ISSN (Online): 2455-7838
SJIF Impact Factor (2017): 5.705

EPRA International Journal of

Research & Development

(IJRD)

Monthly Peer Reviewed & Indexed International Online Journal

Volume: 3, Issue:11,November 2018

Published By :
EPRA Journals

SJIF Impact Factor: 5.705 Volume: 3 | Issue: 11 | November | 2018 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)

A PERSPECTIVE ANALYSIS OF TRAFFIC ACCIDENT USING DATA MINING ALGORITHMS

R. Limya[1]
[1]Research Scholar, Department of Computer Science, Vivekanandha College for Women, Tiruchengode, India,

M. Baskar[2]
[2]Assistant Professor of Computer Science, Vivekanandha College for Women, Tiruchengode,India,

M. Thulasimani[3]
[3]Research Scholar, Department of Computer Science, Vivekanandha College for Women, Tiruchengode, India,

ABSTRACT
Road traffic accidents are the common and cruel issue, it fallout death and injuries of various levels. The traffic control framework is one of the key areas, where basic information in regards to the general public is noted and kept as protected. Different issues of a traffic system like vehicle accidents, traffic volumes and deliberations are recorded at various levels. In association with this, the accident severities are propelled from road traffic accident database. Road traffic accident databases give the cause to road car crash study. In this exploration work, Madurai city road traffic databases is taken to thought, the city having higher number of vehicles and activity and the city having higher number of vehicles and activity and the expense of these misfortune and mishaps greatly affects the financial development of a general public. Conventional machine learning calculations are utilized for building up a choice emotionally supportive network to deal with street auto collision investigation. The calculations, for example, SMO, J48, IBK are actualized in Weka the result of these algorithms are analyzed. In this work, the calculations were tried on an example database of in excess of five hundred things, each with 20 accident attributes. In this paper among the three algorithms, we conclude which algorithm provides more accurate result.

KEYWORDS - *Road Traffic Accident, SMO, J48, IBK.*

1. INTRODUCTION

India is being one of the quickest creating countries on the planet with an immense populace thickness, in view of these the street activity thickness is additionally expanding. As of late, with the development of the volume and travel speed of street activity, number of auto collisions, particularly extreme accidents [1], has been expanding speedily on a yearly premise. The issue of activity wellbeing has lifted incredible worries over the world, and it has turned out to be one of the key issues requesting the maintainable advancement of present day activity and transportation. Movement conditions are a multifaceted framework because of numerous coincidental variables [2], and car crash information has for quite some time been known to be extremely hard to process. Numerous analysts have been made as of late through applying different systems and calculations. Information mining [3] is a blend of strategies, procedures and process in learning disclosure. In other words, it requires a wide assortment of apparatuses running from established and factual procedures to neural systems and other new procedures beginning from machine learning and man-made consciousness for enhancing databases advancement and process enhancement. Six key capacities or exercises of information mining are arranged into coordinated and undirected. Particularly characterization, assessment and forecast are coordinated, when the accessible subtle elements are utilized to assemble models that characterize one specific variable of criticalness as far as the rest of accessible information.

Information mining in car crashes [4] are which finds the concealed learning and guidelines has turned into a fundamental research region in rush hour gridlock security. Lately, the greater part of the movement data examinations are restricted to general factual examination, or, in other words find the principles concealing [5] in car crash data. Measurable examination does not have the ability of guide showing and spatial investigation, and consequently it can't locate the spatial circulation trademark and connection between car crashes [6] and street organize components. Along these lines, through this exploration work an endeavor has been made to apply information mining instruments and strategies in breaking down and deciding fascinating examples particularly as for conceivable outcomes of street auto collision, on street mishaps information at Madurai locale movement control System.

2. BACKGROUND STUDY

Madurai is the third leading city in Tamil Nadu with a population of more than 10 lakhs. The city includes more than 20,000 of tiny, medium and large industries and mills are running popularly in Madurai.

2.1 Accident Statistics

Madurai is ranked 38 rd proportionate to its population in the number of fatalities in road accidents. A report on "Accidental Deaths and Suicides in India 2010" released by National Crime Records Bureau (NCRB) given that a total of 1,131 accidents on the city roads. Ranking is not done on the basis of number of accidents but on the basis of number of accidents proportionate to the population. The following table (Table I) shows the road accident details includes Total number of accidents, number of persons injured and number of persons killed during the year 2016 to 2017

Table 2.1 Road Accident

Year	City	Total No. of Accidents	No. of persons injured	No. of persons killed
2016- 2017	Madurai	4789	6543	3802

2.2 Vehicle Statistics

In the year 2016 alone, the Circle of the Transport Department is comprising Madurai, Theni and the Kambam districts find a registration of 2,17,785 new vehicles, both 2-wheelers and 4-wheelers, as against the registration of 2,15,627 vehicles of the previous year. The table (Table II) describes the number of vehicles registered in Madurai city.

Table 2.2 Number of Vehicles Registered

Vehicle type	No. of registered vehicle
Auto-rickshaw	501
Motor Cab	788
Maxi Cab	364
School Bus	202
Ambulance	30
Light commercial vehicle	3254
Lorry	966
Motor Cycle	64192
Scooter	20743
Moped	7235
Motor Car	15130

3. 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 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. Analysis of Datamining Technique for Traffic

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],

4. DATA MINING CLASSIFICATION ALGORITHM

This study is generally focus on predict potential of road traffic accident in a particular area using machine learning techniques. There are three algorithms are SMO, J48,IBK.

4.1 Support Vector Machine

Support Vector Machines are based on the idea of decision planes that characterize decision boundaries. A decision plane is vital one which isolates between a arrangement of objects having diverse class participations. Support Vector Machine (SVM) is fundamentally a classier method that performs characterization tasks by building hyper planes in a multidimensional space that isolates cases of different class labels. SVM supports two special tasks like regression and classification and can handle various constant and categorical factors. To develop an optimal hyper plane SVM uses an iterative preparation algorithm, for limit the error function. According to error function, SVM models can be classified into two separate groups like Classification SVM and Regression SVM.

1. Classification SVM For this type of SVM, training involves the minimization of the error function:

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i$$

subject to the constraints:

$$y_i\left(w^T\phi(x_i)+b\right)\geq 1-\xi_i \text{ and } \xi_i \geq 0, i=1,\ldots,N$$

where C refers the capacity constant, w refers the vector of coefficients, b is a constant, and refers parameters for handling non separable data (inputs). The index i label the N training cases. It should be renowned that it is larger than C, then the error is penalized. So the C must be chosen with carefully to avoid over fitting problem. The SMO is based on the Support Vector Machine process.

4.2 Decision tree

A decision tree [13] is a simple flowchart that selects labels for input values. This flowchart contains the decision nodes, which will check feature values, leaf nodes, and assign labels. To select the label for an input value, begin at the flowchart's initial decision node, called as its root node. This node contains a condition for checks one of the input value's features, and also selects a branch based on that feature's value. Following the branch that describes the input value, attain at a new decision node, condition on the input value's features. Then continue following the branch selected by each node's condition, until arrive at a leaf node which provides a label for the input value. Once have a decision tree, it is straightforward to use it to assign labels to new input values. It encompasses of one leaf for each possible feature value, identifying the class label that should be assigned to inputs whose features have that value. In order to build a decision stump, first decide which feature should be used. The simplest method is to build a decision stump for each possible feature, and get which one achieves the highest accuracy on the training data. Once feature has been picked, build the decision stump by assigning a label to each leaf based on the most frequent label for the selected examples in the training set. The J48 is based on the Decision tree process.

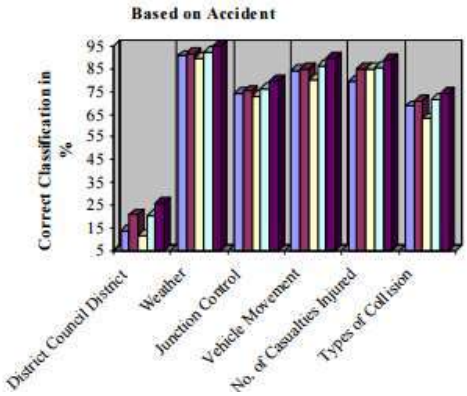

Fig.4.2 Comparison of SMO, J48,IBK.

4.3 K-Nearest Neighbor

The k-nearest neighbor algorithm is the simplest of all categorization algorithms. In pattern recognition, the k-nearest neighbor algorithm (k-NN) is a process for classifying objects based on neighboring training examples in the feature space. k-NN is a form of illustration-based learning, where the task is only approximated nearby and all calculation is late until categorization. It can also be used for regression. If k=1, then the object is basically assigned to the class of its nearby neighbor. In binary (two class) categorization troubles, it is useful to choose k to be an odd number. The similar method can be used for regression, by simply conveying the goods value for the object to be the average of the values of its k nearest neighbors. It can be helpful to weight the offerings of the neighbors, so that the nearer neighbors donate more to the average than the more distant ones. The IBK is based on the K-Nearest Neighbor process.

5. CONCLUSIONS

The objective of this research work is to explore the possible application of data mining technology at Madurai city road traffic accident data for developing a classification model. The classification model could support the traffic officers at Madurai city traffic office for making decisions in traffic control activities. The most best performing J48 classifier was chosen taken into account the reliability of the rules it generated and also the number of false negatives it reduced, finally its predictive accuracy was evaluated and analyzed.. The classification accuracy of the J48 was tested, and it showed an accuracy of 94%.

REFERENCES

1. *H.Nabi, L.R.Salmi, S.Lafont, M.Chiron, M.Zins, and E.Lagarde, "Attitudes associated with behavioral predictors of serious road traic crashes: results from the GAZEL cohort," Injury Prevention, vol.13,no.1, pp.26–31 ,2007.*
2. *B.Yu, W.H.K.Lam, and M.L.Tam, "Bus arrival time prediction at bus stop with multiple routes,"*

Transportation Research Part C , vol. 19, no. 6, pp. 1157–1170, 2011.

3. *L.-Y. Dong, G.-Y. Liu, S.-M. Yuan, Y.-L. Li, and Z.-H. Wu, "Applications of data mining to traffic accidents analysis," Journal of Jilin University Science Edition, vol.44, no.6, pp.951–955, 2006.*
4. *D.-H. Lee, S.-T. Jeng, and P. Chandrasekar, "Applying data mining techniques for traffic incident analysis, "Journal of the Institution of Engineers, vol.44, no.2, pp.90–101, 2004.*
5. *Marie-France Joly, Robert bourbeau and Jacques Bergeron, "What Can We Learn from the Experience of Risk Location Identification?", Proceedings of International Conference on Traffic Safety, New Delhi, India, January 1991.*
6. *Babkov, V.F, Road Conditions and Traffic Safety; Mir Publishers; Moscow.*
7. *T. Tesema, A. Abraham, and C. Grosan, "Rule mining and classification of road traffic accidents using adaptive regression trees. I," Journal of Simulation, vol. 6, no. 10, pp. 80–94, 2005.*
8. *M. Hirasawa, "Development of traffic accident analysis system using GIS," Proceedings of the Eastern Asia Society for TransportationStudies, vol. 10, no. 4, pp. 1193–1198, 2005.*
9. *H. Nabi, L. R. Salmi, S. Lafont, M. Chiron, M. Zins, and E.Lagarde, "Attitudes associated with behavioral predictors of serious road traffic crashes: results from the GAZEL cohort,"Injury Prevention, vol. 13, no. 1, pp. 26–31, 2007.*
10. *PramodAnantharam, KrishnaprasadThirunarayan, AmitSheth, "Tra_c Analytics using Probabilistic Graphical Models Enhanced with Knowledge Bases" fpramod, tkprasad,Kno.e.sis - Ohio Center of Excellence in Knowledge-enabled Computing Wright State University, Dayton, USA.*
11. *S.Krishnaveni, Dr.M.Hemalatha, "A Perspective Analysis of Traffic Accident using DataMining Techniques" in International Journal of Computer Applications (0975 – 8887) Volume 23– No.7, June 2011.*
12. *Tibebe Beshah1, Shawndra Hill2," Mining Road Traffic Accident Data to Improve Safety: Role of Road-related Factors on Accident Severity in Ethiopia".*
13. *Lior Rokach, Oded Maimon, "Decision Trees", Department of Industrial Engineering, Tel-Aviv University.*
14. *Yang Song, Jian Huang, DingZhou, Hongyuan Zha, and C. Lee Giles, "IKNN: Informative K-Nearest Neighbor Pattern Classification", Springer-Verlag Berlin Heidelberg, PKDD 2007, LNAI 4702, pp. 248–264, 2007.*
15. *H. Wan-Jo Yu, "Data Mining via Support Vector Machines: Scalability, Applicability, and Interpretability", Research work.*