

# Road Accident Analysis System using Data Mining

Chanchal Sharma, Mr. Anil Wadhwa

Department of Computer Science and Engineering  
Rao Pahlad Singh Group of Institutions, Balana, Mohindergarh

**Abstract:-** Information Mining is expelling from hid cases from huge database. It is by and large used as a piece of a displaying, perception, coercion recognizable proof and consistent disclosure. In information mining, machine learning is fundamentally connected as research which is normally learnt to see complex cases and settle on cunning decisions in light of information. Globalization has impacted various countries. There has been an extraordinary addition in the monetary activities and use level, provoking augmentation of development and transportation. The extension in the vehicles, movement provoke street mischances. Contemplating the essentialness of the street prosperity, government is trying to recognize the explanations behind street mischances to diminish the mishaps level.

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

\*\*\*\*\*

## I. Introduction

Mischances happened in view of the inconsiderateness of driving vehicle on the streets. There are distinctive reasons accountable for the incident like surrender of action oversees yet street conditions and the development are seen as the one of prime purpose behind loss and causality over the globe. These mischances happen due to dynamic arrangement and progression of vehicle wanders. A pile up happens due particular reasons like pounds of two vehicles on street, walking individual, animal, or some other trademark impediments. It could realize harm, property mischief, and passing. Auto collision examination required examination of the diverse factor affecting behind them.

In survey its seen that harsh 1.2 million end and 50 million injuries evaluated overall reliably. The gathered estimation of causality and wounds on account of poor street establishment is a noteworthy test before the living animals. The ask for to deal with the issue, in computational science, we can get information digging model for different circumstance. In any vehicle setback, it finds out about the driver's direct, street establishment and potential results of atmosphere assess that could be some place related with different disaster events. The standard issue in the examination and examination of setback information is its mix heterogeneous condition and information division which is used for the most part to vanquish accident issue. [2,5,7]

Information Mining is a computational framework to oversee generous and complex informational collection and these informational collections can be of standard, apparent and mixed. It is extremely easy to use in variety of room have a place with science and organization; moreover, it could be used as a piece of deception recognizing evidence

and various more intelligent cases and furthermore in setback reality issue. Bundle of articles in a social occasion of packs or in a homogeneous set is a vital movement of information mining.

Bundling is a technique to divide in a tantamount social occasion. The k-suggests computation having a respectable viability for batching far reaching informational collections yet bound in molding bunches for honest to goodness word information while working just on numerical information since it helps in diminishing the cost work by altering the centrality of the groups [1,3].

Information mining methodology is seen as strong technique for examination of fender bender earnestness issue and finding factors behind them. Damage like property, people in view of street mishap are annoying. Normally, it happened that street incident scenes are more commonplace at particular places that can help in perceiving factors behind them. Connection run mining is a procedure that perceives the relationship in different parameter of street incident. [6]

## II. Related Work

Sachin et, al., (2015), proposed a system for Dehradun, India road mishap (11,574) occurred amid 2009 and 2014 by utilizing K-modes grouping strategy and affiliation lead mining. The investigation of result utilizing blend of these procedure presume that the outcome will be more powerful if no division has been performed preceding produce affiliation rules [2].

On the planet wellbeing association [8], India is taking driving edge with 1,05,000 movement passings in multi year, with correlation with the china with more than 96,000 passings on road. The review was executed with estimated 178 nations. According to the study comes about, it

demonstrated that rough in excess of 300 Indians causality on roads consistently. There are in excess of two million individuals have loss from a car crash. The overview is taken from the report of data accumulation for 2008.

S. Krishnaveni, (2011), work with some of classification models to anticipate the wounds occurred in auto collision in Nigeria's and looked at Naive Bayes Bayesian classifier [3]. This exploration is utilized on the fake neural systems based approach while the choice trees data investigation can be utilized to takes a shot at diminishment of slaughter on the expressways. The data was arranged in nonstop and absolute data where persistent data examined utilizing fake neural systems procedure and the downright data, utilizing choice trees strategy. The outcomes uncover that choice tree approach outflanked the ANN with a lower mistake rate and higher precision rate. This examination in light of three most critical reasons for mishap because of tire burst, loss of control and over speeding. This investigation utilized car crash records from 1995 to 2000, an aggregate number of 417,670 cases. They connected them to a genuine data set acquired from the National Automotive Sampling System (NASS) General Estimates System (GES). Trial comes about uncover that in every one of the cases the choice tree beats the neural system.

This exploration investigation additionally demonstrates that the three most vital factors in deadly damage are: driver's safety belt utilization, light state of the roadway, and driver's liquor use. [4]

### III. Proposed Methodology

The method include data preprocessing and clustering using k-means algorithm. The clustering and classification techniques will be used for data preprocessing and then for analysis of road accident in the year 2016 for various parameters like road type, gender, monthly analysis, age, type of vehicle etc.

The proposed work is planned to be carried out in the following manner

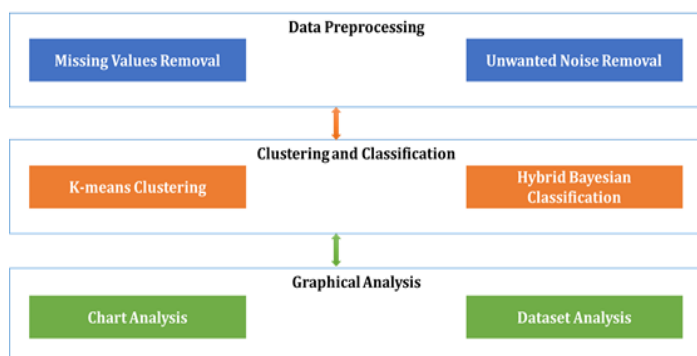


Figure: Proposed Methodology

### Clustering

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 Algorithm for Clustering

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.

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

#### 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. Fatal rate were calculated and binned to two categories: high and low.

#### 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 Bayes classification to find relationships among the attributes and the patterns.

### IV. Conclusion

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. We have provided a brief accidental analysis on UK dataset of year 2016. Using various parameters we have generated association mining rules and preprocessed dataset accordingly. We have also extracted results in form of graphs using Java Freechart API.

Using proposed system it is easy to analyze reasons of accidents with its most probable conditions..

### 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.
- [15] Raffaele Mauro et. al., "Using a K-Means Clustering Algorithm to Examine Patterns of Vehicle Crashes in Before-After Analysis".
- [16] Naina Mahajan and Bikram Pal Kaur, "Analysis of Factors of Road Traffic Accidents using Enhanced Decision Tree Algorithm", *International Journal of Computer Applications* (0975 – 8887) Volume 135 – No.6, February 2016.
- [17] S. Shanthi, "Classification of Vehicle Collision Patterns in Road Accidents using Data Mining Algorithms", *International Journal of Computer Applications* (0975 – 8887) Volume 35– No.12, December 2011.
- [18] Sami Ayramo et. al., "Mining Road Traffic accident", Reports of the Department of Mathematical Information Technology Series C. Software and Computational Engineering No. C. 2/2009, University of Jyväskylä a Department of Mathematical Information Technology P.O. Box 35 (Agora) FI-40014 University of Jyväskylä a FINLAND fax +358 14 260 4980 <http://www.mit.jyu.fi/>.
- [19] Global status report on roadsafety: time for action, WHO, 2009.
- [20] G. Jost, M. Popolizio, R. Allsop, And V. Eksler, 3rd road safety pin report: 2010 on the horizon, report, European Transport and Safety Council (ETSC), 2009.
- [21] Dinesh Mohan et. al., Road Safety in India Status Report, Transpo