e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-13, Issue-3 (June 2023)
https://doi.org/10.31033/ijemr.13.3.28

Accident Severity Detection Using Machine Learning

Nagma Bi[1] and Dr. Halima Sadia[2]
[1]Student, Department of Computer Science & Engineering, Integral University, INDIA
[2]Associate Professor, Department of Computer Science & Engineering, Integral University, INDIA

[1]Corresponding Author: nagma1215@gmail.com

Received: 29-05-2023 Revised: 16-06-2023 Accepted: 30-06-2023

ABSTRACT

Road accidents are one of the most regrettable hazards in this hectic world. Each year, traffic accidents cause a large number of casualties, illnesses, and deaths in addition to suffering huge financial losses.There are numerous things that cause traffic accidents, especially those related to the environment, vehicles and the travelers.By analyzing the severity of the road accidents that happened in the past, and the factors that caused it, it is possible to take precautionary measures to reduce the road accidents rate significantly in the future.This project includes developing a machine learning model that can categorize accident severity depending on the circumstances that affected the accident. A prediction model was created using a variety of machine learning classifiers, including Gradient Boosting Classifiers (GBC), K-Nearest Neighbour (KNN), Random Forest (RF), and Decision Tree (DT). The severity of a road accident can be detected 90% accurately, according to the results of a gradient boosting algorithm.The study makes use of publicly accessible European data.The approach presented in the research is broad enough to be used with various data sets from other nations.Additionally, the web portal's model was used to create an intelligent system for predicting the severity of accidents.

***Keywords*—**Decision tree (DT), K- Nearest Neighbour (KNN), Random Forest (RF), Gradient Boosting Classifiers (GBC)

I. INTRODUCTION

A road traffic accident (RTA) is an unanticipated incident that happens on the road and involves a vehicle and/or other road users and results in casualties or property loss.The economic and social levels are greatly impacted by traffic accidents. In 2030, it's predicted that road accidents would account for a significant portion of fatalities. Many people's lives and entire communities have improved thanks to motorization, however, there are costs associated with the advantages.Road fatalities account for more than 90% of all global deaths.

Low-income nations only make up 48% of the world's registered autos. More money was lost financially—about $518 billion—than was given to these nations in development aid. Developing nations continue to lose 13% of their gross national product (GNP) due to the epidemic of traffic fatalities, despite the affluent nations' stable or dropping road traffic death rates as a result of concerted corrective efforts from many sectors. Road crashes could become the sixth leading cause of death worldwide, according to the World Health Organization (WHO).Authors evaluate several severity levels, such as light injury and severe injury.Accidents are a significant contributor to illnesses, fatalities, long-term disability, and property damage. Because it burdens the hospitals, it has an impact on both the economy and the healthcare system. Accident severity prediction, one of the key topics in accident management, is crucial to the rescuers' assessment of the seriousness of traffic accidents, their potential effects, and the implementation of effective accident management measures.The number of injuries, the number of fatalities, and the quantity of property damage are the three elements that make up an analysis of the severity of an accident. The prediction model looks into the relationships between different crash severity injury categories and factors that may have contributed to the accident, including driver behaviour, vehicle characteristics, road geometry, environmental factors, and accident causes.Therefore, the goal of this research is to examine several ML algorithms and evaluate how well they perform in terms of predicting both the frequency and severity of traffic accidents.

II. METHODOLOGY

This project classifies the road accident as slight or serious.This paper's methodology is displayed in Fig. 1.

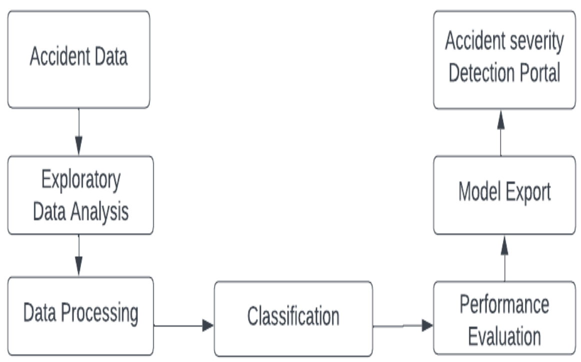

Figure 1

This project consists of five main phases. 1. Conducting descriptive study on the accident data 2. Pre-processing the data using grouping and label encoding 3. Building the machine learning classification model 4. Performance evaluation 5. Developing intelligent web portal for accident severity classification

The steps involved in the intelligence web portal are described in the figure 2.

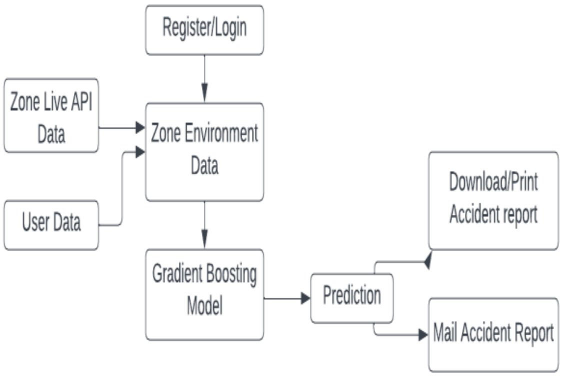

Figure 2

A. Dataset Description

This project was implemented using the Accident Severity dataset from the kaggle. The dataset contains 3057 accident samples. Each sample contains 14 predictive attributes and 1 target attribute. The dataset contains the accident data from 01-01-2009 to 31- 12-2009. There are no null values present in the dataset. The target attribute contains two values: slight accident and serious accident. The dataset contains 321 serious data and 2736 slight accident data.The Table-1 describes the predictive attributes that are available in the accident data and its description.

B. Data Preprocessing

A data mining approach known as data preprocessing is used to transform the raw data into a format that is both useful and efficient. Preprocessing alters the format of the data to make it more quickly and effectively processable by data mining, machine learning, and other data science applications.Usually used at the very beginning of the machine learning and AI development pipeline, the techniques provide accurate results..In the accident data, three data preparation activities were carried out.

1. Data compression
2. Adding a code for the categorical data
3. Eliminating the unnecessary column

C. Build Machine Learning Model

Four machine learning classifiers such as Decision tree learning, K-Nearest neighbor,Random forest classifier and the Gradient boosting classifiers are used to develop machine learning models. These models are imported using the Sci-kit learn machine learning library.

Variables	Description	Data Type	Scale	Null Value
Referene No.	Accident Identity no.	Integer	Serial No.	No
Easting	Easting point	Integer	Map Point	No
Northing	Northing point	Integer	Map Point	No
No. of Vehicles	No. of Vehicle in the spot	Integer	Vehicle Count	No
Accident Date	Date of Accident happened	Date	Date	No
Time(24 hr)	Time of Accident happened	Time	Time	No
1st Road class	Road Type	Varchar	Category	No
Road Surface	Surface type of the road	Varchar	Category	No
Lighting Conditions	Lighting condition in the spot	Varchar	Category	No
Weather Conditions	Weather condition in the spot	Varchar	Category	No
Casualty Class	Casuality type(Driver..,etc)	Varchar	Category	No
Sex of casualty	Sex of the Casualty	Varchar	Category	No
Age of casualty	Age of the Casualty	Integer	Age in Years	No
Type of vehicle	Type of the Vehicle	Varchar	Category	No

Table 1

III. RELATED WORK

Rabia Emhamed Al Mamlook[1] to selected a group of influencing elements and create a model for categorizing the inflexibility of injuries, established models are used. According to the study's findings, the RF model holds promise for predicting the injury rigidity of commercial accidents. Keneth Morgan Kwayu[2] The purpose of this investigation is to evaluate and compare various techniques to modelling collision inflexibility as well as investigate the impact of threat elements on the casualty issues of traffic accidents using machine learning-based driving simulation.The suggested model can be used to study highway crashes. Buket Geyik[3] aimed to establish models to prognosticate the accident inflexibility situations fatal, serious, and slight of business accident injury records for possible accidents by using some data mining bracket styles. Hemanth Kumar[4] The goal of this inquiry is to assess and contrast several methods for simulating collision rigidity as well as examine the effect of danger factors on the casualty issues associated with automobile accidents.The proposed model can be used to research traffic collisions. Bulbula Kumeda[5] The dataset effectively classified using the fuzzy-FARCHD algorithm, which obtains 85.94% delicacy. In this study, he made it clear that the factors to consider when choosing the qualities are Lighting Conditions, First Road Class, and Number of Vehicles.Md. Farhan Labib[6] Sort the inflexibility of accidents into the four categories of fatal, grievous, simple injury, and motor collision. AdaBoost eventually succeeds in achieving the desired performance. Yongbeom Lee[7] proposed a model for valuing the rigidity of business accident injuries relevant to the business accident automatic announcement system. He implemented the prophetic model using a decision tree.

Nejdet Dogru[8] provides a system for intelligent business accident detection whereby moving objects exchange bitsy vehicle variables. The proposed system delivers business warnings to drivers and also makes use of dissembled data from vehicular ad-hoc networks (VANETs) based on the pets and equals of the vehicles. It is demonstrated that if location and haste values for each car are provided, the geste of the vehicles can be dissected, and accidents can be quickly discovered.

Joy Paul[9] created a more effective model to prevent traffic accidents, he presented a multiclass model that incorporated both the vaticination of accidents and their related inflexibility. MUBARIZ MANZOOR[10] Road accident inflexibility was predicted using an ensemble of deep learning and machine learning models dubbed RFCNN, which combined Random Forest and Convolutional Neural Network. Eakapan Boonserm[11] uses Thailand's free government data to address the unbalanced binary classification problem of forecasting injury severity of traffic accidents during Thai new year celebrations. Before fitting any Random Forests models, training instances were rebalanced using random undersampling, oversampling with SMOTE, and the combination of both.

Jian Zhang[12] compare the predicted performance of various machine learning and statistical methods with separate modelling logic for crash severity assessments, including prediction accuracy and estimation of variable relevance.It demonstrated that the conclusions drawn about variables' importance using various methodologies were not always consistent.Hani M Alnami[13] order to created prediction models to estimate the severity of traffic accidents, this study examines actual traffic and accident data for a Florida roadway.

Sanaa Elyassami[14] The proposed study to develop machine learning-based models for analysing crash data, linking key threat indicators, and predicting the injury rigidity of drivers is accepted. In the proposed study, we used three machine learning algorithms to compare the Statewide Vehicle Crashes Dataset provided by the Maryland State Police to Decision Tree, Random Forest, and Grade Boosted Tree. The highest delicacy was reported by the grade boosted- grounded model.

ANG JI[15] Study's objective was to use ensemble machine learning models to capture the predictive potential of crash medium-related characteristics. The findings show that named models are capable of prognosticating rigidity at a high level of delicacy.

IV. RESULT ANALYSIS

The common measures like accuracy, precision, recall, and ROC curve are used to assess the machine learning models. Based on the metrics of each classifier, the comparative study was conducted over the models. The result shows that the ensembler can give the best accuracy among the traditional single machine learning classifier.

PERFORMANCE METRICS

The effectiveness of the classification is measured using the following metrics.

A. Confusion Matrix

A condensed table used to assess how well a classification model performs is known as a confusion matrix, also known as an error matrix. To calculate the total number of accurate and incorrect predictions for each class, count values are employed.The effectiveness of the classification is assessed using the metrics listed below.

True Positives (TP) are outcomes that the model correctly identified as belonging to the positive category.

True Negative (TN): A result that the model correctly anticipated would fall under the category of being negative.

False Positive (FP), also known as a type 1 error, is a result that happens when a model predicts the positive class when it ought to have predicted the negative class.

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-13, Issue-3 (June 2023)
https://ijemr.vandanapublications.com
https://doi.org/10.31033/ijemr.13.3.28

False Negative (FN), also referred to as a type 2 error,when the model correctly predicts the negative class but forecasts the positive class instead.

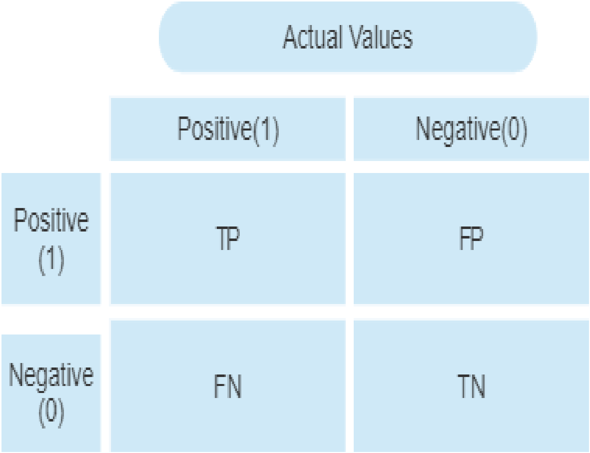

Figure 3: Confusion Matrix

B. Performance Comparison over the Classifiers

The figure 4 describes the accuracies that are gained by the considered classifiers.Gaining high accuracy is the important thing to be noticed. The Decision tree learning has got 89% of accuracy. The KNN got 89.4 % of accuracy and worked better than the KNN. Nearly the Random forest algorithm got 90% of accuracy. The best accuracy gained by the Gradient boosting model. Since it is an ensemble classifier, it employs many weak learners in order to learn the severity classification. Hence, its performance is better than the classifications that are considered.

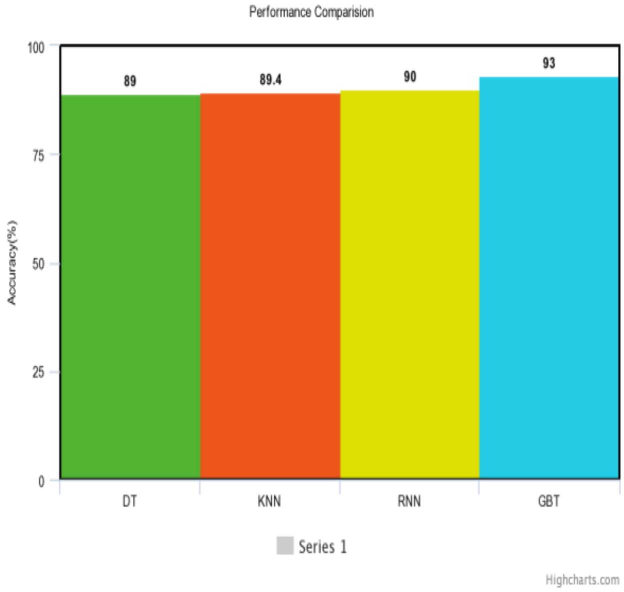

Figure 4

C. ROC Curve

The ROC curve is another performance metric used to rate the machine learning model. A graph that shows how well a classification model performs at every level of categorization is called the receiver operating characteristic curve (ROC curve). On this curve, two variables are plotted: 100% True Positive,False Positive. The fig.5 describes the ROC Curves that are achieved by the considered classifier.The gradient boosting classifier has its curve from less false positive rate.

e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-13, Issue-3 (June 2023)
https://doi.org/10.31033/ijemr.13.3.28

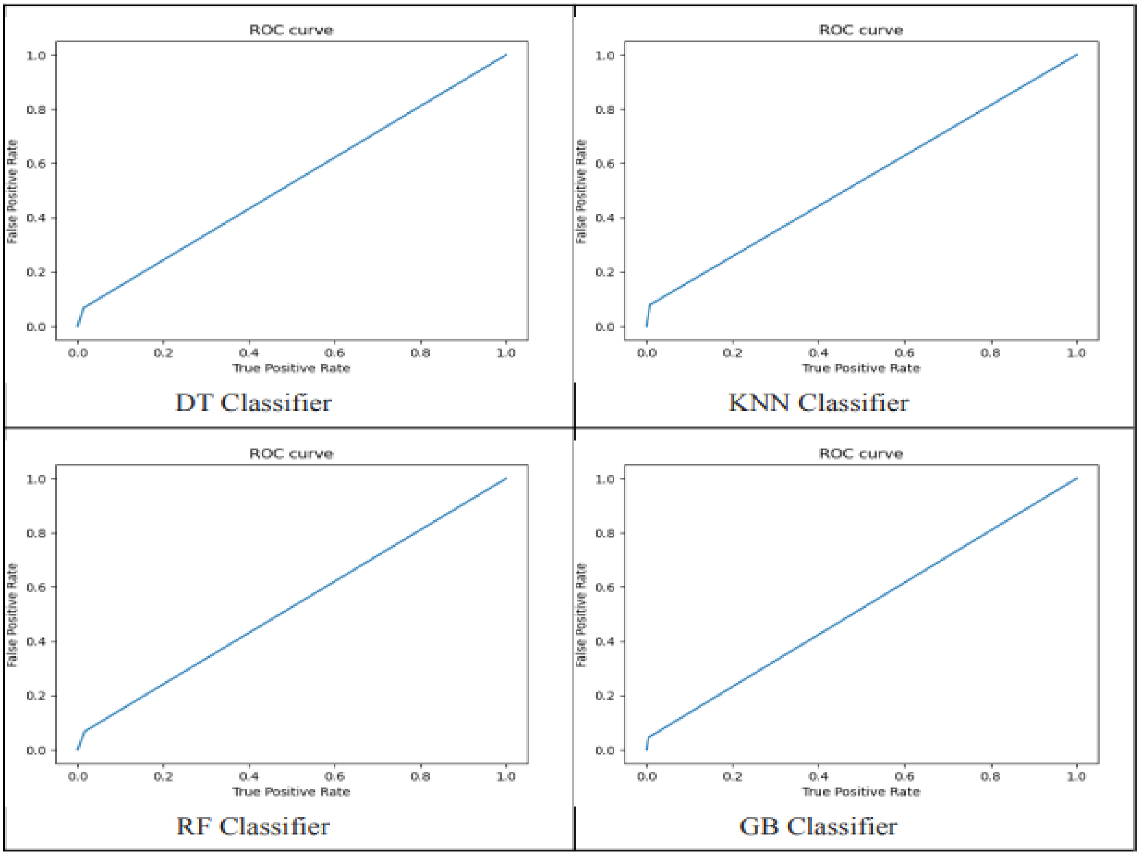

Figure 5: ROC curve

V. CONCLUSION

In this work, the effectiveness of four machine learning algorithms for building trustworthy and precise classifiers was compared.These are Decision tree (DT),K-Nearest Neighbour (KNN), Random Forest (RF) ,Gradient Boosting Classifiers(GBC).The Gradient boosting model outperforms while considering the accident data and it achieved 93% of accuracy. The following factors are utilised to estimate the accident severity: the number of vehicles, the road class, the road surface, the lighting conditions, the spot weather, the casualty class, the sex and age of the casualties, and the kind of vehicle.

This is extraordinarily beneficial for the highway authorities, police departments and for journalists. The key applications of this project are Early accident severity prediction, No expert knowledge required, Can access the model anytime and anywhere, Can access the previous predictions, Can send mail immediately to the respective authorit.The methodology can be used for pre prediction also. Whenever the driver, traveler or passenger starts a journey in a particular area they can predict the accidents happening in that area and the severity of the accident. This can be further used to identify the risk factors, countermeasures. The previous predicted data also can be used to predict the future accident severities and to improve the efficiency of the model.

REFERENCES

[1] Boonserm and N. Wiwatwattana, "Using Machine Learning to Predict Injury Severity of Road Traffic Accidents during New Year Festivals from Thailand's Open Government Data," in *Proceeding of the 2021 9th International Electrical Engineering Congress, iEECON 2021*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 464–467. doi: 10.1109/iEECON51072.2021.9440287.

[2] M. Manzoor *et al.*, "RFCNN: Traffic Accident Severity Prediction Based on Decision Level Fusion of Machine and Deep Learning Model," *IEEE Access*, vol. 9, pp. 128359–128371, 2021, doi: 10.1109/ACCESS.2021.3112546.

[3] J. Zhang, Z. Li, Z. Pu & C. Xu, "Comparing prediction performance for crash injury severity among various machine learning and statistical methods," *IEEE Access*, vol. 6, pp. 60079–60087, 2018, doi: 10.1109/ACCESS.2018.2874979.

[4] A. Ji & D. Levinson, "Injury Severity Prediction from Two-Vehicle Crash Mechanisms with Machine Learning and Ensemble Models," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 1, pp. 217–226, 2020, doi: 10.1109/OJITS.2020.3033523.

[5] H. Kumar, "PREDICTING ACCIDENT SEVERITY USING MACHINE LEARNING," *International Research Journal of Engineering and Technology*, [Online]. Available: www.irjet.net

[6] S. Elyassami, Y. Hamid, and T. Habuza, "Road crashes analysis and prediction using gradient boosted and random forest trees," in *Colloquium in Information Science and Technology, CIST*, Institute of Electrical and Electronics Engineers Inc., Jun. 2020, pp. 520–525. doi: 10.1109/CiSt49399.2021.9357298.

[7] 2019 IEEE National Aerospace and Electronics Conference (NAECON). IEEE, 2019.

[8] N. Dogru & A. Subasi, "Traffic Accident Detection Using Random Forest Classifier."

[9] Institute of Electrical and Electronics Engineers, Chongqing da xue. School of Microelectronics and Communication Engineering, and Chongqing Institute of Electronics, *2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN 2019) : June 12-15, 2019, Chongqing, China*.

[10] Institute of Electrical and Electronics Engineers, *2019 7th International Conference on Smart Computing & Communications (ICSCC).*

[11] Institute of Electrical and Electronics Engineers. Turkey Section. And Institute of Electrical and Electronics Engineers, *HORA 2020 : 2nd International Congress on Human-Computer Interaction, Optimization and Robotic Applications : proceedings : June 26-27, 2020, Turkey.*

[12] J. Paul, Z. Jahan, K. F. Lateef, M. R. Islam & S. C. Bakchy, "Prediction of Road Accident and Severity of Bangladesh Applying Machine Learning Techniques," in *IEEE Region 10 Humanitarian Technology Conference, R10-HTC*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020. doi: 10.1109/R10-HTC49770.2020.9356987.

[13] H. M. Alnami, I. Mahgoub & H. Al-Najada, "Highway Accident Severity Prediction for Optimal Resource Allocation of Emergency Vehicles and Personnel," in *2021 IEEE 11th Annual Computing and Communication Workshop and Conference, CCWC 2021*, Institute of Electrical and Electronics Engineers Inc., Jan. 2021, pp. 1231–1238. doi: 10.1109/CCWC51732.2021.9376155.

[14] Y. Lee, H. Kim, E. Cho, K. Choi, M. Park & S. Park, "A Machine Learning Approach to Prediction of Passenger Injuries on Real Road Situation."

[15] "Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity."

