IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

Road Accident Severity Prediction using Machine Learning

R. Arunachalam[1] , S. Peararulselvi[2] , M. Saraswathi[3] , M. Subasri[4]
Assistant Professor , Department Of Computer Science and Engineering[1]
Students, Department Of Computer Science and Engineering[2,3,4]
Anjalai Ammal Mahalingam Engineering College , Thiruvarur , Tamil Nadu , India
r.arunachala@gmail.com[1] , spearuarul@gmail.com[2] , snegasara16@gmail.com[3], msubasri.mng@gmail.com[4]

Abstract: *Road accidents have become a major concern globally, causing a significant number of fatalities and injuries every year. The aim of this project is to predict road accidents severity using machine learning techniques, in order to reduce their occurrence and mitigate the associated risks. The project uses data collected from various sources such as accident reports, weather conditions, and road infrastructure to train and evaluate various supervised learning algorithms and predict the accident severity. Four algorithms were compared, including Decision Tree, Naive Bayes, Random forest . Most probably occurring road accident locations are identified and that particular region is indicated as black pot. The proposed method can be used to provide real-time risk information to road users, helping them to make informed decisions and avoid potential accidents. The project highlights the importance of using machine learning techniques in road safety analysis, providing a foundation for further research in this field.*

Keywords: Road accident severity prediction, Machine learning Algorithm

I. INTRODUCTION

Public safety worldwide is still very concerned about road accidents. According to estimates, 1.35 million people died in traffic-related incidents worldwide in 2020, and if adequate action isn't taken, this number is predicted to rise. Being able to detect and avoid accidents before they happen is one of the biggest obstacles to minimizing traffic accidents. In this sense, machine learning has shown promise as a tool for predicting the seriousness of traffic accidents, which can assist in giving drivers early warnings and information to help them avoid potential hazards on the road.

Large-scale databases on traffic accidents, together with improvements in machine learning methods and computing capacity, have made it possible in recent years to create sophisticated models for determining the severity of accidents.

In this conference paper, we present a machine learning-based approach for road accident severity prediction. We will discuss the various machine learning algorithms that we have evaluated for predicting the severity of road accidents, along with the factors that we have considered for our prediction model. We will also describe the dataset we have used for training and evaluating our models, along with the preprocessing steps we have taken to prepare the data. Finally, we will present the results of our experiments and discuss the potential applications of our approach for reducing the impact of road accidents on society.

II. BACKGROUND

Several studies have been conducted on predicting the severity of road accidents using machine learning. These studies have used various machine learning algorithms such as decision trees, random forests, and logistic regression. The predictors used in these studies include weather conditions, road characteristics, vehicle characteristics, and driver behavior.

III. RELATED WORKS

There have been several related works in the area of road accident severity prediction using machine learning. Some of the notable studies are:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

[1] In order to forecast the severity of road accidents, this study proposes an ensemble of deep learning and machine learning models dubbed RFCNN that combines Random Forest and Convolutional Neural Network. A number of basic learner classifiers are used to compare the performance of the proposed technique. Accident data from the USA from February 2016 to June 2020 was used in the analysis. Deep learning and machine learning methods had previously been combined. Used for predicting traffic accidents, like Random forest RF-CNN model, and CNN(RF-CNN), anticipate the accident's severity, Accuracy=0.812.

[2] The main objective of this research paper is to examine traffic accidents and evaluate their severity using advanced deep learning techniques (TASP-CNN model for traffic accident's severity prediction is proposed that takes into consideration combination relationships among traffic accident's features).TASP - CNN's traffic accident model. It is tested and trained to predict severity. The accident severity is predicted using a model.

[3] The authors of this study analyze traffic accidents using four of the most sophisticated and widely used supervised learning algorithms in machine learning. Decision Tree, K-Nearest Neighbors (KNN), Naive Bayes, and Adaptive Boosting (AdaBoost) are those methods. The Gaussian and random forest together distributional approach to accident prediction. Map with an error place noted. Users are sent notification information that through the map, specify in the form of a black spot.

[4] In order to determine the factors that influence an accident's severity, data mining and machine learning approaches are used in the analysis of road accident statistics. Displaying an image of examination of the numerous influences the seriousness of an incident. The Association rule mining algorithm is the most well-known and frequently applied one. This can be applied to find the important connections between the data in the big database. a priori, prognostic The two most popular association rule mining techniques are apriori and FP-growth algorithms. The outcomes of these data mining techniques can aid in understanding the most important variables or recurring trends.

[5] The authors have examined data on traffic accidents and made predictions about upcoming collisions using machine learning techniques. The data collection and analysis methods, the prediction algorithms, and the outcomes are all covered in the paper. The authors want to make driving safer and have fewer accidents by using a predictive analytics methodology. The notion of machine learning is used to forecast accident severity and examine the causes of accidents and methods for predicting them, such as Random Forest and Linear Regression.

[6] The main objective of this research paper is to analyze traffic accidents and assess accident severity using cutting-edge machine learning techniques. This research study also does traffic accident analysis utilizing the most widely used supervised machine learning techniques. Visual representation of learning examination of the numerous influences the seriousness of the collision and a traffic accident analysis.

IV. PROPOSED MODEL

In this section, we discuss our proposed model.

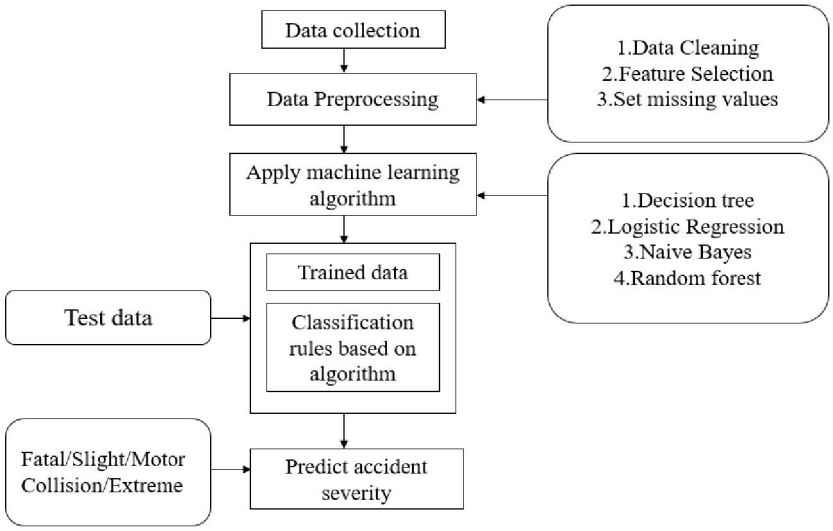

Figure.1 Proposed Model

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

4.1 Data Collection

Three CSV files—Accidents.csv, Vehicle.csv, and Casualties.csv—make up the dataset. The datset is a collection of information about UK traffic incidents that spans the years 2005 to 2015. The data set includes a variety of characteristics about the accident, such as the date, time, and location of it as well as information about the cars that were involved, the severity of the accident, and the number of people who were injured. Information about the weather and road conditions at the time of the accident is also included.

4.2 Data Preprocessing

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.Data preprocessing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model. It involves below steps

- Data Cleaning: The collected data will likely have missing values, inconsistencies and outliers. The data cleaning module is responsible for cleaning and preprocessing the data so that it can be used for training the machine learning model.
- Data Integration: In some cases, data may be stored in different sources and formats, so integrating data from multiple sources is a necessary step in preprocessing.

Normalize the data: Normalize or standardize numerical variables.

4.3. Exploratory Data Analysis

Exploratory data analysis is the graphical representation of information and data. Using visual components like charts, graphs, and maps, data visualization tools make it simple to explore and grasp trends, outliers, and patterns in data.

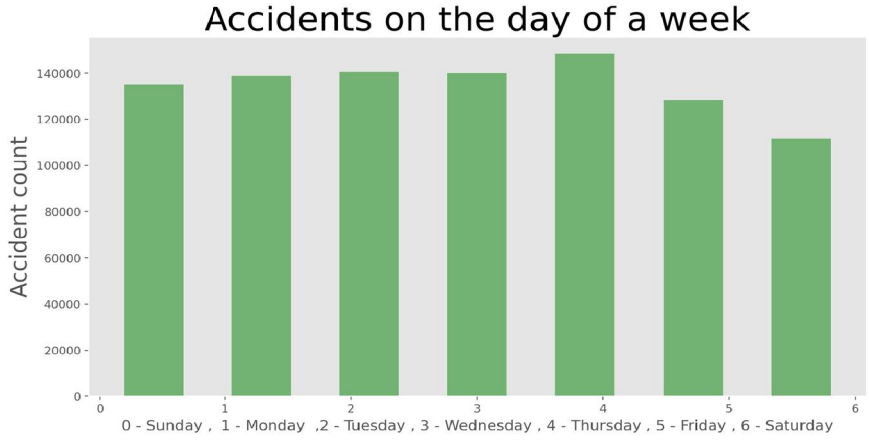

Figure.2 Accidents on the day of the week

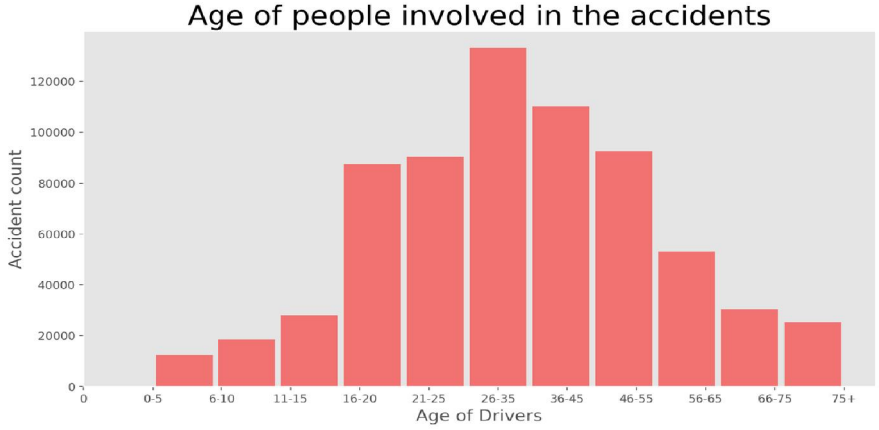

Figure.3 Age of people involved in the accidents.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

4.4 Model Training and Evaluation

The machine learning model is trained on the pre-processed data to learn the relationship between the features and the target variable (i.e., whether an accident is likely to occur or not). A variety of machine learning algorithms can be used for this purpose, such as decision trees, random forests, logistic regression and naïve bayes. Split the preprocessed data into training and testing datasets. Choose a suitable machine learning algorithm based on the problem statement and dataset characteristics. Train the model on the training dataset using the chosen algorithm. The trained model is evaluated using various metrics such as accuracy, precision, recall to determine its performance. Save the trained model for deployment in the prediction module

4.5 Deployment

The model is deploy using flask app and develop the ML powered web app. The model is deployed in the back-end. The input data from the front-end is fed into the Machine Learning model. The input data are User age, gender, vehicle type, vehicle age and engine capacity. The model runs and predicts the severity. The output is sent back to the front-end and displayed to the user.

V. THE METHODOLOGY OF PROPOSED MODEL

5.1 Decision Tree

Decision tree is a popular supervised machine learning algorithm that is widely used in both classification and regression problems. It is a tree-structured model where each internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label or a numerical value. The algorithm learns a set of rules to make decisions based on the features of the input data.

The decision tree algorithm works by recursively splitting the data into subsets based on the values of the features. At each split, the algorithm selects the feature that provides the most information gain or the best split. Information gain is a measure of the reduction in entropy or impurity achieved by splitting the data based on a feature. The goal is to select the features that provide the most discrimination between the classes or the best separation between the regression targets.

5.2 Random Forest

Random forest is a machine learning algorithm that belongs to the family of ensemble methods. It is used for both classification and regression tasks and combines multiple decision trees to improve the accuracy and stability of the predictions.

The algorithm works by building a set of decision trees, where each tree is trained on a random subset of the training data and a random subset of the features. This helps to reduce the variance of the model and improve its generalization performance. To make a prediction on a new data point, each tree in the forest independently predicts the output, and the final prediction is obtained by taking the majority vote for classification or the average for regression.

5.3 Logistic Regression

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.Logistic regression is used for classification tasks where the goal is to predict the probability that an instance of belonging to a given class. It is used for classification algorithms its name is logistic regression. It's referred to as regression because it takes the output of the linear regression function as input and uses a sigmoid function to estimate the probability for the given class.

5.4 Naive Bayes

Naive Bayes refers is a probabilistic machine learning algorithm based on the Bayesian probability model and is used to address classification problems. The fundamental assumption of the algorithm is that features under consideration are independent of each other and a change in the value of one does not impact the value of the other.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

VI. EXPERIMENTAL RESULT

The Accuracy of machine learning models of road accident severity prediction between Random Forest, Decision Tree, Naive Bayes and Logistic Regression is being visualized in Figure 4.

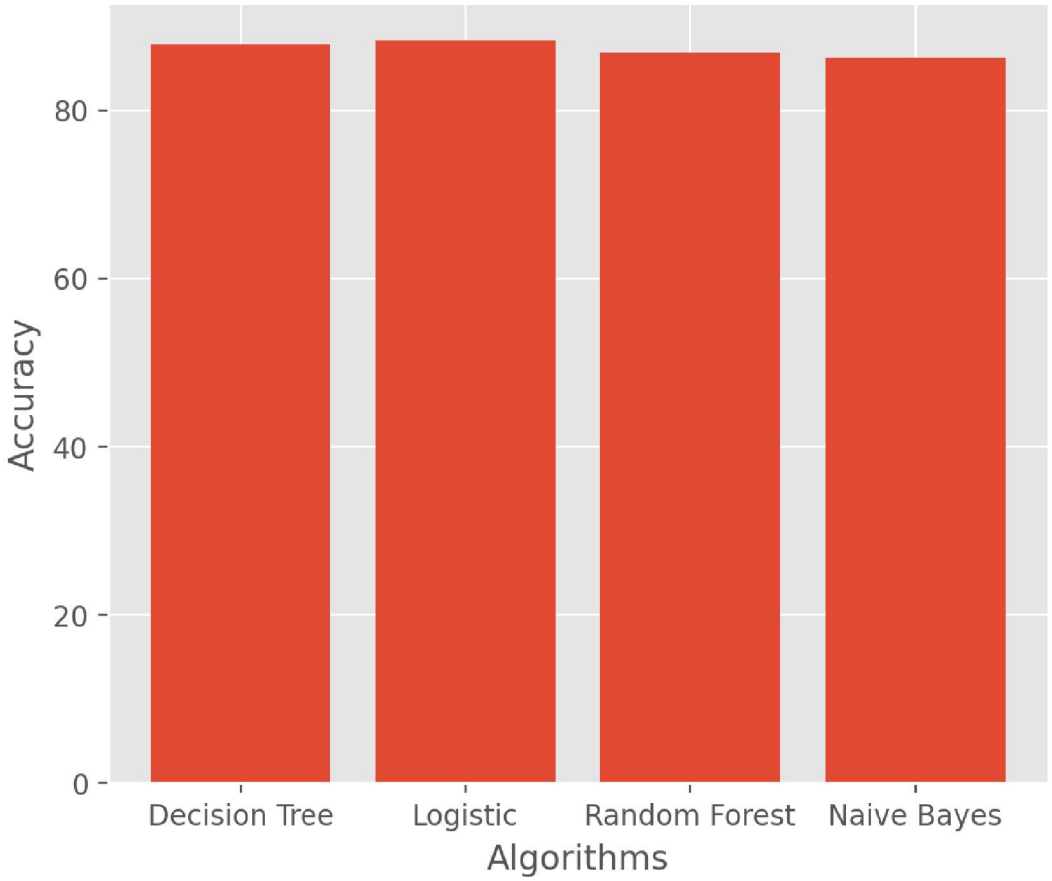

Figure.4 Algorithm accuracy

VII. CONCLUSION

This project aims at using Machine Learning techniques to predict severity of an accident at any particular location. Machine Learning has enabled us to analyze meaningful data to provide solutions with a greater accuracy than with humans. In conclusion, road accident severity prediction using machine learning is a promising approach to reducing the number of accidents and the impact they have on society. By analyzing various features related to road conditions, weather conditions, driver behavior, and vehicle type, machine learning models can accurately predict the severity of accidents and provide early warnings to drivers. This can help drivers avoid potential dangers on the road, reduce the likelihood of accidents, and ultimately save lives. However, the implementation of such models requires careful consideration of the data, feature selection, model selection, and evaluation. Future research in this area should focus on improving the accuracy of the models and integrating them into existing traffic management systems. Overall, road accident severity prediction using machine learning has the potential to significantly improve road safety and should be considered as an important tool in traffic management.

REFERENCES

[1]. "Traffic Accident Severity Prediction Based on Decision Level Fusion of Machine and Deep Learning Model"-Mubariz Manzoor1 , Muhammad Umer, Saima Sadiq, Abid Ishaq, Saleem ullah, Hamza Ahmad Madni, and Carmen Bisogni(2021),IEEE ACCESS Journal.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

[2]. "A Deep-Learning Approach-Based CNN Network "-Ming Zheng, Tong Li, Rui Zhu, Jing chen, Zifei Ma, Mingjing Tang, Zhongqiang Cui and Zhan Wang(2019),IEEE ACCESS Journal.

[3]. "Prediction of Road Accidents Using Machine Learning Technique"-Saravana Alagarsamy,M. Malathi, M.Manonmani(2021), Fifth International Conference on Electronics, Communication and Aerospace Technology (ICECA).

[4]. "A framework for analysis of road accidents"- Shristi Sonal,Saumya Suman(2018), International Conference on Emerging Trends and Innovations in Engineering and Technological Research (ICETIETR).

[5]. "Predictive Analytics of Road Accidents Using Machine Learning"-C.Kalairaja,D.Chitra Devi,Anju Rajan(2022), second International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE).

[6]. Road Accident Analysis and Prediction of Accident Severity By Using Machine Learning in Bangladesh"-Md. Farhan Labib, Ahmed Sady Rifat, Md. Mosabbir Hossain, Amit Kumar Das, Faria Nawrine(2019), Seventh International Conference on Smart Computing and Communications(ICSCC).