

# Roadway Traffic Analysis using Data Mining

Vaishnavi Dhole, Nihal Singh, SrushtiSontakke, Prof. PreetiBhagat

*Department of Computer Science and Engineering, DMIETR Wardha MH India*

**Abstract:** -Roadway traffic safety is a major concern for transportation governing agencies as well as ordinary citizens. 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. Globalization has affected many countries. There has been a drastic increase in the economic activities and consumption level, leading to expansion of travel and transportation. The increase in the vehicles, traffic lead to road accidents. Considering the importance of the road safety, government is trying to identify the causes of road accidents to reduce the accidents level. The exponential increase in the accidents data is making it difficult to analyse the constraints causing the road accidents. The paper describes how to mine frequent patterns causing road accidents from collected data set. We find associations among road accidents and predict the type of accidents for existing as well as for new roads. We make use of association and classification rules to discover the patterns between road accidents and as well as predict road accidents for new roads.

**Keywords:** Data mining, Association rule, Classification rule, Apriori algorithm, Naïve Bayes algorithm

## 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, 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. 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. [2,5,7]

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 algorithm 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.

power based multi mobile charger system is implemented in the proposed system which can be used in public places like railway stations, bus stands, hospitals and parks etc.

## II. OBJECTIVES

The primary objectives of this study can be summarized as follows:

1. To process the dataset.
2. To cluster the dataset based on Parameters and apply k-means for clustering and Bayesian algorithm for classification.
3. To analyzed the graph and Predict the Road Accidents based on various parameters.
4. To develop automatic analytical tool which mines road wise accident patterns
5. To develop a Decision support system for general public and for government which will predict the possibilities of accidents and spread awareness about the roads and take decision about redevelopment of roads

### III. PROPOSED WORK

Roadway traffic safety is a major concern for transportation governing agencies as well as ordinary citizens so, for that purpose we are introducing an analytic tool in which Genetic algorithm will be used for classification. We are taking dataset of a country and analyzing that database month wise for one year, 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 accidents dataset. Before applying the algorithms, the tuples with missing value in chosen attributes were removed. The proposed work is planned to be carried out in the following manner.

Data preparation was performed before each model construction. All records with missing value (usually represented by 99 in the dataset) in the chosen attributes were removed. All numerical values were converted to nominal value according to the data dictionary in attached user guide.

#### Algorithms:

##### 1. Association Rule

To find out how to drive safer, data mining technique could be applied on the traffic accident dataset to find out some valuable information, thus give driving suggestion.

Data mining uses many different techniques and algorithms to discover the relationship in large amount of data. It is considered one of the most important tool in information technology in the previous decades.

Association rule mining algorithm is a popular methodology to identify the significant relations between the data stored in large database and also plays a very important role in frequent itemset mining. A classical association rule mining method is the Apriori algorithm who main task is to find frequent itemsetss, which is the method we use to analyze the roadway traffic data.

Before applying the algorithms, the tuples with missing value in chosen attributes were removed, the numerical values

were converted to nominal values. The clean data was stored in CSV format and ready to be analyzed by the data analyzing tool. The clean data for association rule mining and classification contains number of tuples, 5 condition attributes, and 1 decision attribute After applying Apriori algorithm with minimum support and minimum confidence in analyzing tool, association rules with fatal rate at the right side as decision were generated.

Rule Measures: Support and Confidence

- Confidence( $X \rightarrow Y$ ) = #tuples containing both X & Y / #tuples containing X =  $\Pr(Y/X) = \Pr(X \cup Y) / \Pr(X)$
- Support( $X \rightarrow Y$ ) = #tuples containing both X & Y / total number of tuples =  $\Pr(X \cup Y)$

##### 2. Naïve Bayes Classification

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.

Naive Bayes classifier was built on the cleaned data. The Naive Bayes Classifier shows that the fatal rate does not strongly depend on the given attributes, although they are considered feature in comparison to other attributes in the dataset.

On the basis of these formulas the naïve bayes classification table will be created.

$$T_p = \frac{\text{Number of accidents per event}}{\text{Total number of accidents}}$$

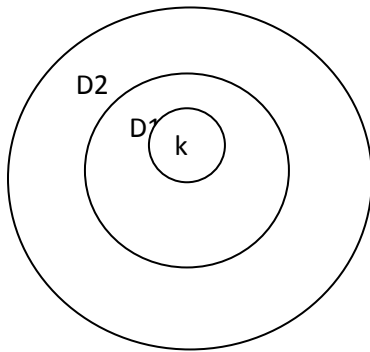
$$F_p = \frac{\text{Number of false accidents per event}}{\text{Total number of accidents}}$$

$$\text{Precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

$$\text{Recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

##### 3. K-means clustering

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. With the fatal accident and the population dataset, fatalities per million people in the state was calculated.



K= Parameter (i.e. weather condition)

For given k cluster evaluation will be done as (M)

$$\text{Mean cal} = \frac{\sum \text{all entities}}{\text{Cluster mean}(k)}$$

Recluster till two consecutive (M) has some values.

### Work Flow

In the modeling We first calculated several statistics from the dataset to show the basic characteristics of the fatal accidents. We then applied association rule mining, clustering, and Naive Bayse classification to find relationships among the attributes and the patterns.

The results of our analysis include association rules among the variables, clustering of states in the USA on their populations and number of fatal accidents, and classification of the regions as being high or low risk of fatal accident.

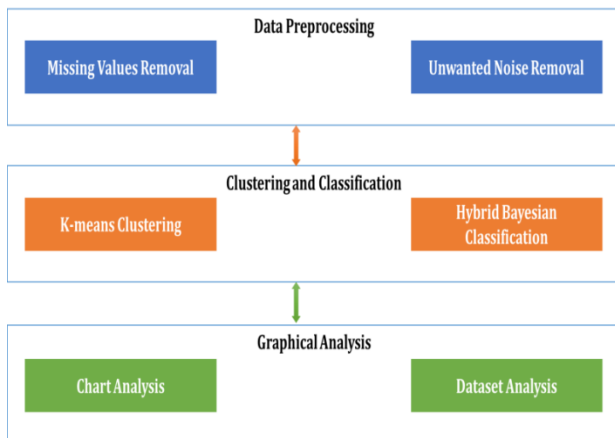


Fig 2. Work Flow

Clustering is a process of collection of objects which are similar between them while dissimilar objects belong to other clusters. A clustering technique is used to obtain a partition of N objects using a suitable measure such as resemblance function as a distance measure 'd'.

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 algorithm 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.

### IV. CONCLUSION

In this paper, we have collected multiple researchers' works together in single document and discussed about the contribution towards impact of road and traffic accident on human life and society. This highlights the number of approaches used to avoid the accident happened in various countries and cities. The paper also discussing about various data mining techniques which is proved supporting to resolve traffic accident severity problem and conclude which one could be optimal technique in road traffic accident scenario. The brief discussion will also help us to find better mining technique in this kind of problem.

### REFERENCES

- [1] Zhexue Huang, "Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values", *Data Mining and Knowledge Discovery* 2, 283–304 (1998).
- [2] Sachin Kumar and Durga Toshniwal, "A data mining framework to analyse road accident data", *Journal of Big Data* (2015) 2:26 DOI 10.1186/s40537-015-0035-y.
- [3] S. Krishnaveni and Dr. M. Hemalatha, "A perspective analysis of Traffic Accident Using Data Mining Techniques", *International Journal of Computer Application*.
- [4] Olutayo V.A and Eludire A.A, "Traffic Accident Analysis Using Decision Trees and Neural Networks", *I.J. Information Technology and Computer Science*, 2014, 02, 22-28 Published Online January 2014 in MECS (<http://www.mecspress.org/>) DOI: 10.5815/ijitcs. 2014.02.03.
- [5] K. Geetha and C. Vaishnavi, "Analysis on Traffic Accident Injury Level Using Classification", *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 5, Issue 2, February 2015, ISSN: 2277 128X.
- [6] Sachin Kumar and Durga Toshniwal, "A data mining approach to characterize road accident locations", *J. Mod.*

Transport. (2016) 24(1):62–72 DOI 10.1007/s40534-016-0095-5.

[7] Tibebe Beshah, Shawndra Hill, “Mining Road Traffic Accident Data to Improve Safety: Role of Road- elated Factors on Accident Severity in Ethiopia”

[8] Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993. [9] K. Jayasudha and Dr. C. Chandrasekar, “An overview of Data Mining in Road Traffic and Accident Analysis”, Journal of Computer Applications, Vol – II, No.4, Oct – Dec 2009.

[10] Miao Chong, Ajith Abraham and Marcin Paprzycki, “Traffic Accident Analysis Using Machine Learning Paradigms”, Informatica 29 (2005) 89–98.

[11] M. Sowmya and Dr.P. Ponmuthuramalingam, “Analyzing the Road Traffic and Accidents with Classification Techniques”, International Journal of Computer Trends and Technology (IJCTT) – volume 5 number 4 –Nov 2013.