Identifying Efficient Road Safety Prediction Model Using Data Mining Classifiers

Durga Karthik, P. Karthikeyan, S.Kalaivani, K.Vijayarekha

Abstract— Road accidents are a major cause of death and disabilities. The aim of the traffic accident analysis for a region is to investigate the cause for accidents and to determine dangerous locations in a region. Multivariate analysis of traffic accidents data is critical to identify major causes for fatal accidents. In this work, accident dataset is analysed using algorithmic approach, as an attempt to address this problem. The relationship between fatal rate and other attributes including collision manner, weather, surface condition, light condition, mobile users and drunken driving are considered. Prediction model using various data mining classifiers such as Bayesian, J48, Random Forest will be constructed to enhance safety regulations for a region.

Keywords: Collision manner, Weather, Surface condition, Bayesian, J48, Random Forest.

I. INTRODUCTION

A road traffic accident (RTA) is any injury due to crashes originating from, terminating with or involving a vehicle partially or fully on a public road. The various reasons for an accident could be weather, light condition, surface condition, drunken driving, over speed, drivers fault, vehicle problems and latest additions being mobile usage while driving. As the usage of private vehicles has increased, it has lead to increased in accidents in certain areas in peak hours and in densely populated areas across the world. The accident data sets can provide greater insight into the problem for identifying accident prone areas.

II.LITERATURE WORK

The accident record in database of police in Iran has analysed based on the CART and LR regression [1]. Road Traffic Accident database of china's public security department has understood the various factor using association rule and binary logistic model [2]. Accident data for rural highways for the province of Granada (South of Spain) predicted using CART and Decision rules (Decision trees) [3]. Traffic dataset on the mainline G60 freeway in Shanghai modelled using ADASYN, SVM and Random Forest [4]. Predictive mining can be applied on previous road accident data with Association rule, Apriori algorithm and Naïve Bayes algorithm [5].

Twitter based data posted by regional traffic agencies where analysed based on data mining algorithms such as Apriori algorithm, Naïve Bayes classifier, K-means clustering algorithms [6]. National Highways Authority of India (NHAI) which covers accident historical data has analysed and predicted using Random Tree, J48 and Naive Bayes [7]. Logistic regression model can be used to identify the prediction factors like crashes and crash–related injuries, to perform a risk assessment of a region [8]. Various combinations of cluster analysis, regression analysis and other techniques can be grouped on homogenous accident data to determine RTA risk [9]. Using the accident data of National Highway in Taiwan, the authors developed classification and regression tree to establish the relationship between traffic accidents and environmental factors [10].

Identified and collected information about hazardous point on the complementary road network of Andalusia and Spain and data mining techniques are applied [11] to determine road safety . Road accidents in India at National, State and Metropolitan city level has been analysed with its issues and challenges [12]. Historical traffic accident data of California, severity factor analysed with Naïve Bayes classifier and Decision Tree classifier [13]. Our work is to predict accident prone areas in a region using data mining classifiers.

III.EXPERIMENTS & RESULTS

A. Description of the Dataset

Kumbakonam is a region in Thanjavur District, India. It is famous pilgrim centre and densely populated as it is located in the river banks. 10 years accident data containing the following informations: Accident Location, Road Bound, Accident Time, Surface Condition, Collision Type, Reported Cause, Accident Zone, Lighting Condition, Persons Injured, Persons Killed.

B. Data Preprocessing

The dataset was preprocessed to remove missing data and the required features were selected as shown in Fig.1.

Date of...	occ time	Latitude	Longit...	Casual...	Sex of...	Age of...	Cau
31/12/...	15:45	10.960...	79.3833	Serious	Male	38	1
04/01/...	17:00	10.960...	79.3833	Slight	Female	50	1
10/01/...	12:10	10.960...	79.3833	Slight	Male	26	1
11/01/...	06:50	10.960...	79.3833	Slight	Female	22	1
11/01/...	10:30	10.960...	79.3833	Slight	Male	57	1
12/01/...	15:00	10.960...	79.3833	Serious	Male	59	1
13/01/...	13:00	10.960...	79.3833	Serious	Female	53	1

Revised Version Manuscript Received on 05 August, 2019.

Durga Karthik, Asst. Prof., Department of Computer Science and Engineering, SRC – SASTRA Deemed University, Tamil Nadu ,India. **(e**-mail: durgakarthik@src.sastra.edu)

P. Karthikeyan, M.C.A Final Year, SRC – SASTRA Deemed University, Tamil Nadu, India.**(e**-mail: sivaselvakarthik96@gmail.com)

S.Kalaivani, B.Tech(CSE), SRC – SASTRA Deemed University , Tamil Nadu, India.**(e**-mail: dexterouskalai@gmail.com)

K.Vijayarekha, Associate Dean/EEE, SASTRA Deemed University Tamil Nadu, India.**(e**-mail: vijayarekha@eee.sastra.edu)

Fig. 1. Feature selection after preprocessing.

C. Proposed Architecture

D. J48:

The datasets were subjected to algorithms such as J48, Naivebayes, Randomforest to identify a good prediction model. The results of prediction for the above dataset using J48 is given in Fig.2 using JAVA.

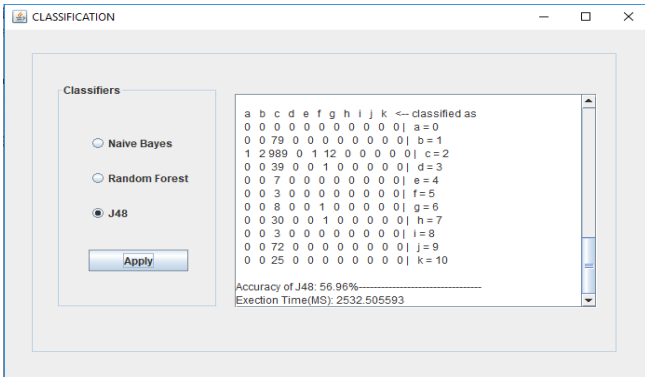

Fig. 2. Showing J48 results for accident datasets

E. Naïve Bayes:

The same datasets were subjected to Naïve Bayes algorithm and Random forest algorithm and the results are shown in Fig.3 and Fig.4.

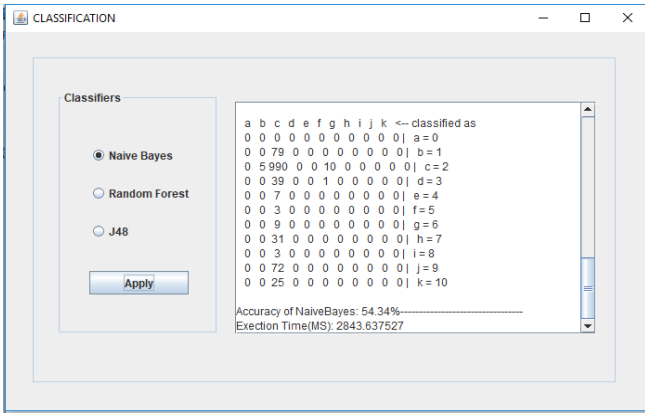

Fig. 3. Classification results using Naive Bayes for accident datasets.

F. Random Forest:

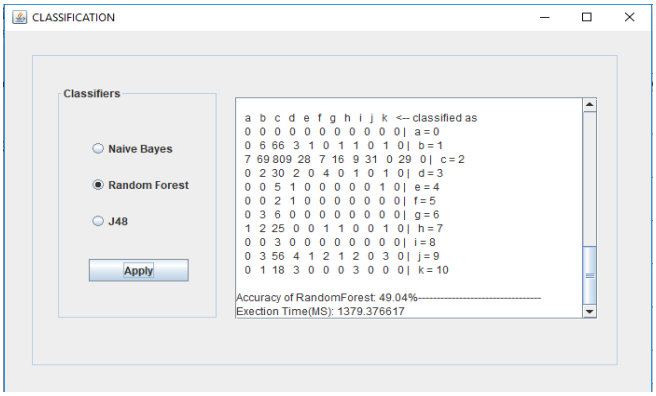

Fig. 4. Classification results using Random Forest algorithm on accident data.

The classification results were compared for accuracy of classification and is found that J48 algorithm provides good accuracy 56.56% than the other two algorithm.

The results of accuracy for the three algorithms for classification of accident data for a good prediction model is given as a table in table.1 and as chart in Fig:5.

TABLE:1 : Classification Accuracy using different algorithms

S.no	Algorithm	Accuracy in %
1.	Naive Bayesian Method	54
2.	Random Forest Method	49
3.	J 48 Method	56.96

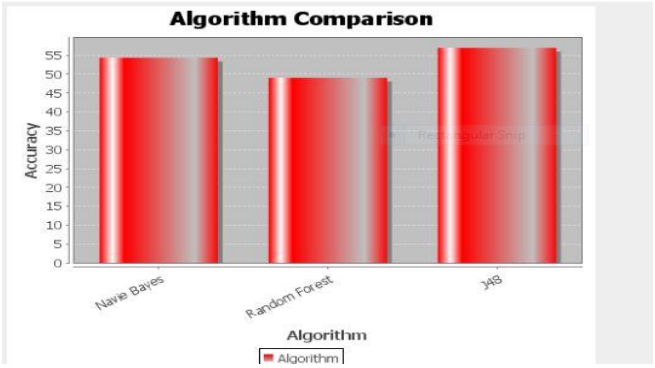

Fig. 5. Classification Algorithms Accuracy

J48 algorithm was used to predict various vehicles such as auto, bus, car, cycle, lorry, can etc that caused more accidents for the dataset of the region. The results revealed that more accidents occurred due to buses followed by lorry and the results is given as a chart in Fig.6.

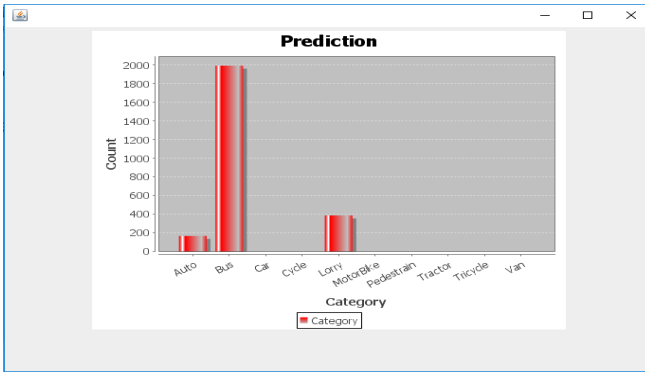

Fig. 6. Prediction results

IV.CONCLUSION

As Classification algorithms can be used for prediction on employing various algorithms on accident datasets, J48 algorithm provides better accurate results with 56.56% , when compared to other algorithms such as Naive Bayes and Randomforest. J48 algorithm can be used to create a prediction model for identifying the various causes of accidents in a region.

REFERENCES

1. **Yanyan Li** and **Toshiyuki Yamamoto et al(**Dec 2017) Understanding factors associated with misclassification of fatigue-related accidents in police record, Journal of safety research.
2. Alireza Pakgohar and Reza Sigari Tabrizi et al(2010) The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach. Published by Elsevier Ltd.
3. López Griseldaa, de Oña Juanb and Abellán Joaquín (2012) Using Decision Trees to extract Decision Rules from Police Reports on Road Accidents. SIIV - 5th International Congress - Sustainability of Road Infrastructures.
4. Jinming You, Junhua Wang, Jingqiu Guo (2017) Real-time crash prediction on freeways using data mining and emerging techniques.
5. Poojitha Shetty, Sachin P C et al (2017) Analysis of road accidents using data mining techniques International Research Journal of Engineering and Technology (IRJET).
6. S. Revathi, Dr.A.Sumithra et al (2017) CERTAIN ANALYSIS ON TRAFFIC DATASET BASED ON DATA MINING ALGORITHMS International Research Journal of Engineering and Technology (IRJET).
7. Baye Atnafu and Gagandeep Kaur (2017) Survey on Analysis and Prediction of Road Traffic Accident Severity Levels using Data Mining Techniques in Maharashtra, India. International Journal of Current Engineering.
8. Tibebe Beshah, Shawndra Hill(2015) Mining Road Traffic Accident Data to Improve Safety: Role of Road-related Factors on Accident Severity in Ethiopia Accidents Analysis and Prevention.
9. Ossenbruggen, P. J., J. Pendharkar, et al. (2001). "Roadway safety in rural and small urbanized areas." Accidents Analysis and Prevention.
10. Chang, L. and H. Wang (2006). "Analysis of traffic injury severity: An application of non-parametric classification tree techniques Accident analysis and prevention " Accident analysis and prevention.
11. Ng, K. S., W. T. Hung, et al (2002). "An algorithm for assessing the risk of traffic accidents." Journal of Safety Research.
12. Luis Martín, Leticia Baena et al (2014) Using data mining techniques to road safety improvement in Spanish roads. Published by Elsevier Ltd.
13. Sanjay Kumar Singh "Road Traffic Accidents in India: Issues and Challenges" World Conference on Transport Research - WCTR 2016 Shanghai.