Traffic Accident Severity Rate Detection

Farha Parveen[1*], Rahil Shaik[2], Shaik Raheem[3], Sarikonda Sree Hari Raju[4]

[1,2,3]*B.Tech. Student, Department of Computer Science and Engineering, Nalla Narasimha Reddy Education Society's Group of Institutions, Hyderabad, India*

[4]*Associate Professor, Department of Computer Science and Engineering, Nalla Narasimha Reddy Education Society's Group of Institutions, Hyderabad, India*

Abstract: **Road accident is most unwanted thing to happen to a road user, though they happen quite often. The most unfortunate thing is that we don't learn from our mistakes on road. Most of the road users are quite well aware of the general rules and safety measures while using roads but it is only the laxity on part of road users, which cause accidents and crashes. Main cause of accidents and crashes are due to human errors. The unbalance of traffic incident data has a great influence on the detection effect. Therefore, a traffic incident detection method based on factor analysis and weighted random forest (FA-WRF) is designed. We have used different Machine Learning Algorithms such as Random Forest algorithm, XG Boost algorithm, Linear Regression algorithm and Support Vector Machine algorithm to predict the severity rate of the accident whether it is slight or fatal or severe.**

Keywords: **XG Boost algorithm, Random Forest, SVM, Logistic Regression.**

1. Introduction

Traffic accident prediction is a crucial and challenging issue in the domain of intelligent traffic safety management system; it is of great significance for analysing the future development trend of traffic accidents and implementing proactive prevention measures under existing road traffic conditions. There is a huge impact on the society due to traffic accidents where there is a great cost of fatalities and injuries. In recent years, there is a increase in the researches attention to determine the significantly affect the severity of the drivers injuries which is caused due to the road accidents. Road traffic injuries cause considerable economic losses to individuals, their families, and to nations as a whole. These losses arise from the cost of treatment as well as lost productivity for those killed or disabled by their injuries, and for family members who need to take time off work or school to care for the injured.

The study on road accidents says that Every year the lives of approximately 1.3 million people are cut short as a result of a road traffic crash. Between 20 and 50 million more people suffer non-fatal injuries, with many incurring a disability as a result of their injury.

2. Proposed System

In our proposed system we have used four different Machine Learning Algorithms. They are Support Vector Machine, Random Forest algorithm, XG Boost algorithm, and Logistic Regression algorithm. Here we will upload a data set which we have collected from the Kaggle website. We will build the four different models and train them each with a Machine learning algorithm. Then we will check for the dataset whether the model is able to predict it or not. At last, we will check for the algorithm which is having highest accuracy for the prediction of the accident severity rate whether it is slight, severe or fatal.

3. Literature Review

Literature review is an important task in any research. Many of the researches have done many great works to ensure the safety at road traffic accidents and also at accident severity rate detection.

H. J. Payne and S. C. Tignor- while developing and evaluating incident detection algorithms, they have discovered that certain type of data appears repeatedly and have also developed the understandings and reasons for such repeated occurance of data patterns. They have developed a prediction model using Random Forest algorithm for analysing the relationships between the parameters and the road accidents.

A. R. Cook, He investigated the responses of using freeway traffic flow by CCTV cameras and control systems for the detection of accidents. He has conducted research on a street in Loas Angels. He had used nearly nineteen detection algorithms in his research process. The best algorithm he used had detected 42 percent of the incidents. His research has told us that the typical occurrences on urban freeways can be detected by the flow perturbations.

S. G. Ritchie and R. L. Cheu, they have presented a wavelet based novel freeway automated incident detection algorithm considering the level of traffic flow. Unlike conventional incident detection algorithms, which apply fixed threshold values and often result in undesirably high false alarm rates, their proposed algorithm varies its threshold values adaptively based on the level of tricolored.

A. Algorithms Used

1) Random Forest Algorithm

A random forest is a machine learning technique that's used to solve regression and classification problems. A random forest algorithm consists of many decision trees. A Random

---
*Corresponding author: farhaparveen606@gmail.com

Forest algorithm is more accurate than the decision tree algorithm. It provides an effective way of handling the missing data. It reduces the overfitting of datasets and increases precision. In every random forest tree, a subset of features is selected randomly at the node's splitting point. It can produce a reasonable prediction without hyper parameter tuning.

2) Logistic Regression Algorithm

Logistic regression is a supervised learning algorithm mainly used for classification purpose. It is a kind of statistical algorithm, which analyse the relationship between a set of independent variables and the dependent binary variables. It is a powerful tool for decision-making. The difference between linear regression and logistic regression is that linear regression output is the continuous value that can be anything while logistic regression predicts the probability that an instance belongs to a given class or not.

3) XG Boost Algorithm

XG Boost stands for "Extreme Gradient Boosting" and it has become one of the most popular and widely used machine learning algorithms due to its ability to handle large datasets and its ability to achieve state-of-the-art performance in many machine learning tasks such as classification and regression. One of the key features of this algorithm is its efficient handling of missing values, which allows it to handle real world data with missing values without any significant pre-processing techniques.

4) SVM Algorithm

Support Vector Machine is one of the most popular supervised learning algorithms. This algorithm works best with both regression and classification. The main purpose of SVM is to create the best line or decision boundary that can segregate N-dimensional space into classes. This algorithm chooses the extreme points or vectors that help in creating a hyperplane. Hyperplane is the best decision boundary that helps in classifying the data points.
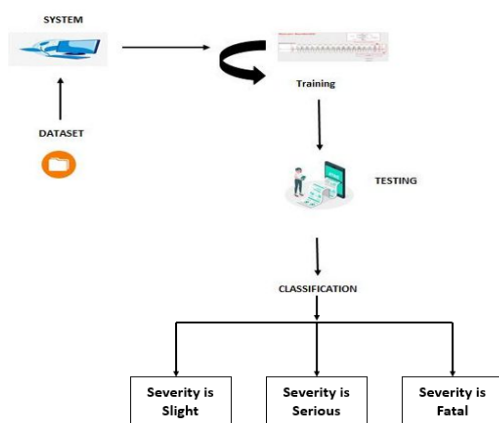

Fig. 1. Architecture

Modules:

There are two main modules in our project User and the System.

The sub modules in User module are viewing home page, upload page, results, scores and input models.

Whereas the sub modules of the System module include

working on accident severity dataset, pre-processing, training the data, model building, generating scores, generating results and graphs.
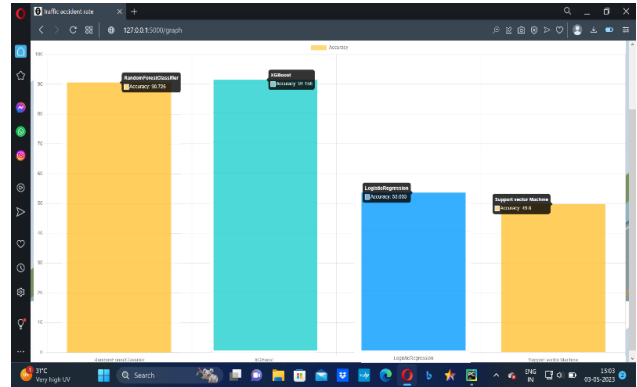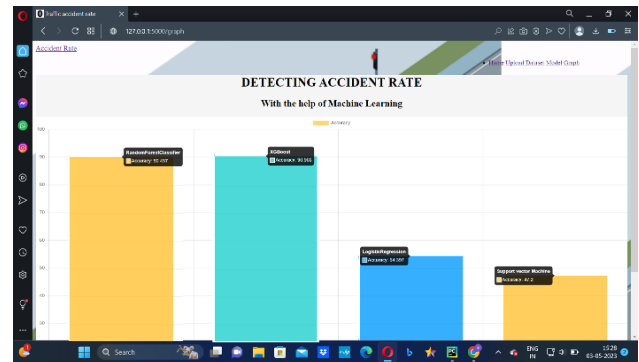
4. Results
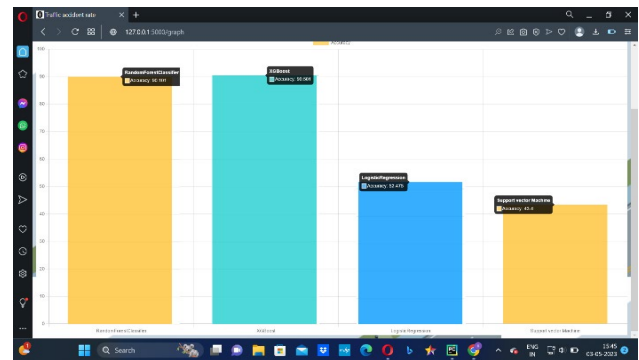

Fig. 2.

Fig. 3. Detecting accident rate

Fig. 4.

5. Conclusion

In this project we have used different Machine Learning algorithms such as Random Forest Algorithm, XG Boost algorithm, Logistic Regression algorithm and Support Vector Machine for predicting the severity rate detection of the accident. Here we will train the models by using four different machine learning algorithms and check for the model of the algorithm which gives the highest accuracy and proceed for prediction of the accident severity rate detection.

References

[1] World Health Organization (WHO), *A Road Safety Technical Package*, World Health Organization, Geneva, Switzerland, 2017.

[2] G. Biau, E. Scornet, and J. Welbl, "Neural random forests," *Sankhya A*, vol. 81, no. 2, pp. 347–386, 2019.

[3] M. Levin and G. M. Krause, ''Incident detection: A Bayesian approach,'' Nat. Res. Council, Washington, DC, USA, Transp. Res. Rec. 682, 1978, pp. 52–58.

[4] Abegaz T., Gebremedhin S., (2019), Magnitude of road traffic accident-related injuries and fatalities in Ethiopia.

[5] Yassin, S.S., Pooja Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach.

[6] Al-Masaeid, H. (2009). Traffic Accidents in Jordan. *Jordan J. Civil Eng*. 3 (4), 331–343.