

A REVIEW ON DESIGN AND IMPLEMENT OF ROAD ACCIDENT ANALYSIS FOR INDIAN STATES USING ASSOCIATION MINING.

Bushra D.Qureshi¹, Dr. S. S. Dhande²

Master of Engineering (M.E.) Scholar ¹

Pph

Department of Computer Science & Engineering,
Sant Gadge Baba Amravati University, Amravati, India

Abstract : Data Mining is a technique which deals with huge and complex data sets which has led to globalization that has affected many countries. There has been an excessive increase in the economic activities and consumption level, leading to increase of travel and transportation. Road traffic accidents are among the foremost cause of death and injury worldwide. To give safe driving suggestions, careful analysis of roadway traffic data is needed. So, there is need of system which provide better solution for such problems. Proper data mining approach has to be applied on collected data sets indicating occurred road accidents to identify possible unknown relationships and connections between various factors affecting road accidents with serious consequences. The paper gives brief introduction of accidents and proposed work.

Index Terms- Data mining, Association rule, Clustering, K-means algorithm, Classification rule, Naïve Bayes algorithm.

I. INTRODUCTION

Accidents happened due to the inattentiveness of driving vehicle on the roads. There are various reasons liable for the accident like abandon of traffic rules but road conditions and the traffic are considered the one of prime cause of fatality across the globe. Dynamic design and development of automobile industries are the reasons of occurrence of these accidents. It results in injury, property damage, and death. In survey it's seen that likely 1.2 million death and 50 million injuries estimated worldwide every year. In any accident, it studies about the driver's behaviour, road substructure and possibilities of weather forecast that could be somewhere linked with different accident incidents. In order to compact with the problem we can adopt data mining model for different scenario. Clustering is a method to divide objects in a similar group. The k-means algorithm having a good effectiveness 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. Data mining technique is known for analysis of accidents harshness problem and finding factors behind them. Damage like property, people due to road accident are adverse. It is fairly easy to use in variety of domain belong to science and management and it also 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. Association rule mining is a method that identifies the connection in different parameter of road accident.

II. LITERATURE REVIEW

In the growing countries in the globe, the motorist, are facing road accidents due to poor management in traffic seeing the common leading cause of injury in body and mortality. Data mining techniques could be used to resolve these issues. In survey, numerous researchers contributed and discussed about various techniques of data mining, few important in the context of our problem are shared in this review paper.

Gower et, al., (1971) showed the importance of similarity coefficient and Gowda et, al., and Anderberg et, al., share dissimilarity measures that specify the standard mechanism of hierarchical clustering methods work with numeric and categorical values. But conversion of categorical data with the numeric dataset which will not produce meaningful result when categorical domains are not in order.

Ralambondrainy (1995) introduced k-means algorithm approach using data mining to cluster categorical data which convert multiple category attributes into binary numeric attributes. But in data mining these attributes are in hundreds and thousands that compulsory make increment in computation as well as in the space costs of the k-means.

Zhexue Huang (1998), proposed two algorithms which is extension of K-means algorithm. This extended k-means based algorithm includes categorical domain with numeric and categorical values. The k-mean algorithm uses a simple matching dissimilarity measure to deal with categorical objects where k-means algorithm extended replaces the means of clusters with modes, and uses a frequency-based method to update modes in the clustering process to minimise the clustering cost function.

Sachin et, al., (2015), proposed a framework for Dehradun, India road accident (11,574) happened during 2009 and 2014 by using K-modes clustering technique and association rule mining. The analysis of result using combination of these technique conclude that the result will be more effective if no segmentation has been performed prior to generate association rules.

In the world health organization [8], India is taking leading edge with 1,05,000 traffic deaths in a year, with comparison to the china with over 96,000 deaths on road. The survey was executed with approximate 178 countries. As per the survey results, it shown that approximate more than 300 Indians causality on roads every day. There are more than two million people have casualty from a traffic accident. The survey is taken from the report of data collection for 2008.

S. Krishnaveni, (2011), work with some of classification models to predict the injuries happened in traffic accident in Nigeria's and compared Naive Bayes Bayesian classifier . This research is employed on the artificial neural networks based approach while the decision trees data analysis can be used to works on reduction of massacre on the highways. The data was classified in continuous and categorical data where continuous data analysed using artificial neural networks technique and the categorical data, using decision trees technique. The results reveal that decision tree approach outperformed the ANN with a lower error rate and higher accuracy rate. This research based on three most important causes of accident due to tyre burst, loss of control and over speeding. This study used traffic accident records from 1995 to 2000, a total number of 417,670 cases. They 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.

This 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.

K. Jayasudha, (2009), shown the effective use of association rule to investigate the accident issue. She also put efforts that systematic deployment of patters and rules shows the positive impact and it helps in understanding the case of fatality in accidents using decision support system.

K. Geetha, (2015), this study works on traffic accident data of Tamilnadu city. The main aim of this study is to reduce the number of road accidents. The traffic accident data is managed in form of text or numerical formats in unsorted manner.

Sachin Kumar et, al., (2016) suggest to apply k-means algorithm and ARM technique to solve traffic accident severity problem. Author divide the different accidental prone location with three different categories which are high, moderate and low frequency to extract the hidden information behind the data set and take some preventive action according to accident location.

Miao Chong. et. al., also proposed the efficient use of ANN and DT prove good result, in support they have used GES automobile accident data from 1995 to 2000, by studying the analysis performance of different data mining technique a significant result visible in support of fatality case study. Direct decision based approach outperforms the direct NN approach in all cases. Author discussed in this theory, if speed limit factor is well known then accident can be controlled.

The purpose of this study is to achieve the fact that there are no fixed or defined set of procedures or evaluation criteria for comparing any programming language which each other. We can define our own criteria with verified results, calculations and observation for analyzing the difference between these languages.

III. PROPOSED WORK

The main objectives of the study are listed below:

1. Dataset Processing
2. Data Clustering based on Parameters
3. Applying k-means for clustering and Bayesian algorithm for classification.
4. Graph Analysis and Prediction of Road Accidents based on various parameters.

The proposed work, shown in below fig, is planned to be carried out in the following manner:

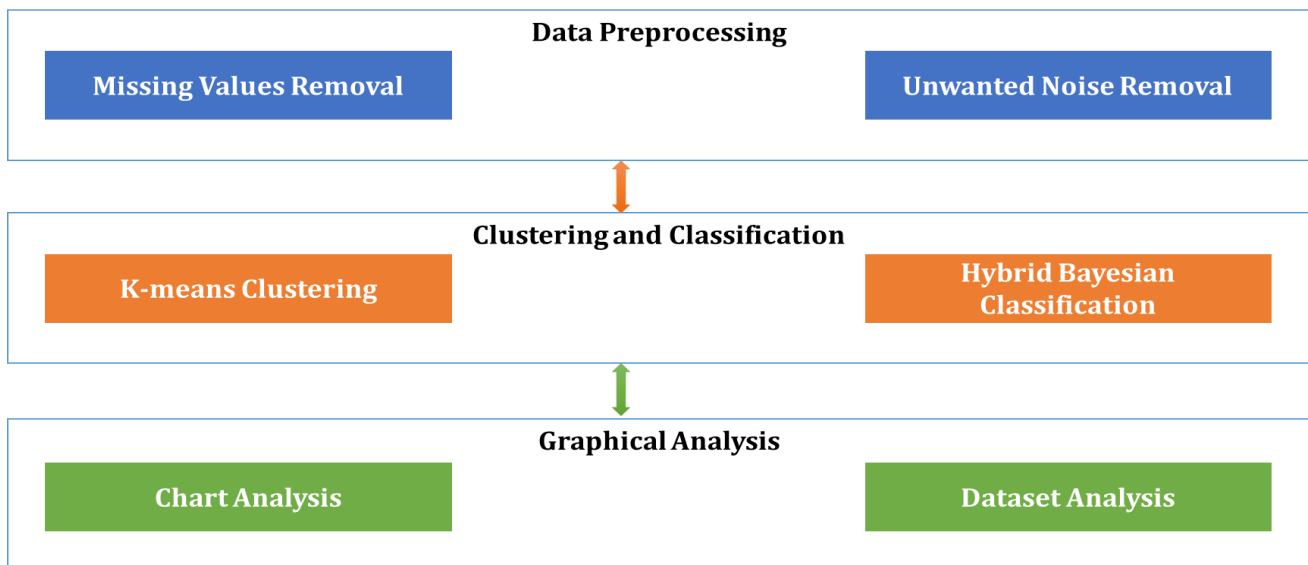


Figure 1: Proposed Work

A. Clustering:

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

B. K-means Algorithm for Clustering:

K-means clustering is a method of vector quantization, initially from signal processing, that is widely held for cluster analysis in data mining. K-means clustering targets to partition n observations into k clusters in which each observation have its place to the cluster with the nearest mean. 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. The 1-nearest neighbor classifier can be applied on the cluster centers obtained by k -means to classify new data into the existing clusters. This is well-known as nearest centroid classifier or Rocchio algorithm.

The approach we took for our study follows the traditional data analysis steps

1) Data Preparation :

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. Mortal rate were calculated and binned to two groups: high and low.

2) Modeling:

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

3) Result Analysis :

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

IV. CONCLUSION

Considering the importance of the road safety, government is trying to identify the causes of road accidents to reduce the accidents level. Roadway traffic safety is a major worry for transportation governing agencies as well as ordinary people. We find associations among road accidents and calculate the type of accidents for existing as well as for new roads. We make use of

association and classification rules to determine the patterns between road accidents and as well as predict road accidents for new roads. Using various parameters we have generated association mining rules and pre-processed dataset accordingly. Using proposed system it is easy to analyse reasons of accidents with its most probable conditions.

V. ACKNOWLEDGMENT

Dedicated our paper work to our esteemed guide, Dr. S. S. Dhande (HOD) whose interest and guidance has helped in the completion of the paper work successfully. Her encouragement, vision and critique made this work possible. I am obliged to the faculty and staff of Sipna College of Engineering and Technology who have helped us to be better acquainted with this work.

VI. 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- related 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.
- [12] Sohn, S. and S. Lee (2002), "Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea", *Safety Science* 41(1): 1-14.
- [13] Depaire B, Wets G and Vanhoof K. Traffic accident segmentation by means of latent class clustering, accident analysis and prevention, vol. 40. Elsevier; 2008.
- [14] Mario De Luca et. al., "Before-After Freeway Accident Analysis using Cluster Algorithms", *Procedia Social and Behavioral Sciences* 20 (2011) 723–731, science direct.