International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT239038

Accident Severity Detection Using Machine Learning A Review

Nagma Bi, Dr. Halima Sadia
Department of Computer Science & Engineering, Integral University, Lucknow, Uttar Pradesh, India

ARTICLEINFO

Article History:

Accepted: 15 April 2023
Published: 10 May 2023

Publication Issue
Volume 9, Issue 3
May-June-2023

Page Number
69-74

ABSTRACT

One of the greatest challenges in today's world is traffic accidents. It results in fatalities, accidents, and property damage. Making a model that can accurately predict traffic accidents is difficult. The objective of this project is to create a classification system for injuries based on a set of influential factors, including the environment, vehicle speed, driver behaviour, etc. Using data related to traffic accidents, several algorithms are utilized, including AdaBoost, Logistic Regression (LR), Naive Bayes (NB), and Random Forests (RF). Some of the best algorithms are the most effective, including Random Forest, Naive Bayes, and Ada Boost. Compared to LR, NB, and AdaBoost, the RF algorithm performed better, with 75.5% accuracy. To employ various machine learning classification algorithms for traffic accident prediction, the goal of this study is to uncover the underlying causes of road traffic accidents. then decide which prediction model is most likely to help decrease these highway accidents. This paper's goal is to review different authentication procedures offered by numerous scholars around the world.

Keywords : AdaBoost, Logistic Regression (LR), Naive Bayes (NB), and Random Forests (RF)

I. INTRODUCTION

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. According to Michigan Traffic Collision, there were around 314,921 traffic incidents in 2017; these accidents resulted in 1,028 fatalities and 78,394 injuries overall. Freeways are one of the places where severe injuries suffered by those engaged in a collision are more likely to occur.

Several studies have recently examined the impact of factors that affect traffic accidents, mostly focused on people, cars, roads, or the environment. Some researchers [8–9] investigated driver behavior and examined the traits.

Machine learning (ML) enables a computer system to learn from the past data without being directly coded by the developer, which is advantageous because it is

difficult to handle every circumstance on an item [1]. Machine learning algorithms are frequently built on mathematical and statistical principles. The main advantage of machine learning over traditional software is that, as it's challenging to handle every circumstance on an item, there isn't any written code instructing the system how to choose between two things. Machine learning is helpful in this regard.

This study's primary objective is to precisely identify the factors that contribute to severe traffic accidents, which could soon contribute to a reduction in accident frequency and severity.

Machine Learning Classification Model

To evaluate the efficacy of machine learning techniques, the algorithms are trained on a portion of the data and their performance is then evaluated on a testing set. For this study, the data were split into 30% for testing and 70% for training. Four well-known machine learning

classification approaches were researched to create a model for injury severity prediction:

Logistic Regression (LR):
Categorization models include logistic regression. The output of linear functions will be transformed into sigmoid functions using the procedure. The linear regression model is an easy-to-understand and practical mathematical approach.

Random Forest Model (RF):
Several decision trees are built using the Random Forest model, an ensemble learning technique, during the training phase. This model then generates a class that is the mean of the classification or mean regression of all the individual trees. The minimum samples needed to split a node were set to two, whilst the minimum samples needed for each leaf were set to one.

Naïve Bayesian Classifier (NBC):
A group of simple "probabilistic classifiers" referred to as Naive The Bayes theorem is the foundation of Bayes classifiers, which also make strong (naive) assumptions about the independence of the features. Given that the feature set contains continuous variables, the Gaussian NB was chosen.

AdaBoost Classification Tree:
A classification method is AdaBoost. that regularly invokes particular weak learner algorithms throughout several rounds. It is a binary boosting algorithm, maybe the most crucial one serving as the fundamental building block for numerous other classification algorithms

II. Methodology

1.Data Preprocessing

Essentially, data preparation is the act of transforming raw data into a format that can be understood. Data from the real world is frequently inaccurate, inconsistent, devoid of particular behaviours or trends, and all of these things. One tried-and-true method for solving these issues is preprocessing data. Data

preparation is the procedure that gets raw data ready for processing.

The information used in this study includes several columns, including year, speed limit, weather, and road workers in action. However, since accidents and injuries are the only topics we are interested in, all unnecessary data are preprocessed (removed), and only pertinent columns are included in the analysis.

2.Association Learning

In large databases, interesting correlations between variables can be discovered by using the rule-based machine learning technique known as association rule

learning. It utilises some interestingness measurements to find strong rules in databases.Association rules can be generated using a variety of techniques. Apriori, Eclat, and FPGrowth are a few of the more well-known algorithms.

3.Generating Rules and patternPrediction

The method processes the aforementioned inputs, rejects the weak rules based on the restrictions, and then iteratively creates strong association rules. These powerful association rules are then chosen with an assurance level of 80%, resulting in various patterns among various incidents and injuries.

4.Result Analysis

The suggested model reveals patterns among various accident types and injuries, such as the link between brain injuries and fractured bones or the link between driving too fast and hit-and- run, etc. In this investigation, the minimum support count and confidence levels were both set at 2. These parameters can be adjusted to different levels depending on the dataset, which enables the production of more accurate results.

III. COMPARATIVE STUDY

S.No.	Year	Author(s)	Keywords	Proposed Method	Findings	Future work
[1]	2021	Hani M Alnami	Data Mining and ML, Classification algorithms, Intelligent Transportation System(ITS), Vehicle Ad Hoc Networks (VANET), Big Data	Severity Prediction of Highway Accidents for Optimal Resource Allocation of Emergency Vehicles	Random Forest has the highest accuracy, suitable for the real-time handlingof traffic data	To build a system that predicts traffic flow on the highway in real time.
[2].	2021	Eakapan Boonserm	Imbalanced Classification, Machine Learning, Traffic Accidents, Random Forest	Using ML to Predict Injury Severity of Road Traffic Accidents During New Year Festivals	random undersampling and oversampling with SMOTE used, analyzing accidents' hot spots and crucial factors	Efficiency Can be increased
[3]	2021	Mubariz Manzoor et al	Road accidents severity, random forest, convolutional neural network, feature importance, ensemble learning	RFCNN: Traffic Accident Severity Prediction	ensemble of machine learning and deep learning models by combining Random Forest and Convolutional Neural Networks called RFCNN	find out the main factors of the accident
[4]	2021	Joy Paulet al	ML, Multilayer Perceptron, Categorical Naive Bayes.	The severity of Bangladesh Applying ML	video data captured	Many facts relating to injuries
[5]	2020	ANG JI	Injury severity, ML,	Injury severity	Two-Vehicle	Large datasets

			vehicle crashes, ensemble technique, crash mechanisms	prediction of Two-Vehicle Crash Mechanisms With ML	Crash Mechanisms	with fewer unknown items should enable
[6]	2020	Sanaa Elyassami	Crash Data Analysis, Decision Tree, Machine Learning, Random Forest, Gradient Boosted Tree	Prediction using Gradient Boosted and Random ForestTrees	Regressive testing, Crash data analysisis indispensable	Incorporating driver behavior data
[7]	2020	Hemanth Kumar et al.	WHO, Non-fatal injuries, Economic loss, ML, Unsupervised Learning, Patterns, Eclat	PREDICTING ACCIDENT SEVERITY USING ML	Focused not only on the type of accidents but also on the different injuries	Combine road-related factors with driver information
[8]	2020	Buket Geyik Et al.	severity prediction, ML, deep learning, IoT, big data analysis, STATS19 dataset, accident severitylevels	Accident Severity Prediction with ML Methods	fatal, serious, and slight.	Implementedinto the motorcycle accidents
[9]	2019	Md. Farhan Labib	Accident Severity, ML, Supervised Learning Feature Analysis, Road Accident.	Prediction of Accident Severity by Using ML in Bangladesh	AdaBoost performed best, determine the intensity of accidents	create a mobile application
[10]	2019	Bulbula Kumeda et al.	ML; road traffic accident; fuzzy FARCHD; classification	Classification ofRoad Traffic Accident Data Using ML	The fuzzy FARCHD algorithm performed well	More features, and clusters
[11]	2019	Abdulla Ali et al.	ML, traffic accident prediction, driving simulation	Freeway Traffic Accidents Prediction	Random Forest performed well, monitor freeway crashes	Efficiency increase
[12]	2019	Rabia Emhamed Al Mamlook et al.	Random Forest, Logistic Regression, Naïve Bayes, AdaBoost, Traffic Accident Severity.	The severity of accident by ML	Random Forest performed well, Identify the Key factors	Add more data characteristics
[13]	2018	Jian Zhang	crash severity; statistical model; machine learning; accuracy; variable importance	Comparing Prediction Performance for Crash Injury Severity	RF method had the best prediction	consider more machine learning methods
[14]	2018	Nejdet Dogru	IoT, ANN, Random forest, SVM, VANETs	Traffic Accident Detection Using Random Forest Classifier	Trained equipment, V2V installed with a mobile communication	More Efficiency
[15]	2018	Yongbe o m Lee	NASS-CDS, machine learning, decision tree, prediction of injury grade	ML Approach to Prediction of Passenger Injurieson Real Road Situation	Degree of injury of passengers	Automati c notification system.

IV. FINDINGS AND DISCUSSION

This research found these gaps in previous research of this Project that required further research on this topic. It has these limitations-

1. One drawback of the current study is that certain variables (such as the characteristics of the driver, passenger, and pedestrian, as well as traffic circumstances), which can have an impact on the severity and duration of accidents, were not taken into account due to a lack of adequate data.
2. In the future, this study can be used to motorbike accidents by gathering additional data and applying new classification models to improve the classification's quality.

In the future, it is preferable to employ deep learning approaches for improved classification and a cluster of the data records due to the frequently growing size of the data sets, more features, and clusters.
4. The majority of the shortcomings of earlier research on accident prediction either had an accident prediction focus or some of them just took the accident severity into account.

V. CONCLUSION

This study compared the efficacy of four machine learning techniques for producing reliable and accurate classifiers. These include the Naive Bayesian Classifier, Logistic Regression, AdaBoost, Random Forest, and Logistic Regression algorithms.. It is necessary to conduct more research to gather relevant data and examine how these elements have an impact. The best model for forecasting motorway crashes, Random forest, is advised to be used in monitoring. fatal and severe wounds. The key cause causing traffic crashes can be quickly and effectively identified using the suggested predictive model.

This paper also points out various research gaps in the existing systems that need to be focused on in further research.

VI. REFERENCES

[1]. H. M. Alnami, I. Mahgoub, and 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, Jan. 2021, pp. 1231–1238. doi: 10.1109/CCWC51732.2021.9376155.

[2]. E. 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, Mar. 2021, pp. 464–467. doi: 10.1109/iEECON51072.2021.9440287.

[3]. 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.

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

[5]. A. Ji and 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, Oct. 2020, doi: 10.1109/ojits.2020.3033523.

[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, Jun. 2020, vol. 2020-June, pp. 520–525. doi: 10.1109/CiSt49399.2021.9357298.

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

[8]. 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.

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

[10]. 2019 IEEE National Aerospace and Electronics Conference (NAECON). IEEE, 2019.

[11]. J. Zhang, Z. Li, Z. Pu, and 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.

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

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

Cite This Article :

Nagma Bi, Dr. Halima Sadia, "Accident Severity Detection Using Machine Learning A Review", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 3, pp.69-74, May-June-2023. Available at doi : https://doi.org/10.32628/CSEIT239038

Journal URL : https://ijsrcseit.com/CSEIT239038