



INTERNATIONAL JOURNAL OF SCIENCE FOR GLOBAL SUSTAINABILITY
**Predictive Analysis of Road Traffic Accidents in Katsina State,
 Nigeria Using Machine Learning Algorithms: A Study on Factors and
 Mitigation Strategies**

Umar Iliyasu, Muhammad Muntasir Yakudu and A.A. Abdulwasii

Department of Computer Science, Faculty of Computing,
 Federal University Dutsin-Ma, Katsina State, Nigeria.

Corresponding Author's E-mail: umariliyasut@gmail.com

Received on: April, 2023

Revised and Accepted on: May, 2023

Published on: July, 2023

ABSTRACT

This study aims to explore the use of machine learning algorithms in predicting the likelihood of road traffic accidents in Katsina State, Nigeria, with the goal of reducing the high rate of road traffic accidents and associated loss of lives and properties. The study collects and analyzes data on road traffic accidents in Katsina State, including the number of accidents, location, time of day, and the type of vehicles involved. Machine learning algorithms such as Decision Trees, Random Forest, and K-Nearest Neighbors, are trained using the data to predict the likelihood of road traffic accidents. The study also identifies the factors that contribute to road traffic accidents in Katsina State by using techniques such as feature selection and correlation analysis, to identify the most important variables. The findings can be used by stakeholders, including the government, law enforcement agencies, and road safety organizations, to develop and implement effective strategies to reducing road traffic accidents in Katsina State. The study utilizes data collected from the Federal Road Safety Corps database in Katsina, Nigeria, and employs data cleaning and feature selection techniques to improve data quality. The Random Forest algorithm achieved the predictive value as 85% while K-Nearest Neighbors (KNN) and Decision Tree algorithms yielded 17% and 42% respectively.

Keywords: Machine Learning, Road traffic accident, K-Nearest Neighbour, Random Forest and Decision Tree.

1.0 INTRODUCTION

The study on predicting road traffic accidents in Katsina State using machine learning aims to explore the use of machine learning algorithms in predicting the likelihood of road traffic accidents in Katsina State, Nigeria. Katsina

State is located in the north-western region of Nigeria and has a high incidence of road traffic accidents. The study is motivated by the need to reduce the high rate of road traffic accidents in Katsina State, which has resulted in the loss of lives and properties. Machine learning algorithms have been shown to be effective in predicting various phenomena, including traffic accidents.

The study will involve collecting and analyzing data on road traffic accidents in Katsina State, including the number of accidents, the location, the time of the day, weather conditions, and the type of vehicles involved. Machine learning algorithms such as Decision trees, Random

Forest, and K- Nearest Neighbour will be trained using the data to predict the likelihood of road traffic accidents (Zheng et al., 2019). The findings of the study will provide insights into the factors that contribute to road traffic accidents in Katsina State and how machine learning algorithms can be used to predict and prevent accidents. This information can be used by stakeholders, including the government, law enforcement agencies, and road safety organizations, to develop and implement effective strategies for reducing road traffic accidents in Katsina State. The study on predicting road traffic accidents in Katsina State using machine learning aims to explore the use of machine learning algorithms in predicting the likelihood of road traffic accidents in Katsina State.

The study will also explore the performance of different machine learning algorithms in predicting road traffic accidents in Katsina State.

This will involve comparing the precision, recall, and F1-score of the algorithms to determine which algorithm is most effective in predicting road traffic accidents in Katsina State.

In addition to predicting the likelihood of road traffic accidents, the study will also explore the use of machine learning algorithms in identifying the factors that contribute to road traffic accidents in Katsina State. This will involve using techniques such as feature selection and correlation analysis to identify the most important variables that influence the occurrence of road traffic accidents in the region. The study will also examine the potential of integrating other data sources, such as traffic flow data and social media data, into the analysis to improve the accuracy of the predictions. This could involve using data from traffic cameras and sensors to monitor traffic flow and identify areas with high traffic volume and congestion, which could be potential hotspots for road traffic accidents (Sharma et al., 2016).

Overall, the study on predicting road traffic accidents in Katsina State using machine learning is expected to contribute to the development of effective strategies for reducing the incidence of road traffic accidents in the region. By identifying the factors that contribute to road traffic accidents and predicting their likelihood, stakeholders can develop targeted interventions that can help prevent accidents and save lives.

2.0 RELATED WORK

Many research studies have used data mining, machine learning, and deep learning algorithms to predict road accidents. Popular clustering and classification algorithms such as K-means, Support Vector Machines, K-Nearest Neighbors (KNN), Decision Tree (DT), Artificial Neural Network (ANN), Convolution Neural Network (CNN), and Logistic Regression (LR) have been used to develop models that predict accident severity.

Kwon et al., (2015) applied Naive Bayes (NB) and Decision Tree (DT) algorithms to a dataset collected in California from 2004 to 2010. They used binary regression to compare the performance of the developed models and found

that NB was more sensitive to risk factors than the DT model.

The development of an optimized radial basis function neural network (RBFNN) was used to predict accident frequency and identify significant risk parameters for road accidents. They discovered that optimizing the RBFNN significantly improved the accuracy of their estimation. However, they found that other variables had little effect on the frequency of crash occurrence. The study also revealed that the RBFNN outperformed the Naive Bayes (NB) and back-propagation neural network (BPNN) models for predicting crash levels when compared to the standard negative binomial (NB) model (Huang et al., 2016).

Zheng et al., (2019) Proposed a deep learning approach using a CNN model to predict road crash severity. They transformed accident data into a grey image representing the weight of traffic accident characteristics, which was then used as input for the CNN model. Using traffic accident data from 2009 to 2016, the Leeds City Council tested the effectiveness of this CNN model and found that it outperformed other models, including K-nearest neighbor algorithm, logistic regression, gradient boosting neural network, support vector machines, and LSTM-RNN. The FM2GI algorithm, Sameen and Pradhan, (2017) was also introduced to facilitate the predictive performance of traffic accident severity by combining the estimation of feature weights, filling the features into the all-zero matrix, and transforming them into images as CNN input.

The analysis of road accident data using SVM and MLP on a limited number of datasets (300 datasets) was considered by Sharma et al., (2016) and only considered two independent variables (alcohol and speed) as key factors. They found that driving at high speed after drinking was the main cause of accidents and that SVM with RBF kernel had a better accuracy rate (94%) than MLP (64%).

Wahab et al., (2019) conducted a study on motorcycle crashes in Ghana using MLP, PART, and SimpleCART classifiers to evaluate their performance and identify the major factors for the

crashes. They used Weka tools to analyze datasets and InfoGainAttributeEval to determine the most influential variable for motorcycle crashes in Ghana. They found that the SimpleCART model had better accuracy than the other classification models.

The use of k-means and Association Rule data mining techniques to uncover the frequency of accident severity locations and extract hidden information was proposed by Kumar and Toshniwal, (2016). They selected 87 out of 158 locations with accident frequency counts greater than 20 and applied k-means clustering to group them into three clusters, determined by gap statistics. A minimum support of 5 percent was used to obtain rules. The results revealed that curved and sloped locations on hilly surfaces were prone to accidents. The study was conducted on FARS dataset during 2007.

Rahman et al., (2019) used machine learning techniques to analyze the risk of pedestrian and bicycle accidents by considering various traffic, roadway, and socio-demographic characteristics. They applied DTR modeling analysis and spatial predictor variables to improve the accuracy of neighboring STAZs assessment. The DTR model revealed the significance of traffic, roadway, and social characteristics as predictor variables for pedestrian and bicycle accidents. The results highlighted the need to improve infrastructure and identify areas with high pedestrian and cyclist traffic. The study also adopted ensemble techniques like sacking, random forest, and gradient boosting to enhance the accuracy of pedestrian and bicycle accident assessment. The results showed that the boosting algorithm outperformed other competitive ensemble methods in terms of good-fit scales.

Hossain et al., (2019) discussed the dangers of driving and the steps being taken to ensure safe travel for road users through advanced ITS-based technologies. RTCPM is a program that aims to improve road safety and become an integral part of the Real Time Proactive Road Safety Management System. The study conducted a review of real-time crash prediction models and outlined six key design requirements for such models. Progress has been made in utility,

performance, and content information production approach.

Annie et al., (2020) conducted a study that analyzes the severity of road accidents and the factors that contribute to them, including lighting conditions. Improvements in these factors can lead to safer roads and fewer accidents. The analysis of data can help government and citizens take precautions to keep themselves safe.

Labib et al., (2019) emphasized the need to control and manage traffic with sophisticated systems to reduce the number of road accidents, which can be prevented by simple precautions based on advanced predictions or warning systems. The implementation of machine learning can help make accurate decisions and reduce the number of traffic accidents.

3.0 METHODOLOGY

Data collection is a crucial step in machine learning algorithms to analyze performance. For road accident analysis, it is essential to have an extensive and accurate dataset. However, obtaining an ideal dataset can be challenging. The quality of the data is crucial for data analysis, and it is essential to pay attention to analyzing and understanding the content and structure of the data. In this research, the dataset used was collected from the Federal Road Safety Corps database in Katsina, Nigeria. The dataset contains various features, such as the command, date, crash time, report time, arrival time, response time, route, location, vehicle number, vehicle type, vehicle category, vehicle model, name of organization/fleet operator, name of driver, driver's license number, probable cause(s) of the road traffic collision, number of injured and killed individuals, and number of individuals involved in the accident, broken down by gender and age group. After data collection, data pre-processing is performed to improve the data quality by removing noisy, incomplete, or irrelevant data. In this research, data cleaning was used for data pre-processing followed by feature selection. The K-Nearest Neighbors (KNN) and Decision Tree algorithms were used for analyzing the factors that contribute to road accidents. Detail on each step carried out are discussed here. And the architectural structure is shown in figure 3.1

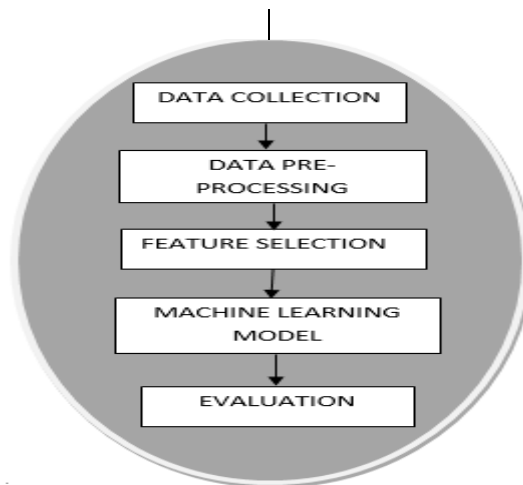


Figure 3.1: Architectural Structure

3.1 Data collection

To analyze performance using machine learning algorithms, it is crucial to have extensive and accurate data records of road accidents. However, obtaining an accurate and ideal dataset can be challenging. The quality of the data is the most important aspect for any data analysis, and it is essential to pay special attention to analyzing and understanding the content and structure of the data. For this research, the dataset used was collected from the Federal Road Safety Corps database in Katsina, Nigeria.

The dataset represents the number of individuals killed, injured, and involved in traffic accidents, and contains information on the time, location, and incident type, including specific details about the traffic conditions at the time of the incident.

The data collected includes various features, such as the command, date, crash time, report time, arrival time, response time, route, location, vehicle number, vehicle type, vehicle category, vehicle model, name of organization/fleet operator, name of driver, driver's license number, probable cause(s) of the road traffic collision, number of injured and killed individuals, and number of individuals involved in the accident, broken down by gender and age group. The table 3.1. gives the detail of the data collected. The size of the dataset refers to the number of instances or rows it contains. Where the size here is 2322 and the time span covered by the dataset refers to the duration during which the incidents were recorded which is from January to December, 2022.

Table 3.1: Data Attribute and Description table.

<i>ATTRIBUTE</i>	<i>DESCRIPTION</i>
<i>Command</i>	Office of the FRSC in string and number
<i>Date</i>	Date when accident occur in string and number
<i>Crash time.</i>	Time of the accident in string and number
<i>Report time</i>	Time reported to the office in string and number
<i>Arrival time</i>	Time the reporter arrived at office in string and number
<i>Response time</i>	Time issue was responded to in string and number
<i>Route</i>	The path follower by driver in string
<i>Location</i>	The address where accident occurred in string
<i>Vehicle number</i>	The number of vehicle involved in number
<i>Vehicle type</i>	The type of vehicle involved in string and number
<i>Vehicle category</i>	The category of vehicle
<i>Vehicle model</i>	The model of vehicle number and string
<i>Name of organization</i>	The organization that own the vehicle in string

<i>Name of driver</i>	The drivers name in string
<i>DL number</i>	This is in Numeric
<i>Probable cause of RTC</i>	This is in string
<i>Number injured adult male</i>	This is in numeric
<i>Number injured adult female</i>	This is in numeric
<i>Number injured child male</i>	This is in numeric
<i>Number injured child female</i>	This is in numeric
<i>Number killed adult male</i>	This is in numeric
<i>Number killed adult female</i>	This is in numeric
<i>Number killed child male</i>	This is in numeric
<i>Number killed child female</i>	This is in numeric
<i>Number involved adult male</i>	This is in numeric
<i>Number involved adult female</i>	This is in numeric
<i>Number involved child male</i>	This is in numeric
<i>Number involved child female</i>	This is in numeric

3.2 Data Pre-processing

After data collection, an important step is data preprocessing, which involves using data mining techniques to remove unnecessary dimensions from the data. The goal is to improve data quality by removing noisy, incomplete, or irrelevant data. This process is crucial because inaccurate or irrelevant data can negatively impact the analysis. The following steps were taken in data preprocessing:

3.2.1 Data Cleaning

Data cleaning is an essential step in data preprocessing to eliminate noisy, incomplete, or irrelevant data that can adversely affect the analysis. In our research, we employed the following techniques for data cleaning:

- a) **Handling Missing Data:** We addressed missing data by replacing blank values with "NA" to indicate the absence of a value. This approach allows for clear identification and handling of missing data during subsequent analysis.
- b) **Dropping Columns with Missing Data:** To ensure data quality, we dropped columns that contained a large amount of missing data. These columns were deemed irrelevant for analysis due to the significant gaps in their values.
- c) **Changing Data Types:** In some cases, it was necessary to modify the data types of certain variables. Specifically, we converted integer values to strings when it was appropriate for their representation and analysis. This conversion ensured consistency and compatibility with subsequent data processing steps.

3.2.2 Feature Selection

To mitigate the negative impact of working with a large number of features and to prevent overfitting, this research work employed feature selection techniques to reduce dimensionality. In our research, we utilized the filter method for feature selection, which involves selecting only the most relevant features based on their statistical properties. This approach enables machine learning algorithms to train the dataset more efficiently and effectively.

3.3 Model Training and Evaluation

The following paragraphs outline the typical steps involved in model training and evaluation for a machine learning analysis:

- a) **Model Training:** The pre-processed dataset were split into training and testing subsets of 70% and 30% respectively. The training subset was used to train the machine learning models, which involved feeding the data into the models and optimizing their parameters using appropriate algorithms.

In the predictive analysis of road traffic accidents in Katsina State, Nigeria, the model training process involved the use of three different machine learning algorithms: Random Forest, Decision Tree, and K-Nearest Neighbors (KNN). Each algorithm has its own specific hyperparameters that were tuned during the training process. Here is a breakdown of the algorithms and their corresponding hyperparameters:

3.3.1 Random Forest: Random Forest

Random Forest is an ensemble learning technique that combines multiple decision trees to improve the accuracy and reduce the overfitting problem. In the case of predicting road traffic accidents, Random Forest can be used to identify the most important factors that contribute to accidents, and predict the severity of the accident based on the collective decision of multiple decision trees. Random Forest can handle both categorical and continuous data, and can handle missing data. Random Forest is computationally efficient and can handle high-dimensional data. Random forest is a popular machine learning algorithm that can be used for classification and regression tasks. It works by constructing a multitude of decision trees and then combining their predictions to make a final prediction. Each decision tree is trained on a random subset of the data, and at each node of the tree, a random subset of features is considered for splitting.

In the context of predicting road traffic accidents, random forest can be used to analyze various factors that contribute to accidents such as weather conditions, road surface, traffic volume, and time of day. By analyzing these factors, the algorithm can identify patterns and make predictions about the likelihood of an accident occurring in a particular location and time. To use random forest for predicting road traffic accidents, the first step is to gather relevant data on accidents and the factors that contribute to them. This data can then be preprocessed and used to train the algorithm. Once the algorithm is trained, it can be used to predict the likelihood of accidents in new locations based on the input factors. Overall, random forest is a powerful machine learning algorithm that can be used to predict road traffic accidents with high accuracy. By identifying patterns in accident data and analyzing relevant factors, it can help to reduce the number of accidents on the roads and improve overall safety.

Hyperparameters that were tuned:

n_estimators: The number of decision trees in the random forest.

max_depth: The maximum depth of each decision tree. It controls the complexity of the model and prevents overfitting.

min_samples_split: The minimum number of samples required to split an internal node during tree building.

min_samples_leaf: The minimum number of samples required to be at a leaf node.

3.3.2 Decision Tree

Decision tree is a machine learning algorithm that can be used for classification and regression tasks. It works by constructing a tree-like structure where each node represents a test on a specific feature and each branch represents the outcome of that test. The algorithm recursively splits the data into subsets based on the feature tests until a stopping criterion is reached. In the context of predicting road traffic accidents, decision tree can be used to analyze various factors that contribute to accidents such as weather conditions, road surface, traffic volume, and time of day. By analyzing these factors, the algorithm can identify patterns and make predictions about the likelihood of an accident occurring in a particular location and time.

To use decision tree for predicting road traffic accidents, the first step is to gather relevant data on accidents and the factors that contribute to them. This data can then be preprocessed and used to train the algorithm. Once the algorithm is trained, it can be used to predict the likelihood of accidents in new locations based on the input factors. Overall, decision tree is a simple and effective machine learning algorithm that can be used to predict road traffic accidents with high accuracy. By identifying patterns in accident data and analyzing relevant factors, it can help to reduce the number of accidents on the roads and improve overall safety.

Decision Tree is a simple yet powerful algorithm that predicts the target class by learning decision rules from the training data.

Hyperparameters that were tuned:

max_depth: The maximum depth of the decision tree.

min_samples_split: The minimum number of samples required to split an internal node.

min_samples_leaf: The minimum number of samples required to be at a leaf node.

criterion: The function used to measure the quality of a split. Common options include Gini impurity and information gain.

3.3.3 K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a popular machine learning algorithm that can be used for classification and regression tasks. It works by finding the k closest data points to a given input point and then using the majority class or the average value of the k neighbors as the prediction for the input point. In the context of predicting road traffic accidents, KNN can be used to analyze various factors that contribute to accidents such as weather conditions, road surface, traffic volume, and time of day. By analyzing these factors, the algorithm can identify similar patterns in the data and make predictions about the likelihood of an accident occurring in a particular location and time.

To use KNN for predicting road traffic accidents, the first step is to gather relevant data on accidents and the factors that contribute to them. This data can then be preprocessed and used to train the algorithm. Once the algorithm is trained, it can be used to predict the likelihood of accidents in new locations based on the input factors. Overall, KNN is a simple and effective machine learning algorithm that can be used to predict road traffic accidents with high accuracy. By identifying similar patterns in accident data and analyzing relevant factors, it can help to reduce the number of accidents on the roads and improve overall safety. However, it is important to choose an appropriate value of k, as choosing a value that is too small or too large can lead to overfitting or underfitting of the data.

K-Nearest Neighbors is a non-parametric algorithm that predicts the target class based on the majority vote of its nearest neighbors in the feature space.

Hyperparameters that were tuned:

n_neighbors: The number of neighbors to consider when making predictions.

weights: The weight function used in prediction. Options include 'uniform' (equal weights) and 'distance' (weights inversely proportional to the distance).

algorithm: The algorithm used to compute nearest neighbors. Options include 'brute', 'kd_tree', and 'ball_tree'.

During the model training process, the hyperparameters of each algorithm were optimized using cross-validation.

b) Model Evaluation: Once the models were trained, we evaluated their performance using appropriate evaluation metrics. These metrics include accuracy, precision, recall, F1 score and ROC curve these are presented under result and discussion section.

4.0 RESULT AND DISCUSSION

In this modelling, several graphs were used to explore relations inside the dataset. Column 'Total number injured' was chosen as target output the later used Total number Killed. Here we have taken different attributes for analysis and comparing the situations in which the graph in figure 4.1 to 4.4 were gotten. Figure 4.1 shows the graph of date against total number injured and figure 4.2 shows graph of date against total number killed.

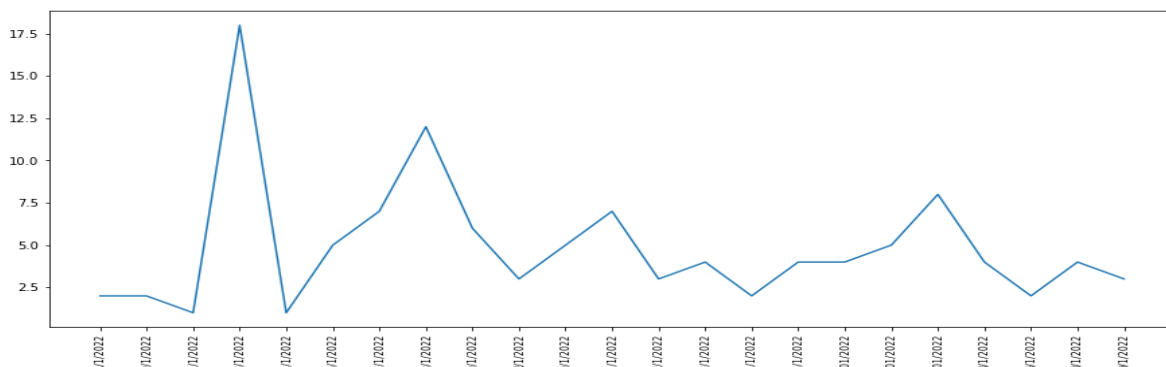


Figure 3.2: Graph of Date against Total number injured

From the graph it was discovered that the number of people injured as a result of the accident between 11th and 14th was very high and followed

by the accident that occurred between 15th and 21st.

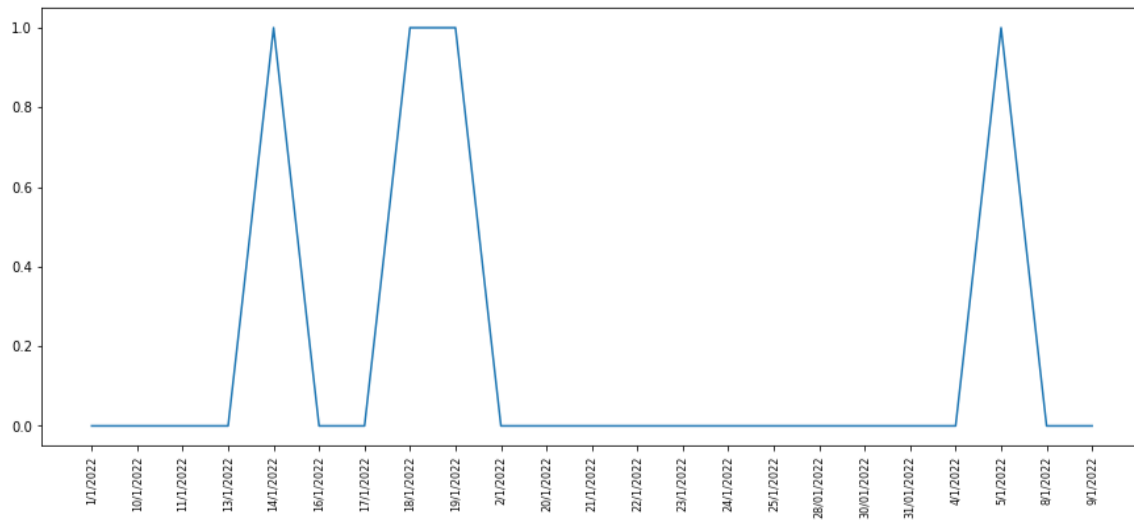


Figure 3.2: Graph of Date against Total number killed

From the graph it was discovered that the number of people killed as a result of the accident between 13th and 16th, 17th and 20th likewise 4th to 5th were very high.

figure 4.3 shows that people involved in the accident count 3.0 for 13th, 17th, 21st and 28th which are the highest, followed by 8th, 18th 19th, 25th and 31st which count 2.0 and every other days counts 1.0 which is not as high as others.

Comparing the total number of people involved in the road traffic accident with date, the graph in

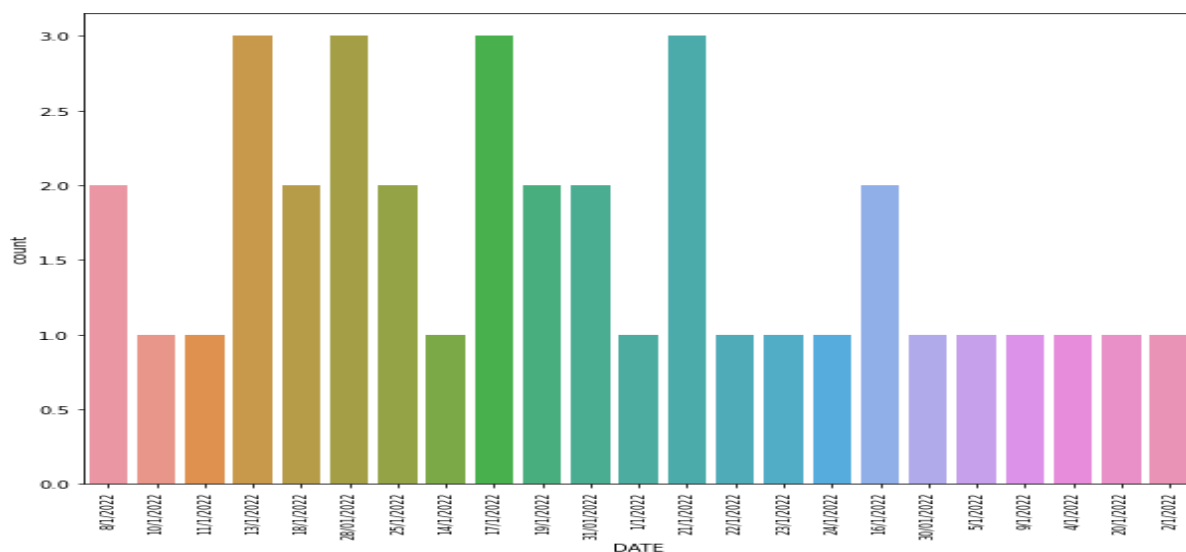


Figure 4.3: Graph of Number of those involved in accident against the Date.

In the modeling phase, three machine learning algorithms, namely Decision Tree, Random Forest, and K-Nearest Neighbors (KNN), were employed. The evaluation metrics used to assess the performance of these models were precision,

recall, F1-score, Accuracy and an ROC Curve. The results obtained from the modeling phase indicate the following performance measures for each algorithm:

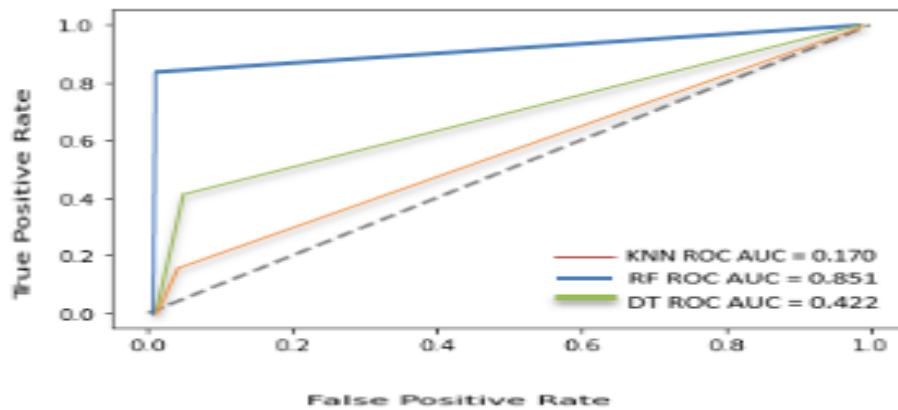


Figure 4.4: ROC Curve

For the ROC Curve, the Random Forest algorithm achieved a True Positive Rate (TPR) of 0.6353 and False Positive Rate (FPR) of 0.2059. These values indicate that the model has a relatively good balance between correctly identifying positive instances (TPR) while minimizing false positive predictions (FPR). This suggests that Random Forest performs well in distinguishing positive and negative instances. With the highest TPR among the three algorithms, Random Forest demonstrates a strong ability to capture positive instances accurately. The Decision Tree algorithm achieved a TPR of 0.3333 and an FPR of 0.25. The relatively low TPR indicates that the model struggles to correctly identify a significant number of positive instances. The FPR value is comparable to Random Forest, implying a similar ability to minimize false positive predictions. However, the

lower TPR leads to a lower overall performance, as evidenced by the lower area under the ROC curve. And the KNN algorithm yielded a TPR of 0.5763 and an FPR of 0.2211. These values suggest a moderate performance in correctly identifying positive instances compared to the other two algorithms. The FPR value is slightly lower than that of Random Forest, indicating a relatively better ability to avoid false positive predictions. However, the lower TPR and overall lower performance, as indicated by the lower area under the ROC curve, suggest that KNN lags behind Random Forest in terms of predictive capacity.

After these, the result gotten from the model (Decision tree, Randomforest and K-Nearest Neighbour) using Precision, Recall and F1-score, Accuracy is given in the table 4.1:

Table 4.1: Prediction Result

	RANDOM FOREST	DECISION TREE	K-NEAREST NEIGHBOUR
PRECISION	1.00	1.00	0.29
RECALL	0.71	0.50	0.67
F1-SCORE	0.83	0.67	0.40
ACCURACY	0.71	0.50	0.67
PREDICTION RESULT	85%	42%	17%

From table 4.1, Random Forest: The Random Forest algorithm achieved a precision of 1.00, recall of 0.71, F1-score of 0.83 and Accuracy of 0.71. These metrics suggest that the model has a high precision in correctly identifying positive Forest demonstrated the highest predictive capacity among the three algorithms.

Decision Tree: The Decision Tree algorithm achieved a precision of 1.00, recall of 0.50, F1-score of 0.67 and Accuracy of 0.50. Although the precision is perfect, the relatively low recall indicates that the model struggles to correctly identify a significant number of positive instances. Consequently, the F1-score is lower compared to Random Forest. The predictive capacity of Decision Tree, resulting in 42%, is lower than that of Random Forest.

K-Nearest Neighbors (KNN): The KNN algorithm yielded precision, recall, and F1-score, Accuracy values of 0.29, 0.67, and 0.40, 0.67 respectively. These metrics indicate that the model has a lower precision in correctly identifying positive instances compared to the other two algorithms. The F1-score suggests a relatively moderate performance. The predictive capacity of KNN, resulting in 17%, is the lowest among the three algorithms.

5.0 CONCLUSION

The study on predicting road traffic accidents in Katsina State using machine learning has demonstrated that machine learning algorithms can predict the likelihood of road traffic accidents and identify contributing factors. Various machine learning algorithms, such as Decision trees, Random Forest, and K-Nearest Neighbour, were used to analyze the factors. Data collection and pre-processing are critical steps in machine learning algorithms, and the quality of the data is crucial for data analysis. The study has explored the performance of different machine learning algorithms and identified the most important variables that influence the occurrence of road traffic accidents. By integrating other data Rahman M. S, Abdel-Aty M, Hasan S, Cai Q. (2019) Applying machine learning approaches to analyze the vulnerable road-users' crashes at

instances, with a relatively good balance between precision and recall. The F1-score, which considers both precision and recall, provides an overall measure of the model's performance. With a result of 85%, Random sources, such as traffic flow data and social media data, stakeholders can develop targeted interventions to prevent accidents and save lives and Random forest shown the best performance which is 85%. The future research should address potential sources of bias that could affect the results and also improve the performance.

REFERENCES

- Annie R., Srihari P, A, Meena P., (2020) "Prediction of Road Accident Severity Using Machine Learning Algorithm", ISSN: IJAST-2020.
- Hossain, M., Abdel-Aty, M.A., Quddus, M.A., Muromachi, Y., and Sadeek, S.N. (2019) Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements. *Accident; analysis and prevention*, 2019. 124, 66-84. <https://DOI:10.1016/j.aap.2018.12.022>.
- Huang, H., Zeng, Q., Pei, X., Wong, S.C., Xu, P., (2016). Predicting crash frequency using an optimised radial basis function neural network model. *Transportmetrica: Transport. Sci.* 12 (4), 330–345. <https://DOI:10.1080/23249935.2015.1136008>
- Kumar S, Toshniwal D, (2016) A data mining approach to characterize road accident locations. *J Mod Transp* 2016. 24(1):62–72 28. <https://DOI:10.1007/s40534-016-0095-5>
- Kwon O. H, Rhee W, Yoon Y, (2025), Application of classification algorithms for analysis of road safety risk factor dependencies. *Accid Anal Prev*, 75:1–15. <https://DOI:10.1016/j.aap.2014.11.005>
- Labib, M.F., Rifat, A.S., Hossain, M.S., Das, A., and Nawrine, F, (2019). Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh. 2019 7th International Conference on Smart Computing & Communications (ICSCC), 1-5. <https://DOI:10.1109/ICSCC.2019.8843640>
- statewide traffic analysis zones. *J Safety Res.* 2019 Sep; 70:275-288. <https://doi:10.1016/j.jsr.2019.04.008>.

- Sameen M. I., Pradhan B, (2017). Severity prediction of trafıc accidents with recurrent neural networks. *Appl Sci* 2017. 7(6):476. <https://doi.org/10.3390/app7060476>
- Sharma B, Katiyar VK, Kumar K (2016) Trafıc accident prediction model using support vector machines with Gaussian kernel. In: *Proceedings of ffth international conference on soft computing for problem solving*, Springer, Berlin, pp 1–10
- Wahab L, Jiang H,Severity (2019). Prediction of motorcycle crashes with machine learning methods. *Int JCrashworthiness..* 24:1–8. <https://DOI:10.1080/13588265.2019.1616885>.
- Zheng, M., Li, T., Zhu, R., Chen, J., Ma, Z., Tang, M., Cui, Z., Wang, A.Z., (2019). Traffic accident’s severity prediction: a deep-learning approach-based CNN network. *IEEE Acc.* 7, 39897–39910. <https://DOI:10.1109/ACCESS.2019.2903319>