

## Analysis of Datamining Technique for Traffic Accident Severity Problem: A Review

Meenu Gupta  
Research Scholar  
Ansal University, Gurgaon  
Email:  
gupta.meenu5@gmail.com

Vijender Kumar Solanki  
Assistant Professor,  
ITS, Ghaziabad  
Email: spesinfo@yahoo.com

Vijay Kumar Singh  
Professor  
Ansal University Gurgaon  
Email:  
vijaykumarsingh@ansaluniversity.edu.in

**Abstract**—This paper is discussing about the road accident severity survey using data mining, where different approaches have been considered. We have collected research work carried out by different researchers based on road accidents. Article describing the review work in context of road accident case's using data mining approach. The article is consisting of collections of methods in different scenario with the aim to resolve the road accident. Every method is somewhere seeming to productive in some ways to decrease the no of causality. It will give a better edge to different country where the no of accidents is leading to fatality of life.

**Index Terms**—Data Mining, Cluster Analysis and Algorithms, Naive Bayes Classifiers, Hybrid Decision Tree, Association Rule Mining.

### I. INTRODUCTION

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 its 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 are undesirable. Many times, it happened that road accident incidents are more common at certain places that can help in identifying factors behind them. Association rule mining is a technique that identifies the correlation in different parameter of road accident. [6]

### II. LITERATURE SURVEY

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 [2].

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 [3]. 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. [4]

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. [9].

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 [5].

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 [6].

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 [10].

Zhexue Huang, (1998), concludes that Ethiopia has the highest rate of Road Traffic Accident (RTA), due to major transport option is only road, instead of train or airways. Report state that approximate one million deaths and nearly 50 million injuries each year. Author has also applied data mining techniques to the connected road feature to accident severity in Ethiopia and develop rule to improve safety. The work support that the accidents are not randomly scattered and alone the drivers are not involved in accident at random but they are based on various circumstances like vehicle speed, road and car condition etc. The objective of this research was to find the applicability of data mining technique in support of road accident analysis in preventing and controlling vehicle accidents, which easily leads to fatality and harm to body. [1, 7].

M. Sowmya, (2013), shown the study work on traffic accident data produced by transport department of government of Hong Kong in 2008. This study applies Naive Bayes, J48, AdaBoostM1, PART and Random Forest classifiers for predicting classification accuracy to analyse the performance. The classification accuracy on the test result reveals for the following three cases such as accident, vehicle and casualty [11].

Sohn S., (2003), studies multiple algorithms on data mining. He also suggested that fusion algorithm could give effective result. He also emphasis that fusion algorithm is better than single classifier techniques. He also claimed that in term of classification accuracy, DSA better than the neural network or decision tree. He has study in Korean environment and claimed that his study proves an effective way in homogenous environment [12].

Depaire B, (2008), analysed whether cluster analysis can be used as a traffic accident segmentation technique or not. Author modelled seven clusters in this theory and make sense and add value to subsequent injury analysis [13].

Mario De Luca et, al., (2011, 2013) shown the study of analysis of after-before approach using cluster and multivariate methods, in a segmented area in Italy. The two methods which is mainly taken in study is cluster analysis and Hard C means Methods proved effective in the study

carried out by author is easy and seen effective in meaningful ways [14-15].

Naina et. al., (2016) used the classification model, with Iterative Dichotomiser 3 and decision tree algorithm. She also compares the existing algorithm with enhanced algorithm C 4.5 with the use of WEKA tool she has shown that its useful in the case when we are using big data sets and the results are quite impressive. [16].

Shanti et. al., (2011) used classification algorithm for vehicle collision patterns. The results prove that random tree approach better than other techniques. The classification algorithm applies to this data set are C4.5, C-RT, CS-MC4, Decision List, ID3, Naïve Bayes and RndTree. The achieved results prove that RndTree technique is better and accurate than other algorithms in collision cases which fatality rate increase in road accidents [17].

Sami Aryamo et, al., (2009) Finland submitted a report which comprise of number of data mining algorithm. The team has very nicely started the data mining algorithm starting from basics to advance level. The long report suggest that the road accidents are due to variety of issue, however the single road lead to more fatality cases. This report present the number of cause of accidents results from the young ages to old drivers. Author also suggest that association rule play significant rule in the study. The report also shown that rule mining is easy in cluster analysis and it can be categorised in different strata, which give effective result as well [18]

Dinesh Mohan et. al., (2015) report state that in India there are huge number of fatality due to poor road conditions. He has discussed number of variable responsible for road accidents, report suggest the number of graphs supporting the accidents causes, authors believe that the data mining can be adopted to find the systematic avoid of road accidents, the number of approaches available in data mining has given a number of futuristic scope in the road accident prevention scheme. The study carried out in IIT Delhi, so it can be considered at one of most trusted document for the subpart of road accident issues [21].

### III. CONCLUSION

In this paper, we have collected multiple researchers' works together in single document as review and discussed about the contribution towards impact of road and traffic accident on human life and society. This survey 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 survey will also help us to find better mining technique in this kind of problem.

In the expansion phase, it's our endeavour to sketch better work to resolve traffic accident severity 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.mecs-press.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 Aryamo 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. Papolizio, 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, Transportation Research and Injury Prevention Programme, IIT Delhi 2015.