International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8, Issue-1, May 2019

An Experimental Research of Accident Report using Data Mining

S K M Sasikomala, P Mayilvahanan

Abstract*: *Traffic accident rates are increasing extremely every year. Though engineers and scholars use the latest technologies to build safe vehicles, accident reports are inevitable. Even though the vehicles are designed with state of the art technology, the accidents are unavoidable. The necessity of developing a model for predicting the dangerous crash patterns that classifies automatically the type of severity of the injuries by using a study of traffic accidents are immense. These roadway models and behaviours are cooperative in building traffic control measurements. To get the highest possible decrease of the accidents with limited economic resources, there should be a detailed study about unbiased and technical surveys and also the root cause of the accident is vital. Data mining is a technique in which from the enormous amount of data in database, hidden patterns will be taken out. It is mainly applied in detection and prediction, surveillance, and fraud analytics, etc. We can learn and recognize about complex patterns by using machine learning techniques in data mining. We can make intelligent decisions and predictions using the customized and available data. In this paper, accident exploration and traffic analysis can be proposed by certain data mining techniques. The idea is to collect traffic accident dataset and to apply data mining techniques. The type of damages are classified.

Index Terms*: *Accident report, CART, data mining, traffic, ANN.

I. INTRODUCTION

Traffic accidents leads to damages and mortalities that affects the society. Clear analysis has to be done for the cause of these accidents. Researchers are mainly concentrating to discover the factors that cause severity of the driver injury during the traffic accidents. India is one of the countries with a huge population among the rapid developing nations in the domain. As a outcome the road transportation is increasing tremendously. Increase of travel speed and the volume of people travelling causes traffic accidents and resulting into severe crashes. The recent traffic and transportation safety problems have raised great concerns and the sustainable development has resulted into one among of all the challenging key issues across the globe.

Revised Manuscript Received on 30 May 2019.
***** Correspondence Author
S K M Sasikomala*, Research Scholar, Vels Institute of Science, Technology & Advanced Studeies, VISTAS, Vels University, Chennai, India.
P Mayilvahanan, Professor, Department of Computer Application, Vels Institute of Science, Technology & Advanced Studeies, VISTAS, Vels University, Chennai, India.

The World Health Organization (WHO) publicized that among the nations in the world, India is one of the leading country in which the reported number of accidents and casualty level is very high. WHO has recorded that India is leading with 105,000 traffic deaths per year whereas in China, more populated than India, only 96000 deaths happening on road. According to a survey conducted in 178 countries, 300 people dying in India every year due to traffic accidents. The traffic accidents also lead to severe injuries and disabilities.
When the vehicle is colliding with another vehicle, with pedestrian or with animal either by serious or negligible mistake, it is resulting into minor or severe injury to human or animal. If the collision is happening with any of the things like bridge lines, barrier gates, road divider, etc., it is resulting in public damage. At any high cost, it is leading to death. The road accident data available is getting doubled every year throughout the world. Data mining approach is mainly accustomed extract the unknown patterns since huge data base by using techniques like sssociation, sorting, etc.,
Lot of importance is given on analysing the accident crash related factors, on the basis of a thorough study about the death ratio growth. The main part of concern is the fatal brutalities caused by the traffic accidents. One of the main thing that influences the fatal rate is the collision pattern. It is very difficult to investigate and extract the critical information without using the advanced data analytics tools when the increase of accident database every year.
The important and the perilous records about the humanity is verified and retained in a system called Traffic Reduction System. It is assisting in providing precautionary measures to save the life. It is possible to analyse the risk areas and factors that leads to accident, death, major injury and minor injuries by leveraging the information available in the Traffic Reduction System. This leads to an influence in the society due to these cruelties of damages. In order to prevent and control the road accidents, the development of model by applying data mining technology is very much required.
The analysis is started in the areas where most of accidents takes place such as signal intersections. The bruise severity is divided into 3 categories: disabling bruise, possible bruise and no bruise. The bruise severity is divided into 3 categories: disabling bruise, possible bruise and no bruise. Decision Tree and Neural Networks are used for classifying the data. To partition or to divide the data into subset and to train the classifiers, each subset is used again by applying a clustering algorithm.

Data is analysed for getting numerous analytical and technical information. This is used mainly to shrink the stream of traffic fates. Data mining techniques are cast-off in analysing many fields to predict and detect the correlations.

II. LITERATURE SURVEY

This survey is used to join the accidents on the road due to traffic information on the urban streets and rural highways. The major reason or factors for the road accident can be determined by using the data mining techniques like K-means and SOM clustering algorithms. It is mostly concentrated to be familiar with the research, machine learning which are done automatically made the smart decisions based on the information. In the recent years, misfortunes are the main cause of bruises and deaths in the world. Road patterns configurations are helpful in the growth of circulation welfare reduction procedure. Spatial dataset and Geographic information systems (GIS) can also be used for identification of information stored with the perturbed spatial patterns.

The study of spatial and time marked information of road accidents can be together with the GIS data, the spatial properties constructions can be empowered and the time-stamped spatial dataset creation are then addressed. Development in overall safety can be attained by reducing and removing such factors. The analysis of crash data referring from the year (2006- 2008) can be analysed by controlling data mining tool and also classification and regression tree (CART).

Checked logistic worsening prototypical with arrangement shrub technique in finding social-demographic jeopardy dynamics which need impacted sadness prestige of females in a detached accidental extents. They designed and implemented a method called classification tree method based on diagnosis assess. Non-parametric classification tree methods has been been implemented by Hang et.al for analysing Taiwan accident information from 2001.The CART method implemented by them helps to identify the relationship between driver vehicle features and bruise severity, highway/environment and accident variables. Weidman et.al evidenced that the hybrid SVM method having improved competence of taking nonlinear relationship involved variables and had best classification rate than MARS, CART and SVM while examining the credit card data. Nojun analysed the restriction of Mutual Information Feature Selector (MIFS) and advised a new method to overcome this limitation. Isabelle et.al explained the basics of feature selection and summarized the steps to resolve a feature selection problem. The execution of several features selection algorithms has been conversed.

III. RESEARCH METHODOLOGY

The technique of Data mining has been classified into two types. It is based on the forecast and recognition of the patterns. Directed data mining and In-directed data mining are the types of data mining. The decisive result can be achieved using directed data mining that is based on a top down approach. Bottom up approach is used in in-directed data mining. The user recognises the patterns based on a set of information and the data. Predictive modelling and descriptive modelling is used in directed data mining and in-directed data mining respectively. The widely used techniques are association, analysis, classification, clustering, prediction and segmentation description in most of the cases. We are describing the following techniques here.

CART: The term **Classification And Regression Tree (CART)** investigation is a method in which both the classification analysis and Regression tree is used. Regression tree is used, if the forecasted outcome is belonging to class in which the data is found. The classification analysis is used, if the forecasted result is measured to be a real number. The both tree analysis, the construction method of the trees will be the same, but the procedure used to control the place of splitting will be different. The trees used in this technique provide a knowledge method for non-parametric decision tree knowledge method based on the dependent variable used.

Decision trees are created by bearing in mind some rules based on the variables in the modelling data set. Originally variable vales are selected to make sure the explanations based on the dependent variable and to find the better split. This process is repeatedly applied when a rule is splits and chosen the node into two and the resultant is said to be child nodes. It's a repetitive or a recursive method. When the CART finds there is no further gain, the splitting of the node is stopped from execution. The splitting is stopped even when pre-set stopping rules are found. The pruned tree will finally be found and terminal node also be found in all the branch of the tree. All of this put together forms a rule sets.

ARTIFICIAL NEURAL NETWORK

A computational designing technique is described as an artificial neural network (ANN). It is referred as neural network. This is a mathematical technique based on biological neural system. In this technique, a group of artificial neurons are connected together and to provide computation, connection technique is used which comprises of methods. This method is an adaptive method, and based on the external and internal information it changes its structure that occurs during the learning period. Here the connection weights are applied between the cells. Information is assembled, saved and generalized based on the weights. This is done using transient learning process. The results will be highly reliable, if the variables are in nonlinear relation.

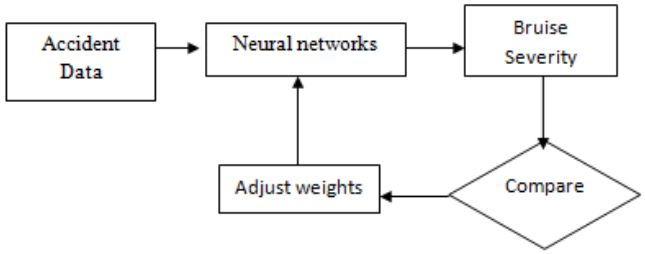

Fig.1 Artificial Neural Networks

HYBRID LEARNING ARTIFICIAL NEURAL NETWORKS

Back propagation technique is applied in the Multilayer perception that comprises the neural network architecture. By using this method in a particular training pattern, the error E will be reduced.

The descent will be in relation with the total training set (E)/W_{mn},where this weight W_{mn} needs to be adjusted after the m-th input toward the n-th output, in the batched mode alternative

$W_{mn}(n)=-(E/W_{mn})+W_{mn}(n-1)$

Momentum and the learning rates are taken into consideration as parameters. The parameters chosen should be the best that results in providing a success in the training. Only few methods are found to select the ideal number of nodes in the unknown layer. This needs to be analysed with all of the parameters in an experimental approach. The best variable will then be selected that helps to divide the data into couple of partitions. The resultant branch will then be of disjoint sets, where each branch is homogenous that contains class variables. The splitting will be applied several times in each of the branch.

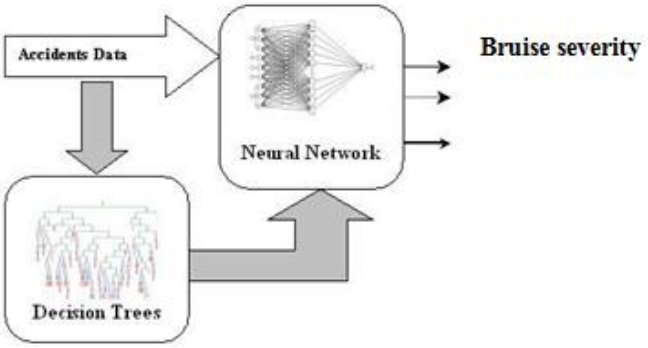

Fig.2 Hybrid Learning ANN model for accident data

ACCIDENT DATA SET

The data used here is based on the survey performed by GES (General Estimate System) and NASS (National Automotive Sampling System).The GES used the data from a sample of more than 6 million accident history in the United States. This is done by a detailed study about the history of accidents from the year 1995 to 2000 that contains a total of more than 400,000 cases. These records contain information about the drivers but not the passenger's. It contains the number describing the police jurisdiction number, labels of year, case number, primary sampling unit, month, person number, travel speed, speed limit pattern of crash, highway exterior form, vehicle number, condition of headlight, region, vehicle model and make; driver's age input, occupation, gender, vehicle body type, alcohol usage (if any), initial point of impact, and the output severity of bruise. The bruise severity can be classified into 5 categories: Fatal bruise, possibility of bruise, Non devastating bruise, No bruise and devastating bruise. Based on the given data, the bruise cases are calculated and categorized. The highest percentage is of no bruises and lowest percentage is of fatal bruises.

IV. EXPERIMENTAL RESULTS & IMPLEMENTATION

Table.1 Assessment Work

Inaccuracy Ratio									
Prediction Value			Misperception Milieus						
Assessment	Evoke	Accuracy		Disastrous	Injure	Nobody	Possible Injury	Non undermine	Indefinite	Amount
Disastrous	0.9898	0.0101	Disastrous	35162	0	1	2	1	4	35170
Injure	0.6265	0.3962	Injure	5	5863	44	260	2543	10	8725
Nobody	0.9462	0.1375	Nobody	3	32	14375	77	187	88	14762
Possible Injury	0.2540	0.3972	Possible Injury	8	850	1200	1430	1543	12	5043
Non undermine	0.6134	0.4115	Non undermine	7	2346	657	576	475	12	4073
Indefinite	0.4553	0.2761	Indefinite	2	79	165	20	43	365	674
			Amount	35187	9170	16442	2365	4792	491	68447

There is a necessity to develop machine learning approach that categorizes the bruise severity based on intelligent models. This provides good understanding about the connection between the driver, road patterns, vehicle design, surroundings, severity of damages and driver injury. The data should be analysed accurately, so that it provides the information to prevent and reduce the road accidents.

To construct the model for predicting the severity of the accidents, input and output variables are taken into consideration. The impact is divided into nine categories: No impact/damage, RH side,LH side,front,backside,Back R corner, Back L corner, Front L corner and front R corner. As previously mentioned, the input variables are considered into account. Heretwo classes are used. Affirmative class which involves the affirmative class and all the other classes that belongs to a harmful class. Here 1 is the value of the affirmative class whereas 0 is the value of harmful classes. Among all of the accident reports considered, 50% records are with no bruises, 25% are with possible bruises, 21% with non-incapacitating bruises, 2% with devastating bruises and 2% with deadly bruises.

V. CONCLUSION

Based on the previous work on the accidental reports, most of the accidents with fatal injuries are occurring on a single roadway main road. In these roads the travelling speed limit is between 80-100 km/hour. Through this survey, the road areas which are risky and the drivers who are responsible for accidents can be detected. But the data needs to improve the road construction plans could not be predicted. The Literature review helps to find the attributes involved in the accidents in the international cases. It is clearly stated that the road accidents are not only caused due to the road construction and also due to the improper road traffic planning. Among all the five classes categorized here, fatality of the driver has the highest influence on the society.

In this paper, though several inputs are considered, speed is not used since the speed of the vehicles was not known. In future we can take speed into consideration to enhance the model performance.

REFERENCES

1. Abdel-Aty, M., and Abdelwahab, H., Analysis and Prediction of Traffic Fatalities Resulting From Angle Collisions Including the Effect of Vehicles' Configuration and Compatibility. *Accident Analysis and Prevention*, 2003.
2. Abdelwahab, H. T. and Abdel-Aty, M. A., Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections. *Transportation Research Record 1746*, Paper No. 01-2234.
3. Bedard, M., Guyatt, G. H., Stones, M. J., & Hireds, J. P., The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities. *Accident analysis and Prevention*, Vol. 34, pp. 717-727, 2002.
4. Buzeman, D. G., Viano, D. C., & Lovsund, P., Car Occupant Safety in Frontal Crashes: A Parameter Study of Vehicle Mass, Impact Speed, and Inherent Vehicle Protection. *Accident Analysis and Prevention*, Vol. 30, No. 6, pp. 713-722, 1998.
5. Dia, H., & Rose, G., Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data. *Transportation Research C*, Vol. 5, No. 5, 1997, pp. 313-331.
6. Nabi H., Salmi L.R., Lafont S., Chiron M., Zins M., Lagarde E., Attitudes associated with behavioral predictors of serious road traic crashes: results from the GAZEL cohort, Injury Prevention 13 (1) (2007), 26–31.
7. Yu B., Lam W.H.K., Tam M.L., Bus arrival time prediction at bus stop with multiple routes, Transportation Research Part C 19 (6) (2011), 1157–1170.
8. Krishnaveni S., Hemalatha M., A Perspective Analysis of Traffic Accident using Data Mining Techniques 23 (7) (2011).
9. Geetha K., Vaishnavi C., Analysis on Traffic Accident Injury Level Using Classification, Accident Analysis & Prevention 5 (2) (2015).
10. M. Sowmya and Dr.P. Ponmuthuramalingam, Analyzing the Road Traffic and Accidents with Classification Techniques, International Journal of Computer Trends and Technology 5 (4) (2013).
11. Ali Tavakoli Kashani, Afshin Shariat-Mohaymany, Andishe Ranjbari, A Data Mining Approach To Identify Key Factors of Traffic Injury Severity 23 (1) (2011) 11-17.
12. Jayasudha K., Chandrasekar C., An Overview Of Data Mining In Road Traffic And Accident Analysis, Journal of Computer Applications 2 (4) (2009).
13. Rajesh, M., and J. M. Gnanasekar. "GCCover Heterogeneous Wireless Ad hoc Networks." Journal of Chemical and Pharmaceutical Sciences (2015): 195-200.
14. Rajesh, M., and J. M. Gnanasekar. "Consistently neighbor detection for MANET." Communication and Electronics Systems (ICCES), International Conference on. IEEE, 2016.
15. Sohn, S. Y., & Lee, S. H., Data Fusion, Ensemble and Clustering to Improve the Classification Accuracy for the Severity of Road Traffic Accidents in Korea. *Safety Science*, Vol. 4, issue1, February 2003, pp. 1-14.
16. Tavris, D. R., Kuhn, E. M, & Layde, P. M., Age and Gender Patterns In Motor Vehicle Crash injuries: Improtance of Type of Crash and Occupant Role. Accident Analysis and Prevention, Vol. 33, 2001, pp. 167-172.
17. Yang, W.T., Chen, H. C., & Brown, D. B., Detecting Safer Driving Patterns By A Neural Network Approach. *ANNIE '99 for the Proceedings of Smart Engineering System Design Neural Network, Evolutionary Programming, Complex Systems and Data Mining*, Vol. 9, pp 839-844, Nov. 1999.
18. Zembowicz, R. and Zytkow, J. M., 1996. From Contingency Tables to Various Forms of Knowledge in Database. Advances in knowledge Discovery and Data Mining, editors, Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. AAAI Press/The MIT Press, pp.329-349.
19. Abraham A., Meta-Learning Evolutionary Artificial Neural Networks, Neurocomputing Journal, Elsevier Science, Netherlands, Vol. 56c, pp. 1-38, 2004.
20. Moller A.F., A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural Networks, Volume (6), pp. 525-533, 1993.