ISSN: 1118-5872

FOS

MULTI-DISCIPLINARY JOURNAL

(Alvan Ikoku Federal University of Education)

Prediction of Road Traffic Accidents using Machine Learning Techniques

[1]Taiwo, E. O., Uzoma, P. O., [3]Odoemene, I. O. & [4]Ihim, K. C.

Department of Computer Science, Alvan Ikoku Federal University of Education, Owerri.
Corresponding author email: emmanuel.taiwo@alvanikoku.edu.ng

Abstract

Road traffic accident is a major concern in the whole world. A large number of people die annually due to road accidents, causing irreparable losses and emotional distress to their family members and society. The study analyzed the effectiveness of three machine learning techniques on the road traffic accidents dataset that was obtained for a period of three years (2020-2022) from the National Bureau of Statistics (NBS). The dataset consists of 27 attributes and 481 instances which were further divided into 75% for training and 25% for testing. The metrics used for evaluation include: accuracy, prediction speed, training time, and area under curve (AUC). The dataset was used to formulate predictive models for road traffic accidents using Support Vector Machine (SVM), Classification Tree (CT), and Ensemble Bagged Tree (EBT) algorithms. The formulation and simulation of the predictive models were carried out using the Matrix Laboratory (MATLAB) statistical tool. The results show accuracy (%) of 56.3, 43.2, and 55.7. Execution Time (Secs) of 66.485, 1.4291, 5.466. Prediction Speed (Obs/Sec) of 120, 14000, 1600, and AUC of 0.79, 0.73, 0.82 for SVM, CT, and EBT techniques respectively. The three models were compared and the best model in terms of accuracy was selected and validated. The study concluded that the SVM model recorded the highest prediction accuracy followed by EBT and CT based on the dataset used. The study, therefore, recommends that the performance of other machine learning techniques could be tested on the dataset.

Keywords: Prediction, Road traffic accident, Support vector machine, Classification tree, Ensemble bagged tree

Introduction

Road traffic accidents is a major concern in the whole world. The alarming rate of road traffic accident daily is worrisome due to its exponential increase. Yearly, numerous people witness human losses such as injury due to traffic accidents or death, causing accident victims, their families, and society to have emotional distress (Tsujimura-Ito, 2019). Road traffic accidents contribute to loss of productivity, road accident fund payouts for the individual involved and country as well as huge financial losses in terms of infrastructure damage (Ogwueleka et al., 2014; Yan & Shen, 2022). Despite of putting in place highway codes and set regulations in Nigeria, a lot of accidents have been caused by speed violations, use of a phone while driving, tyre bursts, mechanically deficient vehicles, brake failure, overloading, dangerous overtaking, wrongful overtaking, dangerous driving, bad road, route violation, sleeping on steering, fatigue and driving under alcohol/drug influence. Although an increase in the number of vehicles is the major cause of road accidents, still we cannot neglect the role played by environmental factors and the condition of the roads. Accidents involving commercial vehicles used for public transportation such as buses and those

involving heavy goods vehicles such as trucks are the most fatal kind of accidents that occur, claiming innocent people's lives. The risk of accidents is also catalyzed by the role play by weather conditions such as rain, harmattan, etcetera. With incessant losses, one of the factors in attaining global sustainability goals in transportation management is the prevention/reduction of traffic accidents or fatalities (Bokaba et al., 2022). This requires the development of a prediction model for road traffic accidents. Due to advancements in Artificial Intelligence (AI), a branch of AI known as Machine Learning (ML) has techniques that have been found useful in a variety of domains for data interpretation purpose in fraud detection, credit risk management, recommendation systems, healthcare informatics, etc. The objectives of this paper are to identify variables causing traffic accidents; formulate the model based on the variables identified; simulate the model formulated; validate the model using evaluation metrics; and select the best model in terms of accuracy.

Literature Review

Support Vector Machine

Support Vector Machine is a supervised learning technique used for either classification or regression. It separates the data points in different classes in the feature space due to finding the optimal hyperplane in an N-dimensional space. It adds a penalty each time a point crosses the hyperplane with previous datasets to find the maximum margin. It also has the characteristics to find the best hyperplane that maximizes the margin and ignores the outlier.

$$W^T = (\phi(x)) + b \qquad (1)$$

Where b is intercept, W is distance, and $\phi(x)$ is the point vector of the hyperplane

Classification Tree

It is a supervised learning technique that could be used for both classification and regression problems. It is a tree structure where nodes, branches, and leaf nodes represent the features of a dataset, decision rules, and the outcome respectively. There are two nodes in a classification tree, which are the decision node and the leaf node. Decision nodes are used to make any decision and have multiple branches, whereas leaf nodes are the output of those decisions and do not contain any further branches. It is called a classification tree because it is similar to a tree. It starts with the root node that expands further branches and constructs a tree-like structure. The classification tree is expressed in equation 2.

$$G(x) = 1 - \sum_{i=1}^{b} p_i^2 \qquad (2)$$

Where G(x) is the Gini impurity at node x and p_i is the proportion of observations of class b at node x

Ensemble Bagged Tree

Ensemble Bagged Tree is a learning technique that could be used for both classification and regression. Bagging means Bootstrap Aggregating which means that the technique creates multiple subsets of the training data (bootstrap sample), trains a classification tree on each subset, and combines the predictions from all the trees to make a final prediction. This approach helps to reduce overfitting by averaging out individual tree errors; improve generalization by considering multiple perspectives; and increase robustness by reducing reliance on a single tree.

Related Works

Esswidi et al. (2023) proposed a model to forecast traffic road accident severity. They deployed random forest, Decision tree, ANN, and KNN algorithms to build models. The experimental results showed that ANN outperformed the other three models. Dattatray et al. (2022) proposed a model to predict road accidents. They utilized support vector machines (SVM), logistic regression (LR), and artificial neural networks (ANNs) to construct models. Results indicated that LR achieved the highest accuracy. Ahmed et al. (2021) developed a model to predict road accident severity. They employed K-Nearest Neighbor (KNN), Naive Bayes (NB), Logistic Regression (LR), Random Forest (RF), Adaptive Boosting (AdaBoost), and Adaptive Boosting (AdaBoost) to build models. Experimental results revealed that the RF model produced the best accuracy. Mane and Rathod (2022) proposed a model to predict road accidents. They used various algorithms such as KNN Classifier, Gradient Boosting, Naïve Bayes, Random Forest, and Logistic Regression to build road accident models. The results showed that the Naïve Bayes model performed better than the others. Venkateswarlu et al. (2023) proposed a model to forecast road accidents. They deployed different data mining and machine learning techniques to build models. The results indicated that the model has high accuracy to predict road accidents. Ardakani et al. (2023) developed a model to forecast road car accidents. They employed multinomial logistic regression, decision trees, naïve Bayes, and random forest to construct models. Experimental results indicated that the Random forest model performed better than the other models.

Kumar et al. (2022) developed a model to predict road accidents. They used a hybrid-based approach to build model. The result showed that the approach produced a good performance. Pardhi et al. (2023) proposed a model to predict the severity of road accidents. They employed various machine learning algorithms to construct models. Experimental results revealed that the Random Forest model performed better than other models. Pooja (2020) developed a model to predict road accidents. They used a hybrid K-means and random forest algorithm to build a model. The result indicated that among different classification techniques used, the approach achieved high accuracy. Vanitha and Swedha (2023) developed a model to forecast road accidents. They deployed Logistic Regression, Decision Tree, and Random Forest to construct models. The results showed that Random Forest outperformed the two other models. Tchoukouegno and Adedeji (2023) proposed a model to predict road traffic accidents. They utilized LightGBM, XGBoost, and Decision Tree to build models. The results revealed that the Decision Tree achieved the highest accuracy. Indraja et al. (2019) proposed a model to predict road accidents in India. They used data mining techniques to build a model. The result showed that the model enhances road safety in India.

Methodology

Datasets and Attributes: The datasets were obtained from National Bureau of Statistics (https://nigerianstat.gov.ng/elibrary/read/1241392#:~:text=The%20total%20number%20of%20road,from%203%2C407%20in%20Q4%202021.). The dataset consists of 27 attributes namely: state, fatal, serious, minor, total_cases, number_injured, number_killed, total_casualty, people_involved, speed_violation, use_of_phone_while_driving, tyre_burst, mechanically_deficient_vehicle, brake_failure, overloading, dangerous_overtaking, wrongful_overtaking, dangerous_driving, bad_road, route_violation, road_obstruction_violation, sleeping_on_steering, driving_under_alcohol/drug_influence, poor_weather, fatigue, sign_light_violation, others. The data consists of information collected from 481 instances.

Data Analysis: Three different supervised learning techniques: Support Vector Machine, Classification Tree, and Ensemble Bagged Tree were predictive models. The values of the 27 features were used to define the input variables of the prediction model. The Matrix Laboratory

(MATLAB) statistical tool was used to develop the predictive models. The percentage splits of 75% training and 25% testing were used for the prediction. Each model was compared and the most efficient model was chosen based on evaluation criteria and results.

Parameter Tuning: This involves fine-tuning the machine learning model's parameters to reach its best performance. 10-fold Cross Validation was employed to assess the effectiveness of the machine learning models by training and testing it on the datasets.

Results and Discussion

Results

The results of the road traffic accidents model developed for three machine learning algorithms in this paper are shown in Figure 1-3. The evaluation of the model is also shown in Table 1.

Support Vector Machine model accuracy is 56.3%. It has execution time, prediction speed, and area under curve (AUC) of 66.485secs, 120 obs/sec, and 0.79 respectively as shown in Figure 1.

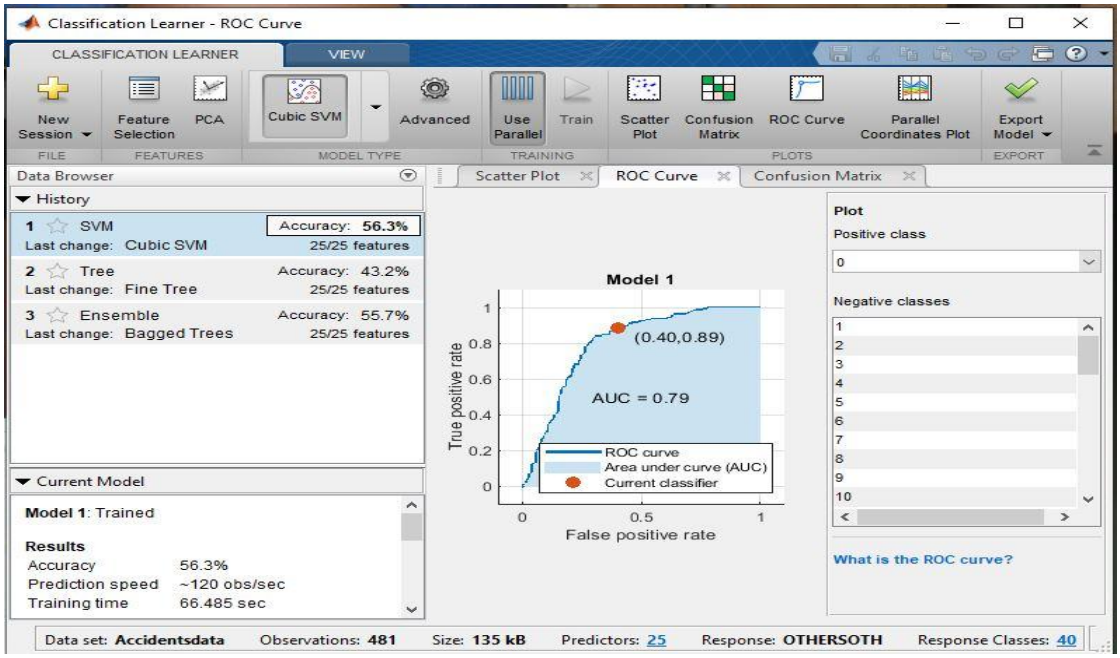

Figure 1: Support Vector Machine Technique for Road Traffic Accident Model

Classification Tree model accuracy is 43.2%. It has execution time, prediction speed, and area under curve (AUC) of 1.4291 secs, 14000 obs/sec, and 0.73 respectively as shown in Figure 2.

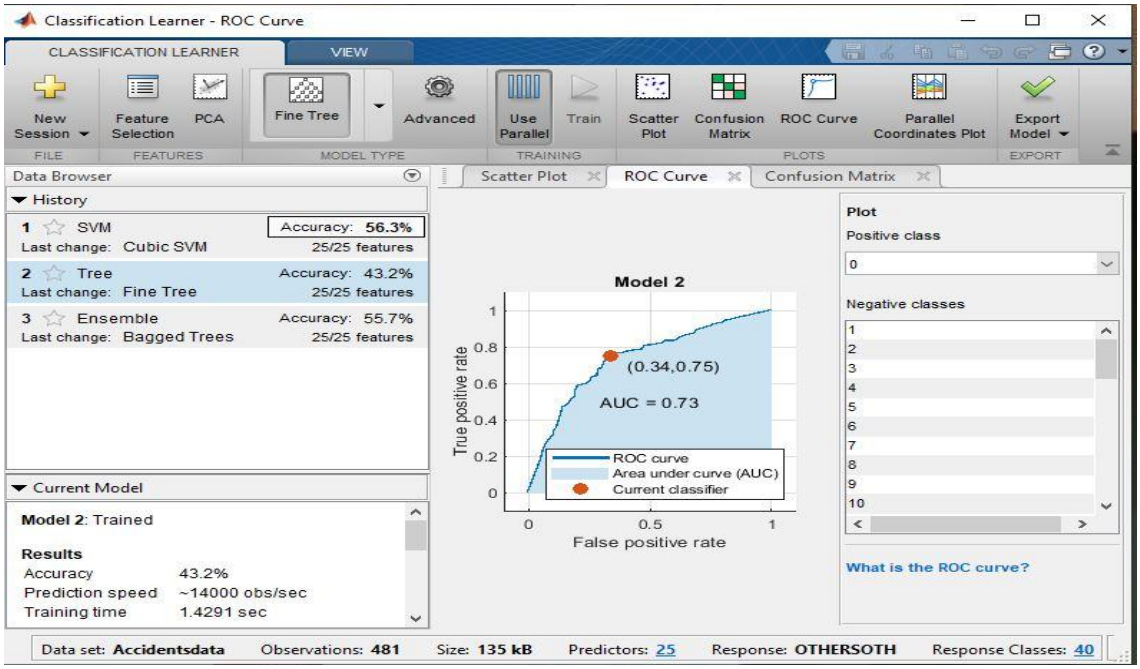

Figure 2: Classification Tree Technique for Road Traffic Accident Model

Ensemble Bagged Tree model accuracy is 55.7%. It has execution time, prediction speed, and area under curve (AUC) of 5.466 secs, 1600 obs/sec, and 0.82 respectively as shown in Figure 3.

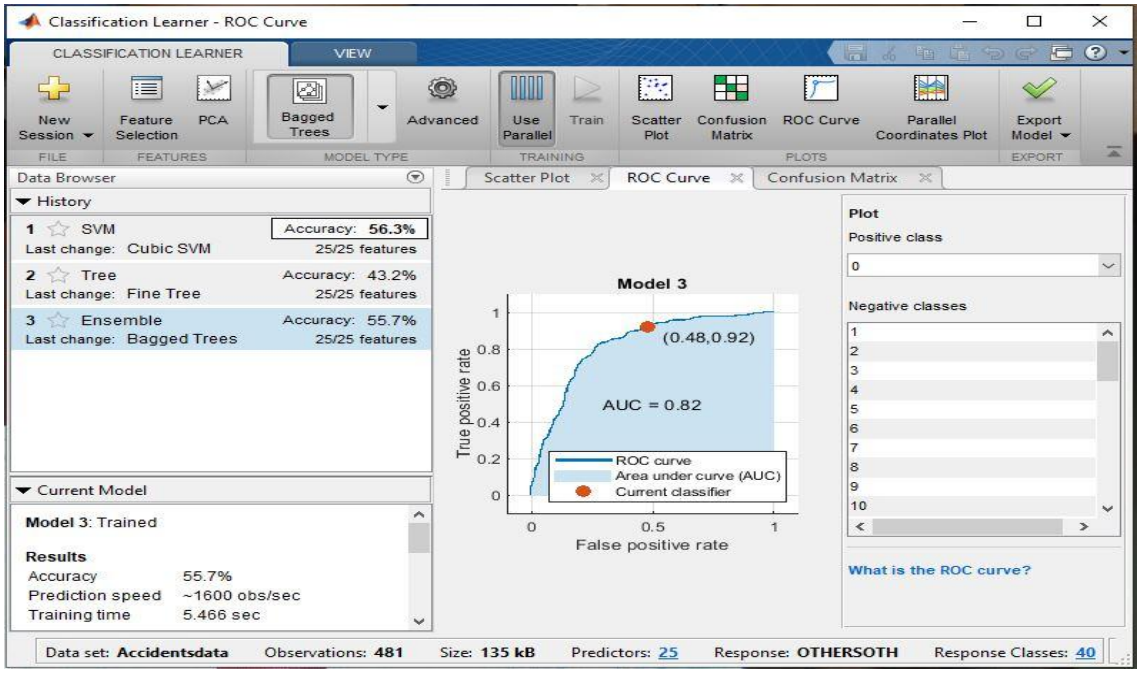

Figure 3: Ensemble Bagged Tree Technique for Road Traffic Accident Model

Table 1: Results for the three Techniques used for road traffic accident predictive model

	Support Vector Machine	Classification Tree	Ensemble Bagged Tree
Traffic Dataset	**75% 25%**	**75% 25%**	**75% 25%**
Number of Instances	481	481	481
Accuracy	56.3%	43.2%	55.7%
Execution Time (Seconds)	66.485	1.4291	5.466
Prediction Speed (Obs/Sec)	120	14000	1600
Area under curve (AUC)	0.79	0.73	0.82
Validation	10-fold Cross Validation	10-fold Cross Validation	10-fold Cross Validation

Discussion

Ensemble Bagged Tree accuracy is higher compared to Support Vector Machine and Classification Tree techniques for the road traffic accident dataset divided into 75%, 25%. The execution time of the Classification Tree technique is faster compared to the other two techniques as shown in Table 1. In terms of prediction speed, the Support Vector Machine algorithm has the fastest speed compared to the Classification Tree technique and Ensemble Bagged Tree technique as shown in Table 1. The AUC is 0.79, 0.73, and 0.82 for the Support Vector Machine, Classification Tree, and Ensemble Bagged Tree respectively. The model prediction was right based on the values of AUC for the three techniques which is close to 1.

Conclusion

We presented three machine learning techniques: Support Vector Machine, Classification Tree, and Ensemble Bagged Tree in this paper to build a model that can be used to predict road traffic accidents. The traffic dataset was split into two, 75% for training and 25% for testing. The predictive model was implemented on the MATLAB statistical tool. The Ensemble Bagged Tree technique achieved the highest prediction accuracy followed by the Support Vector Machine and Classification Tree respectively. The model can be used in transportation management and other machine learning techniques could be tested on the dataset.

References

Ahmed, S., Hossain, M. A., Hossain, M. M. I. & Ray, S. K. (2021). A comparative study of machine learning algorithms to predict road accident severity. *20th International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI/SmartCNS), 392-399. 978-1-6654-6667-7/21/$31.00 ©2021 IEEE DOI 10.1109/IUCC-CIT-DSCI-SmartCNS55181.2021.00069.*

Ardakani, S. P., Liang, X., Mengistu, K. T., So, R. S., Wei, X., He, B. & Cheshmehzangi, A. (2023). Road car accident prediction using a machine-learning-enabled data analysis. *Sustainability*, 15(5939). https://doi.org/10.3390/su15075939

Bokaba, T., Doorsamy, W. & Paul, B. S. A. (2022). Comparative study of ensemble models for predicting road traffic congestion. *Appl. Sci.* 12, 1337.

Dattatray, G. T., Shubhangi, D. G., Vajid, N. K., Atul, R. & Satish, N. G. (2022). Road accident prediction model using data mining techniques. *NeuroQuantology*, 20(16), 2904-2101. doi: 10.48047/NQ.2022.20.16.NQ880299

Esswidi, A., Ardchir, S., Daif, A. & Azouazi, M. (2023). Severity prediction for traffic road accidents. *Journal of Theoretical and Applied Information Technology*, 101(8), 3198-3204. ISSN: 1992-8645. E-ISSN: 1817-3195.

Indraja, E., Smitha, M., Tabitha P. R. & Chandini, S. B. (2019). Development of a predictive model for road accident prediction in India. *International Journal of Innovative Science and Research Technology*, 4(5), 1261- 1263. ISSN No: 2456-2165.

Kumar, G. A., Viswanath, P. K., Wamique, M. & Kumar, I. (2022). Road accident prediction model using data mining techniques. *International Journal of Creative Research Thoughts (IJCRT)*, 10(5), 41-50. ISSN: 2320-2882.

Mane, G. & Rathod, V. U. (2022). Road accident prediction using machine learning techniques.

International Journal of Research in Engineering and Science (IJRES), 10(71), 808-815. ISSN (Online): 2320-9364, ISSN (Print): 2320-9356.

Ogwueleka, F. N., Misra, S., Ogwueleka, T. C. & Fernandez-Sanz, L. (2014). An artificial neural network model for road accident prediction: A case study of a developing country. *Acta Polytech. Hung.*, 11, 177-197.

Pardhi, P., Saw, V. K., Sahu, D., Thakur, P. S. & Sinha, V. K. (2023). Predicting the severity of road accidents and classifying them into three categories using machine learning. *International Journal of Research Publication and Reviews*, 4(5), 4921-4925. ISSN 2582-7421.

Pooja. Y. S. S. (2020). Road accident prediction and model interpretation using a hybrid k-means and random forest algorithm approach. SN Applied Sciences, 2(1576). https://doi.org/10.1007/s42452-020-3125-1

Tsujimura-Ito, T. (2019). State of damage to and support for victims of motor vehicle accidents in Japan. *IATSS Res.*, 43, 97-107.

Tchoukouegno, M. M. & Adedeji, J. A. (2023). Machine learning for road traffic accident improvement and environmental resource management in the transportation sector. *Sustainability*, 15(2014), 1-19. https://doi.org/10.3390/su15032014

Vanitha, R. & Swedha, M. (2023). Prediction of road accidents using machine learning algorithms. *Middle East Journal of Applied Science & Technology (MEJAST)*, 6(2), 64-75. https://doi.org/10.46431/MEJAST.2023.6208

Venkateswarlu, K., Bhavani, P., Reddy, V. V., Sathish, N. & Devi, G. R. (2023). A road accident prediction model using data mining techniques. *Internal Journal for Recent Development in Science and Technology*, 7(6), 297-305. ISSN 2581 – 4575.

Yan, M. & Shen, Y. (2022). Traffic accident severity prediction based on random forest. *Sustainability*, 14, 1729.