysis. Based on accident type it forms 6 clusters by using k-mean clustering algorithm. Cluster 1 gives the data of two wheeler accident happend in congested area, cluster 2 gives the data of two wheeler accident happend in less famous area, cluster 3 gives the data of vehicle drop down from long height, cluster 4 gives the data t of accident during night, clusters 5 gives the data accident take place in market place, cluster 6 gives the data about accident on slope area. The relationship between different clusters are findout by association rule .

[2]. E.Suganya, S. Vijaya Rani “ANALYSIS OF ROAD ACCIDENTS IN INDIA USING DATA MINING CLASSIFICATION ALGORITHMS” 2017 IEEE International Conference on inventive computing and informatics (ICICI).

The classification of the Naive Bayesian is based on the theorem of Bayes with assumptions of autonomy between the predictors. It is easy to build a Naïve Bayesian

Revised Manuscript Received on July 18, 2019.

S.P. Godlin Jasil, Assistant professor, Dept. of CSE, Sathyabama University, Chennai, Tamilnadu, India.(E-mail: spgodlinjasil@gmail.com)

K. Sai Preeti, UG Student, Dept. of CSE, Sathyabama University, Chennai, Tamilnadu, India

Shaik Arifa Banu, UG Student, Dept. of CSE, Sathyabama University, Chennai, Tamilnadu, India.

model and can be used for very big information sets. It can also use supervised discretization to handle numerical attributes. The algorithm of Naive Bayes is based on probabilities that are conditional and dependent. Bayes' Theorem determines the probability of an case taking place considering the probability of another case already taking place. P(c) is calculated from P(c), P(x), and P(x) to describe the later probability. The impact on a specified class (c) of the value of a predictor(x) is unconventional of the values of other predictors present. This hypothesis is called independence conditional class. One way of classification is to determine the posterior probability for each class and assign c to the greatest probability class.

[3]. Liling Li, Sharad Shrestha, Gongzhu Hu, "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques", 2017 IEEE 15th International Conference on Software Engineering Research Management and Applications (SERA).

Eric studied the response on highway velocity to the accident in Washington State. Some researchers claim that those countries that increased their speed limit from 55 mph to 65 mph after 1974 had a 27 percent rise in the fatality rate compared to a 10 percent rise in states that did not raise or boost the speed limit. It is asserted that as the impact of change varies between urban and rural regions in elevated and maximum velocity. In rural regions, after 1987, the accident risen while urban regions remained fairly continuous, but the rate of conflict in urban built-up junction is twice as big as in rural junction. Accident is a sort of road (intersection, road) based on region (urban / rural).

[4]. Alyssa Ditcharoen, Bunna Chhour, Tunyarat Traikunwaranon, "Road Traffic Accidents Severity Factors", 2018 5th International Conference on Bussiness and Industrial Research (ICBIR).

In an earlier study,[2 ] reviewed a number of literature on traffic and road features that influence road accident severity. They found that traffic features such as: traffic velocity ; density ; congestion ; and flow have a important impact on the severity of road accidents. On the other side, it is discovered to be important for road characteristics such as: road facilities; and geometry.

[5]. Abhirup Das, abhisek Ray , Abhishek ghosh, "Vehicle accident prevent cum location monitoring system", 2017 8TH Annual Industrial Automation on Electromechanical Engineering Conference(IEMECON)..

We need to evaluate the reasons behind the accidents in order to decrease road accidents. If we see the documents, it is discovered that there are many accidents due to rash driving triggered by the drunken driver's alcoholic state. Once drunk, the driver loses their driving control. Second sort of accident happens as a result of driver's tiredness whiler iding a lengthy distance on a stretch or driving at night without adequate sleep. According to the study submitted by the Government of India's Ministry of Road Transport and Highways in 2011 4.97 road accidents were occurred, which is 1 accident per minute

III. METHODOLOGY & RESULTS

Prior to each model building, data preparation was conducted and managed. All records with the missing value in the selected attributes will be deleted or removed (usually

represented by 99 in the dataset). All numerical values have been transformed into nominal values in the attached user guide according to the data dictionary. Calculated and split fatal rates into two classifications: elevated and low. Several variables from other independent variables are calculated. Two instances are as follows:

Fatal Rate

This variable shows the proportion of death in a fatal incident calculated as  $FATAL\ RATE = \frac{FATALS}{PERSONS}$ , where FATALS is the amount of deaths and PERSONS is the amount of people engaged in the incident. It is also indicated in the assessment as a "rate."

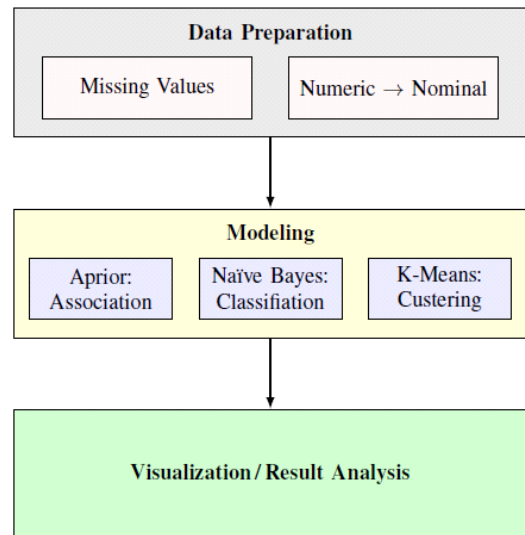


Fig 1. Architecture

Arrival time:

This variable is the time of arrival of emergency staff in minutes calculated as  $ARRIVAL\ TIME = 60 [ARR\ HOUR - HOUR] + ARR\ MIN - MINUTE$ .

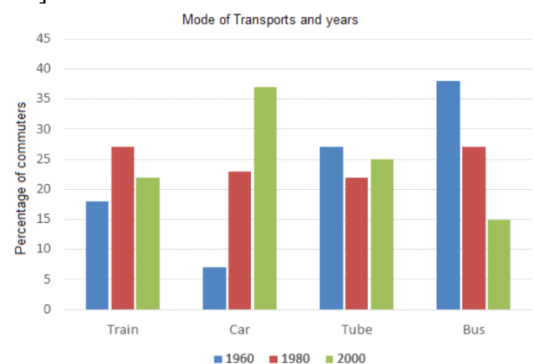


Fig 2. Bar Graph

Algorithms Used:

- Apriori Algorithm
- Simple k-means Clustering Algorithm
- Naïve Bayes.

### 1. Apriori Algorithm:

Apriori is a repetitive item set mining algorithm and association rule for studying transactional databases. It starts by acknowledging the frequent individual objects in the database and expanding them to many bigger item sets as long as these item sets appear in the database system frequently enough. Apriori's frequent item sets can be used to determine and determine rules of association that highlight overall trends. Apriori algorithm utilizes the "bottom-up" strategy, where regular frequent subsets are extended to one element at a moment (a step called candidate generation) and candidate groups are tested against the information. When no additional effective additions are discovered, the algorithm ends.

### 2. Simple k-means Clustering Algorithm:

K-means is one of the easiest and most unsupervised learning algorithms for well-known common problem clustering. The method follows a straightforward manner of classifying a specified information set into the fixed Apriori by a minimum amount of clusters (assuming it to be k clusters). Defining k centers is the primary concept, one for each cluster. These centers should be put in the cunning manner because they cause distinct outcomes due to distinct locations. So, the best way is to keep them as far away from each other as possible. The next stage is to take each point that belongs to a collection of information and connect it to the closest center. The first pace is finished when no point is pending and an early group era is prepared. We need to recalculate k fresh centroids at this stage as the barycenter of the clusters arising from the pre-step. A fresh binding has to be made between the same information set points and the closest new center after we have these k fresh centroids. A loop must be created. As a consequence of this loop, we may notice that the k centers change their position step by step until no other modifications are made or segments do not move in alternative words.

### 3. Naïve Bayes Classification:

The Naive Bayes classification algorithm is a probabilistic classifier that involves variety of opportunity. It is based on anticipated models incorporating very powerful assumptions of independence. Often, the assumptions of independence may not affect real-time circumstances. They are therefore regarded as naive. Using the theorem of Bayes (credited to Thomas Bayes) you can obtain these probability models. You can be created to train the Naive Bayes algorithm in a supervised learning scenario depending on the events of the probability model. For Naive Bayes models, data mining is mostly dependent on the much higher probability of parameter estimation. The Naive Bayes model produced ensures the standard for Predictive Model Markup Language (PMML). A model from Naive

## IV. EXPERIMENTAL RESULTS:

### STATISTICAL RESULTS:

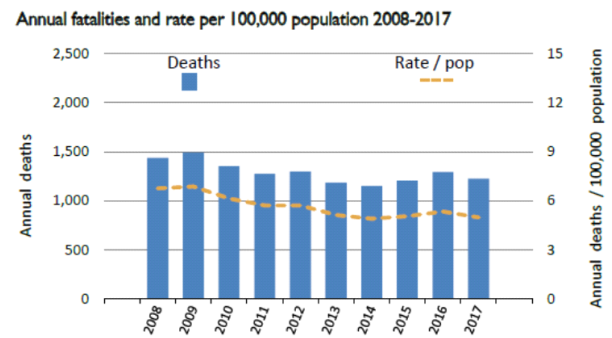


Fig 3. Bar Graph

Annual fatality rate variation from 2008-2017 are shown in Fig 3. The most of the fatal accidents have happened in the year 2009.

## V. CONCLUSION

As shown in statistics, association rule mining, and classification, climate and light conditions do not have a strong impact on the fatal rate, whereas human factors like being drunk or not, and the type of crash has a greater effect on the fatal rate. The primary objective of this research is to assess and investigate algorithms of classification. The information set for road accidents is used to evaluate the efficiency of the selected classifiers on the bigger information set. The algorithm with the smallest absolute error mean and greater precision value is selected as the best algorithm. By considering distinct precision parameters and error rate, the KNN classification algorithm is discovered to be the best algorithm with a maximum precision of 93.7 compared to other classification algorithms.

## VI. ACKNOWLEDGEMENT:

We wish to acknowledge the Department of Science and Technology, India and School of Computing, Sathyabama Institute of Science and Technology, Chennai for providing the facilities to do the research under the DST-FIST Grant Project No.SR/FST/ETI-364/2014.

## VII. REFERENCES

1. Prajakta S.Kasbe, Apeksha V sakhare, "Review on road accidents data analysis using data mining techniques", 2017 IEEE International conference on innovations in information embedded and communication systems (ICIIECS).
2. E.Suganya, S. Vijaya Rani "ANALYSIS OF ROAD ACCIDENTS IN INDIA USING DATA MINING CLASSIFICATION ALGORITHMS" 2017 IEEE International Conference on inventive computing and informatics (ICICI).
3. Liling Li, Sharad Shrestha, Gongzhu Hu, "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques", 2017 IEEE 15<sup>th</sup> International Conference on Software Engineering Research Management and Applications (SERA).

4. Alyssa Ditcharoen, Bunna Chhour, Tunyarat Traikunwaranon, "Road Traffic Accidents Severity Factors", 2018 5<sup>th</sup> International Conference on Business and Industrial Research (ICBIR).
5. Abhirup Das, abhisek Ray , Abhishek ghosh, "Vehicle accident prevent cum location monitoring system", 2017 8<sup>TH</sup> Annual Industrial Automatation on Electromechanical Engineering Conference (IEMECON).
6. Ms. Godlin Jasil S. P, Shaik Asif Moinuddin, Shaik Baba Ibrahim, M. Sakthivel and B. Sakthi Arjun, "Home Security Alert System Using Moving Object Detection In Video Surveillance System", ARPJ Journal of Engineering and Applied Sciences, Vol. 11, No. 15, August 2016
7. Subhashini, R., Jeevitha, J.K., Samhitha, B.K., "Application of data mining techniques to examine quality of water", International Journal of Innovative Technology and Exploring Engineering, 2019.
8. Sethuraman, R., Sathish, E. "Intelligent transport planning system using GIS" International Journal of Applied Engineering Research 01/2015; 10(3):5887-5892. 547.
9. Ashmitha, R., Jeberson Retna Raj, R., Chennai city explorer an android application for locating services for everyday essentials, International Journal of Pharmacy and Technology, 2016