



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 9, Issue 11, November 2020

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.512**

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Analysis of Road Traffic Fatal Accidents using Data Mining Techniques

Kamali R, T Karthikeyan

Department of Computer Science and Engineering, Knowledge Institute of Technology, Salem, India

Assistant Professor, Department of Computer Science and Engineering, Knowledge Institute of Technology,

Salem, India

**ABSTRACT:** Roadway traffic safety is a major involvement for transportation agencies as well as ordinary citizens. Critical to find out the roadway traffic data variables that are closely related to fatal accidents. In this paper apply statistics analysis and data mining algorithms on the FARS(Fatal Accident Rate Study).The relationship between fatal rate and other attributes including collision manner, weather, Surface Condition, Light Condition and drunk driver were investigated. Association rules were discovered by Apriori Algorithm, Classification Model was built by Naïve Bayes classifier and Clusters were forms by simple K-Means clustering Algorithm. The information is preprocessed. After preprocessing Apriori and Predictive Apriori Association rules Algorithm were applied to the dataset to investigate the connection between recorded accidents and factors. First calculated several statistics from the dataset to show the basic characteristics of the fatal accidents then applied association rule mining, clustering and Naïve Bayes classification to find relationships among the attributes and the patterns.

**KEYWORDS:** Data Mining, Association Rules, Classification Model, Clustering.

## I. INTRODUCTION

There are a lot of vehicles driving on the roadway every day, and traffic accidents could happen at any time anywhere. Some accident involves fatality that means huge people die in that accident. Nearly 3,500 people die on the world's roads every day. Tens of millions of people are injured or disabled every year. Children, pedestrians, cyclists and the elderly are among the most vulnerable of road users.To find out how to drive safer, data mining technique could be applied on the traffic accident dataset to find out some valuable information, thus give driving suggestion. Data mining uses many different techniques and algorithms to discover the relationship in large amount of data.

Association rule mining algorithm is a popular methodology to identify the significant relations between the data stored in large database and also plays a very important role in frequent item set mining. A classical association rule mining method is the Apriori algorithm whose main task is to find frequent itemsets, which is the method we use to analyze the roadway traffic data. Classification in data mining methodology aims at constructing a model (classifier) from a training data set that can be used to classify records of unknown class labels.

Partition based clustering and density based clustering were performed to group similar accidents together. Most of the data K-modes algorithm was used. To find the correlation among various sets of attributes association rule mining was performed. First the data set is classified into clusters and each of them are studied to predict some patterns. Among the various rules that are generated those which seemed interesting were considered based on support count and confidence. Association rules mining algorithm apply to dataset about traffic accidents. The information is preprocessed. After preprocessing Apriori and Predictive Apriori Association rules Algorithm were applied to the dataset to investigate the connection between recorded accidents and factors. The relationship between fatal rate and other attributes including collision manner, weather, Surface Condition, Light Condition and drunk driver were investigated. Association rules were discovered by Apriori Algorithm, Classification Model was built by Naïve Bayes classifier and Clusters were forms by simple K-Means clustering Algorithm.

## II. RELATED WORKS

One of the useful and efficient algorithms of Association mining named as APRIORI algorithm. Association rule of data mining is used in all real life applications of business and industry. Using this we gets an effective results rather than traditional results. Association rules are the main technique for data mining and APRIORI algorithm is a classical

algorithm. Lots of algorithms for mining association rules and their mutation (change/transformation) are proposed on basis of APRIORI algorithm, but traditional algorithms are not efficient. The main intension of this paper is to understand the concept of association rule and how to implement the APRIORI algorithm and improved APRIORI algorithm.

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

Rural mayday systems can reduce the time between the occurrence of an accident and the notification of emergency medical services—called the accident notification time. Reductions in this time, in turn, may affect the numbers of fatalities. A statistical analysis is used to estimate the quantitative relationship between fatalities and the accident notification time. The elasticity of rural fatalities with respect to the accident notification time was found to be 0.14. If a rural mayday system were fully implemented (i.e. a 100% market penetration) and the service availability were 100%, then we would expect monetary benefits of about \$1.83 billion per year and comprehensive benefits (which includes the monetary value attached to the lost quality of life) of \$6.37 billion per year.

Data mining is a promising area for dealing with the increased, stored data that has been generated in our times .It is the extraction of implicit, previously unknown and useful data. In this paper we have analyzed some of the data mining techniques, tools, applications and search engines for accident investigation and traffic analysis. Most of the accident investigation methodologies are based on scenarios of the accident occurrence and simulation of accident situation. The costs of fatalities and injuries due to traffic accident have a great impact on society. Engineers and researchers in the automobile industry have tried to design and build safer automobiles, but traffic accidents are unavoidable. In recent years, researchers have been utilizing real-life data in studying various aspects of traffic accidents. So measures have to be taken to reduce accidents. It is important that the measures should be based on scientific and objective surveys of the causes of accidents and severity of injuries.

Data Mining is taking out of hidden patterns from huge database. It is commonly used in a marketing, surveillance, fraud detection and scientific discovery. In data mining, machine learning is mainly focused as research which is automatically learnt to recognize complex patterns and make intelligent decisions based on data. Nowadays traffic accidents are the major causes of death and injuries in this world. Roadway patterns are useful in the development of traffic safety control policy. This paper deals with the some of classification models to predict the severity of injury that occurred during traffic accidents. I have compared Naive Bayes Bayesian classifier, AdaBoostM1 Meta classifier, PART Rule classifier, J48 Decision Tree classifier and Random Forest Tree classifier for classifying the type of injury severity of various traffic accidents. The final result shows that the Random Forest outperforms than other four algorithms.

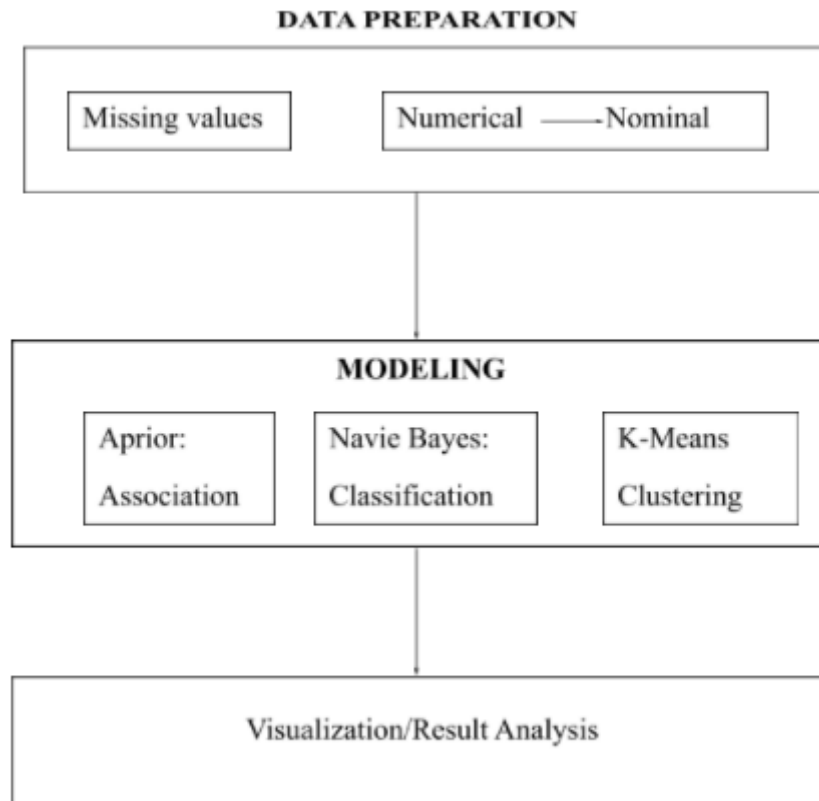
### III. EXISTING SYSTEM

The analysis of roadway traffic data is critical to find out variables that are closely related to fatal accidents. Partition based clustering and density based clustering were performed to group similar accidents together. Based on the categorical nature of most of the data K-modes algorithm was used. To find the correlation among various sets of attributes association rule mining was performed. First the data set is classified into clusters and each of them are studied to predict some patterns. Among the various rules that are generated those which seemed interesting were considered based on support count and confidence.

**Problem identification**

- In this existing system the large data value is calculated. sometime the calculated data is missing so Fatal Rate value is not accurate
- First data set is classified into clusters and the clusters are predict some patterns.
- This patterns generated the rules based on support count and confidence.

**IV. ARCHITECTURE DIAGRAM**



The proposed work is analysis of real time data for road traffic accident. To overcome the existing problem of data missing and produce inefficient results also. To find further classification of data are categorized for different factors. The statistical analysis shows that the number accident in a selected city during the year and the causes of road traffic accident.

Thus we have causes of accident records with age and gender information, the type of vehicle and the accident severity moreover in some city. This analysis to predict the accuracy of Road Traffic Accident given in the graphical representation. We apply association rules mining algorithm on a dataset about traffic accidents. After information preprocessing, Apriori and Predictive Apriori association rules algorithms were applied to the dataset to investigate the connection between recorded accidents and factors.

Two sets of class association rules were generated using the Apriori and Predictive Apriori algorithms and summarized to get the most interesting rules using technical measures. Exact results demonstrated that the class association rules created by Apriori algorithm were more viable than those created by Predictive Apriori algorithm. More relationship between accident factors and accident severity level were investigated while applying Apriori algorithm

**ADVANTAGES**

- We first calculated several statistics from the dataset to show the basic characteristics data of the fatal accidents.
- In this characteristics data the Fatal Rate value is accurate.

## V. MODULES

- Emergency medical services
- National highway transportation safety administration
- Clustering of States
- Association Rule Mining

### MODULE DESCRIPTION:

#### Emergency medical services:

It is assumed that the fatality rate of an accident might be reduced with the introduction of an express emergency system. Reducing the time of delivery for emergency medical services (EMS) accident victims could be treated in time saving their lives. Accident notification time, the difference between crash and EMS notification time is the most crucial. Trauma is time dependent disease and trauma victims could be saved if treated on time. Trauma victims could be stabilized if treated soon. The first 10 minutes is called golden hour. The time was 12.3 minutes for the one who died and 8.4 minutes for the one who survived. The fatality rate was also much higher in rural areas because of unavailability of rapid EMS response in those areas. There are also several other factors affecting the fatality such as vehicle kilometers traveled, alcohol consumption, driver age distribution, accident notification time, personal income per capita and so on .

#### National highway transportation safety administration

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 Naïve Bayes technique is one of the very basic probability-based methods for classification that is based on the Bayes' hypothesis with the presumption of independence between each pair of variables. We used the FARS dataset for our study. The Fatal Accidents Dataset contains all fatal accidents on public roads in 2007 reported to the National Highway Transportation Safety Administration (NHTSA).The dataset is downloaded from California Polytechnic State University and all data originally came from FARS (Fatal Accident Reporting System).The dataset contains 37,248 records and 55 attributes.

#### Clustering of States

To find out which states are similar to each other considering fatal rate, and which states are safer or more risky to drive, clustering algorithm was performed on the fatal accident dataset. To perform the clustering, total number of fatality per state was calculated. Also the population data for each states in 2007 was obtained from U.S. Census Bureau. With the fatal accident and the population dataset, fatalities per million people in the state was calculated. This allowed us to compare relative fatal rate in a state regardless of population of state. The simple K-means algorithm with Euclidean distance as the dissimilarity measure was applied to the data of 48 states (without Wisconsin, Wyoming, and District of Columbia, for some reason) with two variables: population (in 100,000) and number of fatal accidents.

#### Association Rule Mining:

Before applying the algorithms, the tuples with missing value in chosen attributes were removed, the numerical values were converted to nominal values according to data dictionary in the user guide. The clean data was stored in CSV format and ready to be analyzed by the data analyzing tool Weka. The clean data for association rule mining and classification contains 36,789 tuples, 5 condition attributes, and 1 decision attribute. A small partial sample of the dataset is shown in Table I. All values were converted to nominal values.

## VI. CONCLUSION

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



#### REFERENCES

- [1] D. Fisher, “‘venom’ flaw in virtualization software could lead to vm escapes, data theft,” <https://threatpost.com/venomflaw-in-virtualization-software-could-lead-to-vm-escapes-datatheft/112772/>, 2015, accessed: 2015-05-20.
- [2] Z. Durumeric, J. Kasten, D. Adrian, J. A. Halderman, M. Bailey, F. Li, N. Weaver, J. Amann, J. Beekman, M. Payer et al., “The matter of heart bleed,” in Proceedings of the 2014 Conference on Internet Measurement Conference. Vancouver, BC, Canada: ACM, 2014, pp. 475–488.
- [3] K. Cabaj, K. Grochowski, and P. Gawkowski, “Practical problems of internet threats analyses,” in Theory and Engineering of Complex Systems and Dependability. Springer, 2015, pp. 87–96.
- [4] J. Oberheide, E. Cooke, and F. Jahanian, “Clouday: N-version antivirus in the network cloud.” in USENIX Security Symposium, San Jose, California, USA, 2008, pp. 91–106.
- [5] X. Wang, Y. Yang, and Y. Zeng, “Accurate mobile malware detection and classification in the cloud,” Springer Plus, vol. 4, no. 1, pp. 1–23, 2015.



**INNO SPACE**  
SJIF Scientific Journal Impact Factor

**Impact Factor:  
7.512**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details