IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

TRAFFIC ACCIDENT RISK PREDICTION USING MACHINE LEARNING

Ms.Ganga D Benal[1], Darshan G B ,Jayapal reddy S [3], Manoj G V and Nandan K

[,1]Assistant Professor, Department of Computer Science and Engineering, Cambridge Institute of Technology (CITech), Bengaluru, India

[2 ,3,4 5]Student, Department of Computer Science and Engineering, CITech, Bengaluru, India

Abstract: Traffic accidents pose significant threats to public safety and infrastructure. Predicting accident risk is crucial for implementing preventive measures and enhancing road safety. This study proposes a machine learning approach to predict traffic accident risk.The methodology involves the collection of extensive historical accident data, including factors such as weather conditions, road types, time of day, and traffic volume. Various machine learning algorithms are employed, including decision trees, random forests, and neural networks, to analyze and model the complex relationships between these factors and accident occurrence. Feature engineering techniques are applied to extract meaningful patterns and improve model performance. The performance of the models is evaluated using metrics.

Keywords:Traffic Accident Prediction Using Machine Learning.

INTRODUCTION

The project aims to create a system that uses Machine Learning algorithms to forecast the risk of traffic accidents. Traffic accidents area major public safety problem, with millions of incidents happening each year throughout the world, resulting in thousands of fatalities and injuries. Existing traffic accident prediction systems frequently depend on statistical models and heuristic-based techniques that may be incapable of capturing the underlying data's complexity. Machine Learning algorithms, on the other hand, provide a more data- driven approach that can analyse vast datasets and uncover patterns and correlations that are not immediately obvious.reducing accidents and ultimately saving lives. The suggested system would analyse historical traffic accident data using Random Forest Classifier algorithm, which will include elements such as location, time, weather conditions, road conditions, and other important aspects. The algorithm will be optimised to deliver reliable forecasts of traffic accident risk, and it will feature a user-friendly interface that allow users to enter pertinent information and obtain a risk score indicating the chance of a traffic accident occurring. The technology has the ability to improve road safety and minimise the frequency of traffic accidents by giving useful insights on the risk variables associated with traffic accidents to drivers, transportation authoriies, and other stakeholders.

EXISTING SYSTEM

The current system study analysed andstudied several approaches for applying it andfound that traffic accident risk projectionsgenerated from previously known to drivers elements such as personal descriptors, vehicle descriptors, and location made a lot ofintuitive sense. Once individuals are aware ofsuch high risk variables, they have a certain degree of power to lessen the danger. Because drivers are the ones in charge on the road, having this information has help them make better decisions about their trips, reducing the likelihood of traffic accidents and saving lives. The current system in use The K-means clustering technique is used. The present method analyses accident geolocation data using clustering algorithms such as K-means to categorise them into high risk hotspots in a specific region. Once gathered, clusters can be subjected to a

classification algorithm to discover their characteristics are responsible for raising the risk. These factors might include Sensitivity to initialization: K-means clustering is extremely sensitive to centroids' initial location. If the starting centroids are not correctly positioned, the method may converge to a suboptimal solution does

not adequately represent the underlying data distribution K-means clustering has a limited application because it is only applicable to datasets with a spherical or circular form. It might not be appropriate for datasets with uneven forms or clusters with changing densities.

PROPOSED SYSTEM

The suggested system is a Traffic Accident Risk Prediction system that predicts the risk of traffic accidents using a Random Forest Classifier algorithm. The system tries to increase the accuracy of traffic accident risk prediction compared to previous systems, and it also includes a geo-location component, which is significant

The suggested system would make a collection of historical traffic accident data, including geo-location, time, location, weather conditions, road conditions, and other pertinent aspects, sourced from the Kaggle repository. The data cleaned and transformed into features that capture the data's underlying patterns and trends.

METHODOLOGY

Predicting learning involves several key steps and methodologies. Firstly, data collection is essential, which includes gathering historical accident data, road and safety.

1. Input accident Data set
2. Preprocessing
3. Training
4. Randam forest classifier
5. Predected result:slight,serious,fatal

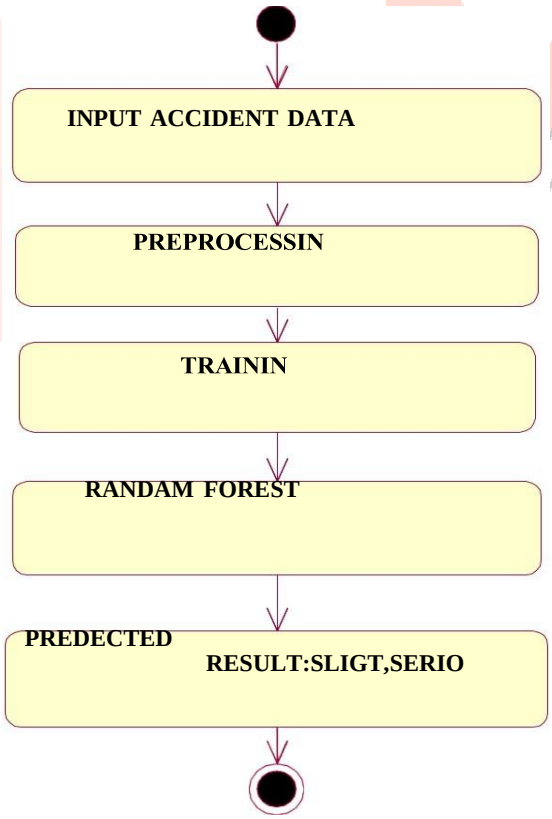

Fig 1. Activity Diagram

algorithms such as decision trees, random forests, support vector machines, or neural networks are trained on the collected data to build predictive models be utilized to perform a risk factor and minimize it by predicting the probabilities of accidents and injuriesBy contrasting two scenarios based on out-of-sample projections, it is possible to demonstrate how a statistical method based on directed graphs works. The model is used to find statistically significant factors that can be utilized to perform a risk factor and minimize it by predicting the probabilities of accidents and injuries.

Logistic Regression:

Logistic regression (LR) is a widely used supervised machine-learning technique is capable of being utilized in both classification and regression issues. The LR method uses probability to predict how categorical data will be labeled. LR's learning and prediction mechanism relies on binary classification probability measurements. The class variables in logistic regression models must be binary classified. Similar to the target column in our study dataset, which contains two separate binary values. In the dataset, the zero represents patients who have no chance of developing heart failure, whereas the one represents anticipated heart patients.

Random Forest:

Random forests, additionally referred to as random decision forests, are methods of collaborative learning that learning that generate the class-the training mode of the classes(classification) or mean prediction(regression) of the individual trees-by building a large number of decision training times and doing other tasks. The inclination of trees of choice to overfit their training set is compensated for by random decision forests.

KNN:

The controlled learning technique K-Nearest Neighbors(KNN) is utilized primarily for problems with regression and classification. KNN algorithm is non-parametric, which means it doesn't make any assumptions about the underlying data. The KNN algorithm places the new instance into a category comparable to the available classes, assuming that the new and available cases are similar. Most sample information is retrieved using the Euclidean distance metric in KNN.

RESULT

HOME PAGE

FEATURE ENHANCEMENT

A random forest is an effective feature selection indication. With the model, Scikit-learn includes an additional variable that reflects the relative value or contribute of each attribute to the prediction. During the training phase, it automatically computesthe relevance score of each feature. The relevance is then scaled down so that the total of all ratings is 1.This score will assist you in selecting the most vital characteristics and eliminating the least important ones for model development.

Random forest calculates the relevance of each feature usingi importance or mean reduction in impurity (MDI). Gini significance is often referred the complete decrease in node impurity.

CONCLUSION

Finally, the research has demonstrated considerable promise for enhancing road safety and lowering the frequency of traffic accidents. The suggestedmethod can reliably forecast the likelihood oftraffic accidents and identify the most relevant risk variables by analysing historical traffic accident data and applying powerful Machine Learning techniques as Random Forest Classifier. Future work might concentrate on increasing system performance by combining other data sources, such as traffic flow data and driver behaviour data, as well as integrating real-time data to offer up-to-date risk rankings. Furthermore, the system might be improved to deliver more personalised risk ranking based on individual driver profiles, perhaps encouraging safer driving behaviour.Overall project has shown the promise of Machine Learning in enhancing road safety and reducing traffic accidents, as well as the relevance of adopting data-driven.

REFERENCES

1. Ait-Mlouk, A., and T. Agouti (2019).A case study on a road accident using DM-MCDA, a web-based tool for data mining and multiple criteria decision analysis.
SoftwareX, vol. 10, no. 100323

2. S. Alkheder, M. Taamneh, and S. Taamneh, "Severity prediction of traffic accident using an artificial neural network," Journal of Forecasting, vol. 36, no. 1, pp. 100–108,2017.

3. Q. Wang and H. Chen, "Optimization of parallel random forest algorithm based on distance weight," Journal of Intelligent andFuzzy Systems, vol. 39, no. 2, pp. 1951– 1963, 2020.

4. J. Gan, L. Li, D. Zhang, Z. Yi, and Q. Xiang, "An alternative method for traffic accident severity prediction: using deep forests algorithm," Journal of Advanced Transportation, vol. 2020, 13 pages, 2020.

5. M. Schonlau and R. Y. Zou, "The random forest

algorithm for statistical learning," STATA Journal, vol. 20, no. 1, pp. 3–29, 2020.

6. M. Yan and Y. Shen, "Traffic accident severity prediction based on random forest," Sustainability, vol. 14, no. 3, 2022.

7. S. H. A. Hashmienejad and S. M. H. Hasheminejad, "Traffic accident severity prediction using a novel multi-objective genetic algorithm," International Journal ofCrashworthiness, vol. 22, no. 4, pp. 425– 440, 2017.

8. S. Alkheder, M. Taamneh, and S. Taamneh, "Severity prediction of traffic accident using an artificial neural network," Journal of Forecasting, vol. 36, no. 1, pp. 100–108, 2017.

9. T. Lu, Z. H. U. Dunyao, Y. Lixin, and Z. Pan, "The traffic accident hotspot prediction: based on the logistic regression method," in Proceedings of the 2015 International Conference on Transportation Information and Safety (ICTIS), IEEE, Wuhan, China, June 2015.

10. Suganya, E. and S. Vijayarani. "Analysis of road accidents in India using data mining classification algorithms." 978-1-5386- 4031- 9/17/ IEEE (2017).