International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 29 Issue 7 – AUGUST 2022.

Road accident prediction using Machine Learning

Y. Lakshmana pandian[#1] and G. Balamurugan[*2]

[#]*Student, Department of Master of Computer Application. Mohamed Sathak Engineering College. Ramanathapuram, India*

[*] *Assistant professor, Department of Master of Computer Application, Mohamed Sathak Engineering College. Ramanathapuram, India.*

***Abstract*— The study of Influencing factors of traffic accidents are an important research direction in the field of traffic safety. In this paper, the traffic accidents of Shanghai Expressway from April to June 2014 were excavated using association rule mining which generated lots of frequent item sets. The strong rules hidden in these frequent item sets often uncover the association between influencing factors of accidents, which can be used to reduce the occurrence of accidents by breaking them. The rules can also be used to probe usual scenes of accidents, and some corresponding security improvement measures can be taken to prevent the accidents, and ultimately improve the city's traffic safety level. General speaking, association rule mining can produce tons of weak rules, the study first designed a method to calculate minimal Support value of training parameters, and further put forward away to extract strong rules automatically. The results of the experiments showed that these methods proposed in the paper are effective. Therefore, an automatic modeling algorithm using association rules was finally established to promote the effective application of association rule mining on intelligent transportation system.**

***Index Terms*—Influencing factors of accidents, training parameters, intelligent transportation system.**

I. INTRODUCTION

The rapid development of urban expressways not only brings the convenience of transportation, but also causes traffic safety problems. From April to June 2014, over 6,000 traffic accidents occurred on the Shanghai Expressway and the city needs a safer traffic environment. Exploring the impact of the influencing factors of accidents and taking corresponding effective measures to reduce the occurrence of accidents is an urgent need. In recent years, scholars studied the impact of influencing factors of traffic accidents, mainly focusing on people, cars, roads or the environment. Some scholars [1-2] studied on driver's behavior and analyzed the characteristics of the process during changing the lanes to identify dangerous driving behaviors. There are also studies [3] on the impact of road conditions on traffic accidents, they proposed a point that high and steep roadbed will undermine the traffic safety. Also other studies [4] focused on the impact of weather or dynamic traffic flow on accidents. However, most of these studies focused on a single factor (people, cars, roads, the environment)on the impact of traffic accidents. With the development of data mining technology, scholars used a variety of data mining approaches in traffic safety research. Among them, the association rule mining was often used to analyze the relationship between the influencing factors of traffic accidents. The strong association rules can be used to find the network hidden in the accident data. To get them, we could measure the importance and credibility of the rules with the two thresholds Support and Confidence, and sort the validity of the rules by Lift.

The existing association rule mining generally determined the model parameters (such as the minimum Support, etc.) by repeated experiments. For the massive results excavated, it is necessary for the experts to screen useful rules according to personal expertise manually. The method is inefficient and the subjective screening process cannot be translated into an objective algorithm, so it hinders the direct application of association rule mining in intelligent transportation system. In this paper, we proposed a method to calculate the minimum Support in the modeling parameters, and put forward a way to extract strong rules from the massive rules by a clustering method, or automatically filtering out the weak rules based on an expert experience related method. Finally, we built up an automated modeling algorithm using association rules which would better promote the practical application of association rule mining in existing intelligent transportation system.

II. PROPOSED SYSTEM

Classification techniques will be using for identifying the accident prone areas. The accident data records which can help to understand the characteristics of many features like drivers behaviour, roadway conditions, light condition, weather conditions and so on. This can help the users to compute the safety measure which is useful to avoid accidents. The data set can be analysing based on the algorithms will gives the accurate dataset. The models are performed to identify statistically significant factors which can be able to predict the probabilities of crashes and injury

III. ADVANTAGES OF PROPOSED SYSTEM

Proposed system is a real time application.

Analyze the previously occurred accidents in the locality which will help us to determine the most accident-prone area and warn them.

To make predictions based on constraints like weather,

pollution, road conditions, etc.

This can help the users to compute the safety measures which is useful to avoid accidents.

IV. MODULES DESCRIPTION

A. *Kaggle*

The main data set we used for this project was a detailed accident record published by the UK government and hosted on Kaggle. **T**his data set contains details about 1.6 million traffic accidents that took place in the UK between 2000 and 2014. It consists of 33 columns which capture details such as the location, time, severity of the accidents as well as various meteorological and traffic backdrops. In this project, we restricted ourselves to analyzing traffic accidents in Greater London between 2012 and 2014.

B. *Dark Sky*

We believe that weather is a particularly important factor in road accidents. Although the aforementioned Kaggle data set contains meteorological information, we didn't think that it's sufficient. There were the two issues about using the weather condition column of the Kaggle dataset:

It assumes that weather is constant throughout the day. If you've ever been to London, you know why this assumption is problematic — the weather there changes very frequently!

It only contains historical weather records (obviously!). In order to use weather condition as a predictor for *future* accidents, we need a way to get a weather *forecast*. Because of these reasons, we decided to instead make use of the meteorological data provided by Dark Sky, an American company that specializes in weather forecasting. Dark Sky provides both past records and future weather forecast updated on a half-hourly basis, which perfectly serves our purpose.

C. *Feature Extraction:*

Here shows a correlation matrix among the numerical features of the dataset. For a set of highly correlated features, it is standard practice to exclude all but one feature to reduce impact of multi collinearity. However, certain models such as Random Forest, are relatively unaffected by multi collinearity. Therefore, the need to exclude correlated features will depend on the model used. The features and their importance scores for the Random Forest Classifier.

D. *Model Selection:*

Table1 shows the models attempted and metrics used to compare the quality of the model. The best performing model so far is Random Forest with only numerical / floating point predictors.

E. *Prediction:*

The "***Prediction***" section contains an interactive map which will carry out RTA prediction. This visualization will allow users to input a specific particular date/time. Upon making this selection, the website will fetch weather information that correspond to the chosen date/time. These three inputs (date, time and weather) will be sent to our trained model, which in turn will predict probabilities on accident-prone spots. These spots will then be displayed on the map.

V. SYSTEM ARCHITECTURE

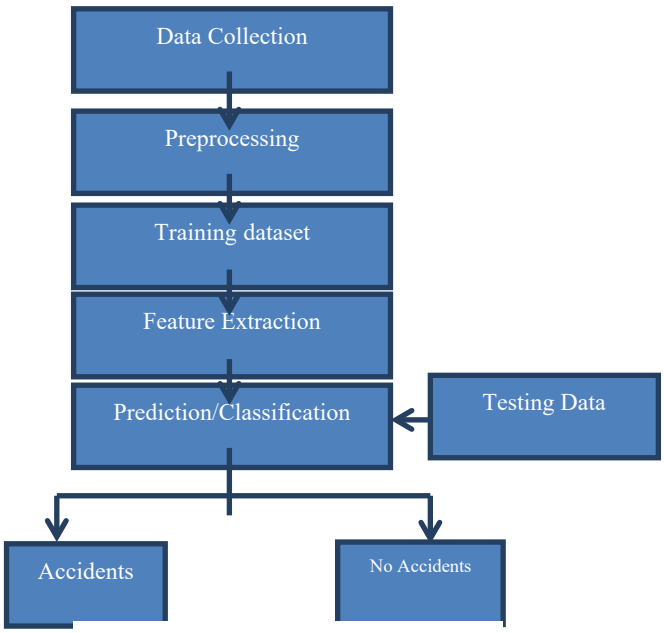

Fig 1: System Architecture

VI. SYSTEM TESTING ANDIMPLEMENTATION

System testing is the stage of implementation, which aimed at ensuring that system works accurately andefficiently before the live operation commence. Testing is the process of executing a program with the intent of finding an error. A good test case is one that has a high probability of finding an error. A successful test is one that answers a yet undiscovered error.

A. *White Box Testing*

White Box testing is a test case design method that uses the control structure of the procedural design to drive cases. Using the white box testing methods, we derived test cases that guarantee that all independent paths within a module have been exercised at least once.

B. *Black Box Testing*

Black box testing is done to find incorrect or missing function

Interface error
Errors in external database access
Performance errors
Initialization and terminationerrors

In 'functional testing', is performed to validate an application conforms to its specifications of correctly performs all its required functions. So this testing is also called 'black box testing'. It tests the external behaviour of the system. Here the engineered product can be tested knowing the specified function that a product has been designed to perform, tests can be conducted to demonstrate that eachfunction is fully operational.

C. *Validation Testing*

After the culmination of black box testing, software is completed assembly as a package, interfacing errors have been uncovered and corrected and final series of software validation tests begin validation testing can be defined as many, but a single definition is that validation succeeds when the software functions in a manner that can be reasonably expected by the customer.

VII. WORKING PRINCIPLE

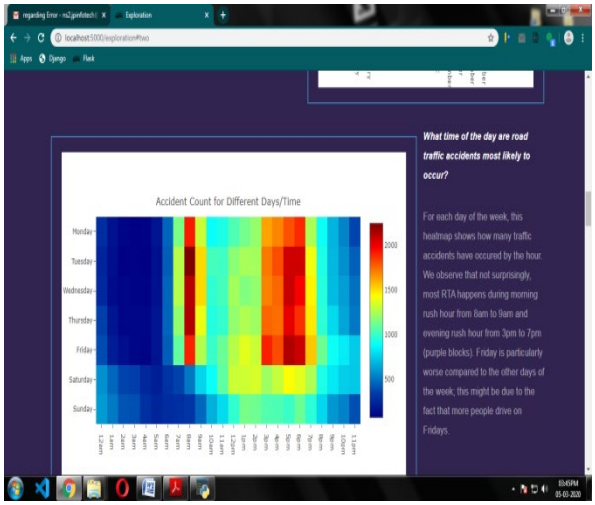

Fig 2: Accidents most likely to occur

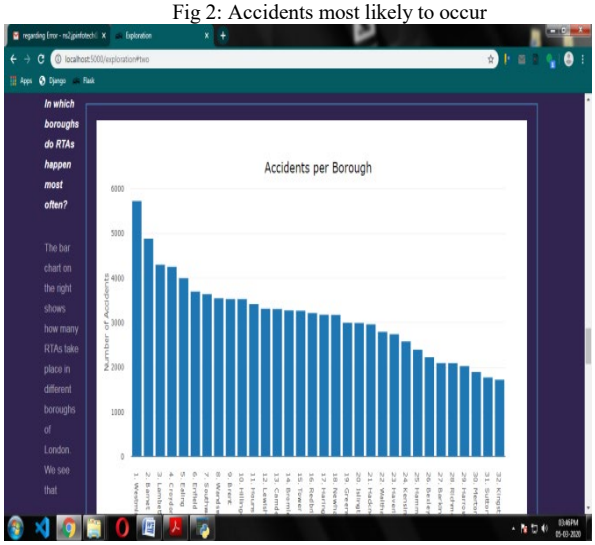

Fig 3: Accidents per bought

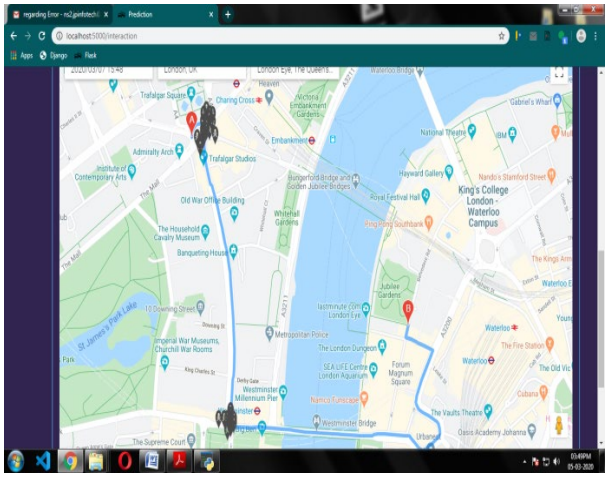

Fig 4:Route

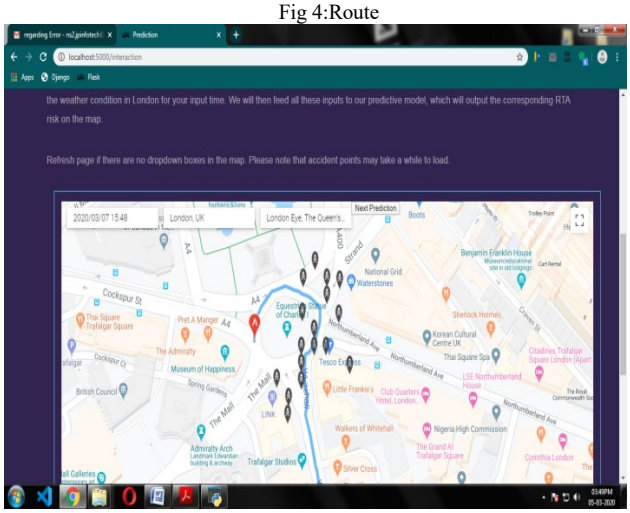

Fig 5: Prediction

VIII. CONCLUSION

In this project, we predicted probabilities of RTA for 32 boroughs of London for 48 hours in advance. We had successfully created an interactive web application that integrates Random Forest model (for prediction), Google API (for route suggestions) and Darksky API (for weather forecast).

IX. FUTURE WORK

This analytical work can provide a way to study the severity of road accidents and consequently the factors that lead to them. It was entirely decided that features such as lighting conditions could be improved to create safer roads that could result in lower levels of road accidents. Providing information containing such type of external size of information such as 3 categories of accident and weight loss conditions as well as police-related information at the scene, may be extensively analyzed to provide useful information and contribute to road safety. Although the incidence of related qualifications is uncontrollable, an analysis of this information will do change the govt. and its voters to require preventive steps towards keeping themselves safer.

REFERENCES

[1] K. Young, M. Pistner, J. O'Mara, and J. Buchanan.Cyber-disorders: The mental health concern for the newmillennium. Cyberpsychol.Behav., 1999.

[2] J. Block. Issues of DSM-V: internet addiction. AmericanJournal of Psychiatry, 2008.
[3] K. Young. Internet addiction: the emergence of a newclinical disorder, Cyberpsychol. Behav., 1998.
[4] I.-H. Lin, C.-H.Ko, Y.-P. Chang, T.-L.Liu, P.-W.Wang, H.-C.Lin, M.-F.Huang, Y.-C.Yeh, W.-J.Chou,and C.-F. Yen.The association between suicidality andInternet addiction and activities in Taiwaneseadolescents.Compr.Psychiat., 2014.
[5] Y. Baek, Y. Bae, and H. Jang.Social and parasocialrelationships on social network sites and theirdifferential relationships with users' psychologicalwell-being.Cyberpsychol.Behav. Soc. Netw., 2013.