ISSN No:-2456-2165

Road Accident Prediction and Analysis Using Machine Learning

Ms.Preshita Bhortake, Vivek Barhate
Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering,
Savitribai Phule PuneUniversity of Pune, Pune, India

Abstract:- India has a substantial population, which contributes to a significant daily automobile commuter population. This leads to several accidents occurring every day. These mishaps frequently result in severe financial hardship for families as well as the possibility of fatalities. The goal of this article is to identify accident-prone areas and alert regular commuters to the incidents that are occurring there. Accidents can occur at any time and without warning, but as users of this interface, we can be more cautious in locations where accidents occur frequently. The user interface will alert a user to the high-medium accident risk areas.

Keywords:- *Random Forest Algorithm,GaussianNaïve Bayes algorithm, Logistic regression, Machine Learning.*

I. INTRODUCTION

According to the study conducted by WHO, millions of people die as a result of traffic accidents every year. Accidents are the primary problems facing the world today since they frequently result in numerous injuries, fatalities, and financial losses. Road accidents are a problem that has impacted the general public's well-being and the economy of the country.A fundamental task for road transportation systems is to develop precise models to identify the cause of accidents and provide recommendations for safe driving. This research effort creates models based on the factors that cause accidents, such as weather, causes, road characteristics, road conditions, and accident type. Likewise, select a number of significant elements from the best model in order to build a model for describing the cause of accidents. In more detail, these are algorithms like Random Forest Algorithm, Gaussian Naïve Bayes algorithm, Logistic Regression (LR) implemented to discover the impact of various factors and gives a safe driving protocol.

II. ABBREVIATIONS

- KNN- K-Nearest Neighbor
- LR-Logistic Regression
- 3.GNB-Gaussian Naïve Bayes algorithm
- 4.VR-Virtual Machine

III. METHODOLOGY

- **The Virtual Machine (VR):** It has the trained and tested Machine learning algorithm implemented. The frontend and backend server are deployed on it.
- **Front-End:** Users input for the prediction factors are taken and sent to the backend server.
- **Back-End:** The model is deployed here and the input data is fed into the Machine Learning model.
- **Machine Learning Model:** We have used decision tree, random forest and logistic regression and also applied hyperparameter tuning to increase its efficiency.

➢ Modules Used:

- **Data Collection:** Collection of data from repository dataset is the initial step for the commencement of the research work in which data is collected from various previously used dataset andpatient history.
- **Attribute Selection:** Attributes are properties of the dataset which are used for determining factorslike driver behaviour, road condition etc. It is a cumbersome process to select the attribute as itsselection affects the accuracy and continuousness of the training model.
- **Data Pre-processing:** Transforming data into required format and adjusting the missing dataset values, attributes scaling etc. It adds or alter the dataset
- **Key Management:** This module is not only responsible for the generation and maintenance of the key, but also uses the hierarchical key management approach to protect the key.
- **Balancing of Data:** Imbalanced dataset are balanced
  1. Under Sampling- Reduction of dataset; When ample amount of dataset is there.
  2. Over Sampling- Increase in dataset; When Dataset isless.

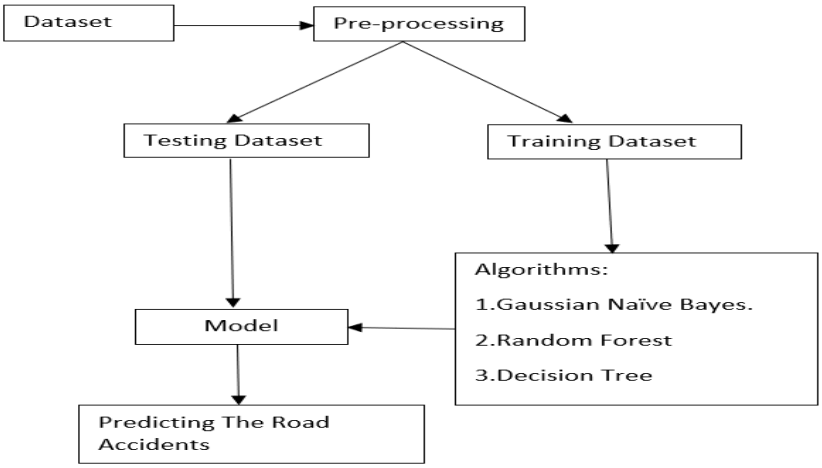

Fig 1: Flowchart Diagram

IV. ALGORITHMS

A. *Random Forest algorithm*

It is a supervised learning technique that is utilized for both classification and regression.However, classification issues are its primary usage. Random forest basically builds decision trees based on data samples, obtains predictions from each one, and then votes to determine which is the best answer. Because it averages the results, the proposed technique is superior to a decision tree classifier excessively.

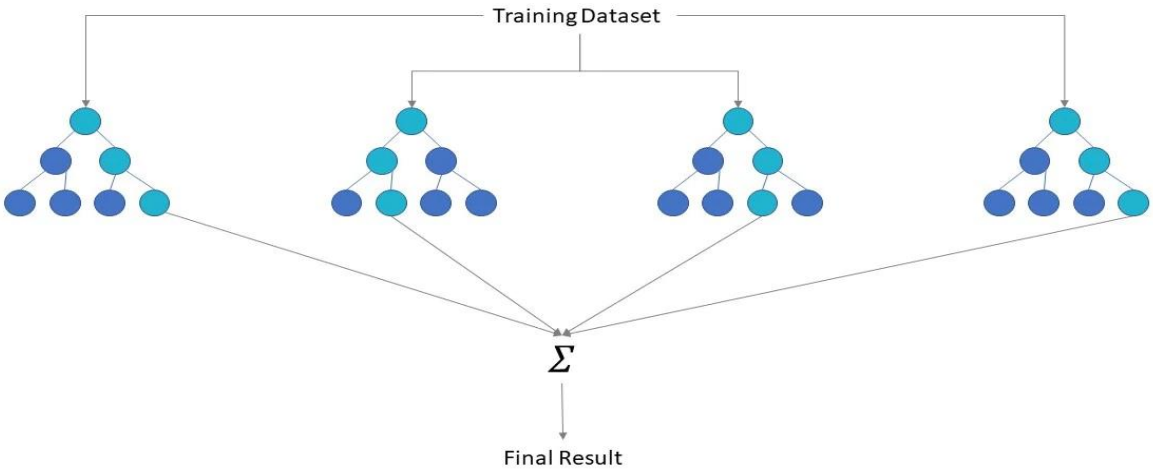

Fig 2: Random Forest

B. *Decision Trees*

In decision analysis, a decision tree can be used to represent decisions and decision making visually and explicitly. As the name goes, it uses a tree-like model of decisions. Though a commonly used tool in data mining for deriving a strategy to reach a particular goal, it's also widely used inmachine learning.

C. *Gaussian Naïve Bayes algorithm*

It is the Naïve Bayes extension. A probabilistic machine learning approach called Naive Bayes can be applied to a variety of categorization applications. Naive Bayes is frequently used for document classification, spam filtering, prediction, and other tasks. This method takes its name from Thomas Bayes' discoveries, on which it is based.The method includes features in its model that are independent of one another, hence the term "Naïve." Any changes to the value of one feature of the algorithm have no direct effect on the value of any other feature. The NaïveBayes algorithm's key benefit is that it is an easy-to-use yet effective method.

D. *Decision Tree Algorithm*

It is a subset of the supervised learning algorithm family. Unlike the other supervised learning techniques, the decision tree methods can be implemented to both classification and regression issues. The main goal of employing decision trees is to build a training model that can be applied to forecast the type or value of a target variable by learning choice rules inferred from historical data. The decision tree approach uses a tree representation to attempt to address the issue. Each leaf node of the tree corresponds to a class label, whereas each internal node of the tree relates to an attribute.

V. CONCLUSION

The training of a model that can detect multiple parameters to predict and assess road accidents can be done using machine learning techniques. It defined the requirements for each technique and developed a comprehensive framework to identify and stop road accidents. It is evident that a variety of elements, including vehicle kinds, vehicle ages, weather conditions, road conditions, and others, have a significant impact on incidents of traffic accidents. Creating an application that accurately predicts traffic accidents based on the aforementioned parameters is what future need is.

ACKNOWLEDGEMENT

I take this to express my deep sense of gratitude towards my esteemed guide Prof. J. A.Wagh for giving me this splendid opportunity to select and present this seminar and also providingfacilities for successful completion.I thank Dr. K. S. Thakre, Head, Department of Computer Engineering, for opening the doors of the department towards the realization of the seminar, all the staff members, for their indispensable support, priceless suggestion and for most valuable time lent as and when required. With respect and gratitude, I would like to thank all the people, who have helped me directly or indirectly.

REFERENCES

[1.] M. F. Labib, A. S. Rifat, M. M. Hossain, A. K. Das and F. Nawrine, "Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh," *2019 7th International Conference on Smart Computing & Communications (ICSCC)*, 2019, pp. 1-5, doi: 10.1109/ICSCC.2019.8843640

[2.] J. Patil, V. Patil, D. Walavalkar and V. B. Lobo, "Road Accident Analysis and Hotspot Prediction using Clustering," 2021 6th International Conference on Communication and Electronics Systems (ICCES), 2021, pp. 763-768, doi: 10.1109/ICCES51350.2021.9489074. Ding Chen, Qiseng Yan, Chunwang Wu, Jun Zhao "SQL Injection Attack Detection and Prevention Techniques Using Deep Learning", School of Cyberspace Security, Chengdu University of Information Technology, Chengdu, 2020.

[3.] Priyanka A. Nandurge, Nagaraj V. Dharwadkar,Analyzing Road Accident Data using Machine LearningParadigms, International conference on I-SMAC (IoT inSocial, Mobile, Analytics and Cloud), 2017, pgno.604-610, IEEE.

[4.] S. Ahmed, M. A. Hossain, M. M. I. Bhuiyan and S. K. Ray," A Comparative Study of Machine Learning Algorithms to Predict Road Accident Severity,", Proc. IEEE Int. Conf, pp.1464-1473, Dec. 2020.

[5.] Dr. Shivashankara S, Archana B J, Vijayalakshmi D, Meghana Gowda D, Anusha "Road Accident Analysis and Prediction Using Data Mining.