Okoh Kate Imanyi, 2024, 12:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Accident Prediction System Using Decision Tree Algorithm

Okoh Kate Imanyi, Agaji Iorshase
Department of Computer Science,
Joseph Sarwuan Tarkaa University, Makurdi, Benue State, Nigeria

Abstract- Road accidents cause enormous human and financial losses all around the world. This research presents an Accident Prediction System (APS) that predicts the probability of traffic accidents by applying the Decision Trees algorithm. To build decision trees for prediction, the APS incorporates weather, traffic patterns, road infrastructure, and past accident data. A dataset containing numerous accident-related characteristics, such as location, vehicle speed, road type, time of day, and environmental factors, is used to train the system by applying the decision trees algorithm. Performance evaluation of this system shows promising accuracy in accident prediction through rigorous evaluation and validation processes. In addition to identifying high-risk locations that are prone to accidents, the APS also helps develop preventative strategies for traffic control and accident avoidance. By giving stakeholders, such as traffic authorities, urban planners, and law enforcement agencies, a trustworthy forecasting tool to lessen the frequency and severity of traffic incidents, this research helps to improve road safety.

Keywords- Accident, Decision Trees, Prediction and Transportation.

I. INTRODUCTION

Transportation as a means of conveying man, goods and services from one point to another has been a breakthrough in the existence of mankind and has several means such as road, air and sea. Road transport is the commonly used means in Nigeria because it is fairly cheap and can be easily accessed. The restrictive nature of the water ways, coupled with the near collapse of the rail system, and the high cost of air travels have further exerted a lot of pressure on the road, as over 75 percent of the total movements in the country are made by road (Chidoka, 2009). The Economic Times (2018) defined road transport as the transportation of goods and personnel from one place to the other on roads. It involves the use of motor cars, lorries, buses, motor cycles, bicycles, trucks and even animals. Road transport compared to other modes of transportation is more flexible, no specialized machinery or technique is necessary, hence all countries no matter how backward, possess them in some form. In primitive regions, they may be simple parts suitable only for foot or hoof traffic, but in advance regions, they are more likely to be well made and suitable for heaviest and the most modern vehicles (Etim, 2018).

Road transport system is important to our day-to-day activities. It provides people with access to workplaces and education facilities, shops, and social, recreational, community and medical facilities (New Zealand Ministry of Transport, 2014). Road transport is one of the important employments generating sectors especially in the rural areas (Srija, 2011). Youth who are not formally employed have engaged themselves as taxi drivers

© 2024 Okoh Kate Imanyi. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

which serve as a means of livelihood. Road transport has provided connections between businesses and their input sources, between businesses and other businesses, and between businesses and their markets (New Zealand Ministry of Transport, 2014). It has always played an important role in influencing the formation of urban societies (Mathew and Rao, 2007).

Road transport has also contributed immensely to the economic development of the country. It is an important component of the economy and a common tool used for development (Rodrigue and Notteboom, 2018). It increases the geographical and occupational specialization as well as interdependence of countries based on the principle of comparative advantage (Wilson and Felix, 2015).

In spite of all the benefits that are inherent in the road transport, road traffic accident is one menace that has devastating effect on this mode of transportation. Ohakwe et al. (2011) defined Road Traffic Accident (RTA) as the collision of a vehicle with another vehicle, pedestrian, animal or geographical or architectural obstacle. Road accidents are among the major causes of death, injury and disability all over the world both in developed and developing countries. With a broad estimation, in every 1 min, 2 people are killed and 95 people are severely injured or permanently disabled in traffic accidents worldwide (Ali et al., 2009). The rate of fatalities, disabilities and injuries caused by road traffic accidents has become an economic, mental as well as physical burden to its victims. Many families are driven into poverty by the cost of prolonged medical care, the loss of a family breadwinner or the extra funds needed to care for people with disabilities (Mathers and Loncar, 2013).

Data published by the World Health Organization and the European Union demonstrated that the main cause of death is attributed to road accidents which numbered 1.2 million fatalities per year-round the world. According to National Bureau of statistics (2016), Nigeria recorded 11,363 road accidents of which a total of 30,105 people got injured, 5053 Nigerians got killed during the year. Nigeria with a population rate of 193 million people, and a total road length of 194,000km and total vehicle population of 10 million vehicles (as cited by FRSC vehicle inspection and certification head), this translated to 995 persons per kilometer of road. This results to extreme pressure of on the road traffic and increases the risk of accidents. These figures give credit to World Health Organization reports which ranked Nigeria second-highest in the rate of road accidents among 193 countries of the world.

Benue state is located in the central zone of Nigeria with a population of 4,219,244 (NPC, 2006) and total land area of about 33,955 square kilometers. Access to all parts of the state is through any of the following Federal Government trunk A roads: Wukari to Katsina-Ala to Vandeikya and Jos to Makurdi, Gboko, Otukpo, and Otukpa, trunk B roads that link all local Government headquarters as well as laterite roads that link various agricultural communities. This attracts purchase of farm produce from other parts of the country thus, increasing traffic on the high way. Benue state is identified as one of the areas with highest accident burdens in Nigeria of which the rate of death from road accident is on the rise and thus an issue of major concern. In Nigeria, the Federal Road Safety Corps has been mandated as the lead and coordinating agency for road safety management. This status has conferred on the agency the responsibility of playing its role as the key driver of all road safety efforts in Nigeria. While this has helped to achieve some safety on the roads, playing this role has been a herculean task. In order to efficiently and effectively capture required information so as to monitor the state of road traffic accidents in Nigeria, there is a need for a system that can be used to collect, store, analyse and predict the likelihood of road traffic accident.

It is in view of the above that this research is proposing an accident prediction system to predict future occurrence of accidents in Benue State based on the previous accident data in the State. This system uses data mining technique called Decision Tree to build an accident prediction system that will predict the likelihood of accidents occurrence and help the road safety stakeholders to put adequate safety measures in place to avert such occurrences. A decision tree is a graphical representation of possible solutions to a decision based on certain conditions (Khurana, 2017). It's called a decision tree because it starts with a single box (or root), which then branches off into a number of solutions, just like a tree. Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes (Sayad, 2018).

1. Statement of Problem

There is no known road accident prediction system with focus on road accidents in Benue State using Decision Tree. However, several research have used other data mining techniques such as neural networks to build road accident prediction system. Road transport is of high important to the people of Benue State because most of the agricultural produce in the rural area are transported to the urban area via road.

The non-availability of road accident prediction system in Benue State has posed serious challenges. Several lives and properties worth millions of naira have been lost due to the menace of road accident. According to the Federal Road Safety Commission (2016), in 2016, Benue State recorded over 183 road accidents within the first six months of the year, ranking it among the highest in the country. The figure places Benue State sixth in the list of states with cases of road crashes across the country. The system will use the risk factors contributing to the previous road accidents as input data. The output will provide useful patterns and trends that will serve as guide to road transport safety stakeholders in Benue State in making critical decisions about the safety of Benue roads.

2. Aim and Objectives of the Proposed System

The aim of this work is to develop a road accident prediction system for Benue State using Decision Tree. The objectives of the study are to:

- Collect, clean and prepare previous road accidents data for prediction
- Design the system using Use Case Diagram and Flow Chart
- Use data mining tool called WEKA to build and implement the system
- Test and validate the Prediction accuracy of the system using test dataset
- Interpret and analyze the result of the prediction

3. Significance of the Study

In Nigeria today, hardly a day goes by without the occurrence of a road traffic accident leading to generally increasing incidence of morbidity and mortality rates as well as financial cost to both society and the individual involved (Agbonkhese et al., 2013); of which Benue State is not left out. This proposed road accident prediction system will be useful to the road management agencies in understanding the patterns of road accidents in Benue State. It will help identify the major factors that are responsible for road accidents in the State. The system will be able to predict future occurrence of road accidents and help the stakeholders to take necessary decision to avert the occurrence. It will help reduce the rate of death as a result of accident in the State to the barest minimum. It will improve the safety of agricultural products on roads and ensure efficient delivery from one point to the other. The system will improve the economy of the State and provide speedy access to goods and services.

II. LITERATURE REVIEW

According to Marker (2017), Prediction in data mining is to identify data points purely on the description of another related data value. It is not necessarily related to future events but the used variables are unknown. Prediction derives the relationship between a thing you know and a thing you need to predict for future reference. Prediction makes it possible to identify best courses of action without requiring understanding of the underlying mechanisms (Bzdok, 2018).

Amorim et al, (2023) used a Machine Learning approach to classify road accident hotspots in Brazil. The model worked by predicting the risk areas and alert drivers. Advanced research was used to carry out and identify accident-influencing factors and potential highway risk areas to mitigate the number of road accidents. Machine learning techniques were used to build the prediction model using a supervised classification based on a labeled dataset. The research work experimented with many machine learning algorithms to discover the best classifier for the Brazilian federal road hotspots associated with severe or non-severe accident risk using several features. The model was developed with SVM, random forest, and a multi-layer perceptron neural network. The dataset used to develop the model contains a ten-year road accident report by the Brazilian Federal Highway Police. The input variables include features such as spatial footprint, weekday and time when the accident happened, road type, route, orientation, weather conditions, and accident type. The results of the model were promising; however, the neural network model provided the best results, achieving an accuracy of 83%, a precision of 84%, a recall of 83%, and an F1-score of 82%.

Nagesh, et al. (2021) used machine learning techniques to analyze and Predict Road Accident. The study identified road accidents as one of the most relevant causes of injuries and death, and also one of the serious issues, which can possibly cause disabilities, injuries and even fatalities. The study used Logistic Regression Analysis to estimate the relationships among variables. The packages which played a major role in the analysis are pandas and numpy. The Pandas was used for data manipulation and analysis. The datasets used in the study were obtained from kaggle and other sources. Some of the attributes in the dataset include area, alarm type, visibility, ecarttime, weather condition, accident severities and pothole severities. Alarm type dataset collected by CAS (collision on avoidance system) device for particular area, time, weather condition and visibility was also used in the process. The model used unsupervised learning approach to predict the results of road accident prediction zone at the end last column that's in high/low using k-means clustering technique.

Ahmed, et al. (2023) developed a road accident prediction and contributing factors system using explainable machine learning models. The paper evaluated a set of machine learning (ML) models to predict road accident severity based on the most recent New Zealand Road accident dataset. The predicted results were analyzed and an explainable ML (XML) technique was applied to evaluate the importance of road accident contributing factors. The system worked by predicting road accidents with different injury severity. The work considered different ensembles of ML models, like Random Forest (RF), Decision Jungle (DJ), Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (L-GBM), and Categorical Boosting (CatBoost). New Zealand Road accident data from 2016 through 2020 obtained from the New Zealand Ministry of Transport was used for the study. The comparison results show that RF is the best classifier with 81.45% accuracy, 81.68% precision, 81.42% recall, and 81.04% of F1-Score. The study also employed the Shapley value analysis as an XML technique to interpret the RF model performance at global and local levels. While the global level explanation provides the rank of the features' contribution to severity classification, the local one is for exploring the use of features in the model. Furthermore, the Shapley Additive exPlanation (SHAP) dependence plot was used to investigate the relationship and interaction of the features towards the target variable prediction. Based on the findings, it can be said that the road category and number of vehicles involved in an accident significantly impact injury severity. The identified high-ranked features through SHAP analysis were used to retrain the ML models and measure their performance. The result shows 6%, 5%, and 8%, increase, respectively, in the performances of DJ, AdaBoost, and CatBoost models.

Gururaj, et al. (2022) developed models for predicting traffic accidents and their injury severities using machine learning techniques The study established models to select a set of influential factors and to build up a model for classifying the severity of injuries. The system worked by applying machine learning to model and predict the severity of injury that occurs during road accidents. One such way is to apply unsupervised learning models such as Apriori, Apriori TID (transaction id), SFIT (set operation for frequent itemset using transaction database) and ECLAT (equivalence class clustering and bottom-up lattice traversal) which analyze the unlabeled traffic accidents dataset and determine the relationship between traffic accidents and injury. This research work was helpful for traffic departments to decrease the number of accidents and to distinguish the injury's seriousness extensive simulations were carried out to demonstrate the unsupervised learning algorithms for predicting the injury severity of traffic accidents. The results showed that Apriori algorithm predicted the patterns in 962 milliseconds, Apriori TID (transaction id) algorithm predicted the pattern in 557 milliseconds, SFIT algorithm predicted the pattern in 516 milliseconds and ECLAT algorithm predicted the pattern in 124 milliseconds. ECLAT algorithm took less time compared to all the other algorithms.

Amuche et al., (2017) used neural networks to develop a road accident predictive system for Benue State as function of road characteristics. The system was design using Multi-Layer Perceptron (MLS). Back propagation algorithm was used to create and train the network to predict the number of accidents that will occur in Benue State and whether the accidents will be fatal, serious or minor in any given year. Also, the neural network model was used to determine the number of persons killed, and number of persons injured in any giving year. The dataset used to create and trained the network was data of about 8376 road accidents that occurred in Benue State within six years which was obtained from the FRSC, Benue State Command. The number of predicted data after testing was 8305 which is proportional to an accuracy of 99.15%. This shows that Artificial Neural Network is an efficient model for prediction.

Aci and Ozden (2018) used K-Nearest neighbor and other machine learning algorithms to predict the Severity of Motor Vehicle Accident Injuries in Adana-Turkey. The main objective of the study was to determine how important weather and other phenomena are for the occurrence of traffic accidents. The model worked by predicting the severity of the accident whether the accident will be fatal or non-fatal. The traffic accident data used for the model was obtained from the reports kept by Regional Traffic Division and the weather data provided by the Regional Directorate of Meteorology during years 2005-2015. Five different major machine learning methods (k-Nearest Neighbor, Naive Bayes, Multilayer Perceptron, Decision Tree, Support Vector Machine) and one statistical method, Logistic Regression, were employed to develop prediction models based on the dataset. The Decision Tree, k-Nearest Neighbor, and Multilayer Perceptron based models yielded higher accuracy in classification of the accidents dataset compared to other models.

Arafeh et al.,(2010) developed a linear regression model approach that can be applied to crash data to predict vehicle crashes. The proposed approach involved novice data aggregation to satisfy linear regression assumptions; namely error structure normality and homoscedasticity. The model was created, tested and validated using dataset from database of crash, traffic, and geometric data obtained from 186 access road sections in the state of Virginia. The model worked by predicting the crash rate per unit distance for a period of time. The regression model was applied to the data to derive the model intercept and slope. The optimum slope of the line was significant ($p << 0.0005$) with a value of -4.135. Similar models were developed using the traditional negative binomial and zero inflated negative binomial. Although the Poisson model produced a slope that is closest to 1.0 (0.549), however, the Sum of Square Error (SSE) was 20% which was higher than that of Linear Regression Model's SSE. Consequently, the proposed Linear Regression Model's model appeared to offer the best compromise in terms of SSE and model prediction.

III. METHODOLOGY

Agile methodology is adopted in this work with specific implementation of the Scrum method which is a subset of Agile for the design, development and implementation of the proposed system. This is because of the iterative and incremental nature of the methodology. Scrum is a lightweight process framework for agile development and it is different from other agile processes separating concepts and practices into categories namely roles, artifacts and time. It is a framework within which people can address complex adaptive problems, while productively and creatively delivering products of the highest possible value (Schwaber and Sutherland, 2017). The use of scrum methodology in this research work will increase quality of the research outcome (product) and provide better development schedules and management of various phases of the research work. According to Littlefield(2016), Scrum is made of the Product Owners which are the customers or end users of the product, the Backlog which is the list of tasks to be accomplished in the project, the Sprint which is the predetermined time frame within which the team completes sets of tasks from the Backlog, the Daily Scrum Meeting which give the project team opportunity to give project update, and the Retrospective where the team reviews their work and discusses ways to improve the next Sprint.

The breakdown of the Scrum framework illustrates the key activities involved in this research work. The customers of this research work are the road transport stakeholders in Benue State. The list of tasks represents the activities involved in the various phases of this research work such as data collection and preprocessing, data classification and model interpretation, and display of the model results on the web interface. The Sprint provides opportunity to allocate time to the various phases of this research work. The Daily Scrum Meeting offers the need to analyze the progress made so far and to examine each step taken to accomplish the various task. The Retrospect enables us to review the entire research and decide the way forward.

The system uses the J48 decision tree algorithm. According to Kaur and Chabra (2014) J48 is an extension of ID3 decision tree algorithm. The additional features of J48 are accounting for missing values, decision trees pruning, continuous attribute value ranges, derivation of rules, etc. The J48 algorithm was used to create the model that will predict whether an accident will be Fatal or Serious. About 500 accident records from 2015 to 2018 in Benue State were used to train and test the model. The input variables to the model consist of the major factors that are responsible for the accidents in the record. These factors are Human Factor, Road Condition, Head on Collision, Environmental Factor, Mechanical Failure and Improper Maintenance.

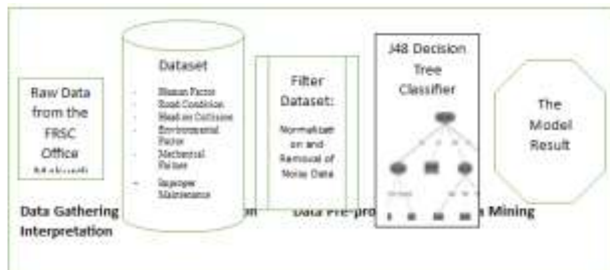

Figure 1: The Model Architecture

The architecture of the model consists of data gathering section. In this section, the raw data are collected from the accident reports in the FRSC office, these reports were in form of files with each report file in a separate folder. Some of the data extracted from these accident files include the accident contributing factor(s) or cause(s) of the accidents which serve as input variables to the system, the year of the accidents and the severity of the accidents which serves as the output variables to the system. The dataset was extracted into excel file and converted to CSV format using Microsoft Excel 2007; the WEKA feature selection features was used to normalized and preprocessed the converted dataset to remove unwanted features; and the preprocessed dataset was fed into the WEKA J48 decision tree classifier to train and test the model. The predicted results of the model will be displayed and interpreted. The classification accuracy and the confusion matrix of the classifier will be used to determine the accuracy of the model. Figure 1.0 shows the architecture of the model.

1. The Model J48 Algorithm

According to Gupta et al. (2012), a decision tree is a flow-chart-like tree structure. The internal node denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distribution. The top most node in a tree shown by oval is a root node.

J48 is a tree-based learning approach. It is developed by Ross Quinlan which is based on iterative dichtomiser (ID3) algorithm. J48 uses divide-and-conquer algorithm to split a root node into a subset of two partitions till leaf node (target node) occur in tree. Given a set T of total instances the following steps are used to construct the tree structure.

Basic Steps in the Algorithm

Kaur and Chhabra (2014) identified the basic steps of the J48 algorithm as follows:

- In case the instances belong to the same class the tree represents a leaf so the leaf is returned by labeling with the same class.
- The potential information is calculated for every attribute, given by a test on the attribute. Then the gain in information is calculated that would result from a test on the attribute.
- Then the best attribute is found on the basis of the present selection criterion and that attribute selected for branching.

Mathematical Expression of J48 Algorithm

According to Suleiman et al. (2015), J48 handles both categorical and continuous attributes to build a decision tree. In order to handle continuous attributes, J48 splits the attribute values into two partitions based on the selected threshold such that all the values above the threshold as one child and the remaining as another child. It also handles missing attribute values. J48 uses Gain Ratio as an attribute selection measure to build a decision tree. It removes the biasness of information gain when there are many outcome values of an attribute. At first, calculate the gain ratio of each attribute. The root node will be the attribute whose gain ratio is maximum. J48 uses pessimistic pruning to remove unnecessary branches in the decision tree to improve the accuracy of classification as shown in equation (1).

$$Gain(S,A) = Entropy(S) - \sum_{v \epsilon}^{a} values\ (A) \frac{|S_v|}{|S|} Entropy|S_v| \quad (1)$$

Where Values (A) is the set of all possible values for attribute A, and Sv is the subset of S for which attribute A has value v (i.e.,

$$Sv = \{s \epsilon S \mid A(s) = v\}).$$

The first term in the equation for Gain is just the entropy of the original collection S and the second term is the expected value of the entropy after S is partitioned using attribute A. The expected entropy described by this second term is simply the sum of the entropies of each subset, weighted by the fraction of examples that belong to Gain (S, A) is therefore the expected reduction in entropy caused by knowing the value of attribute A.

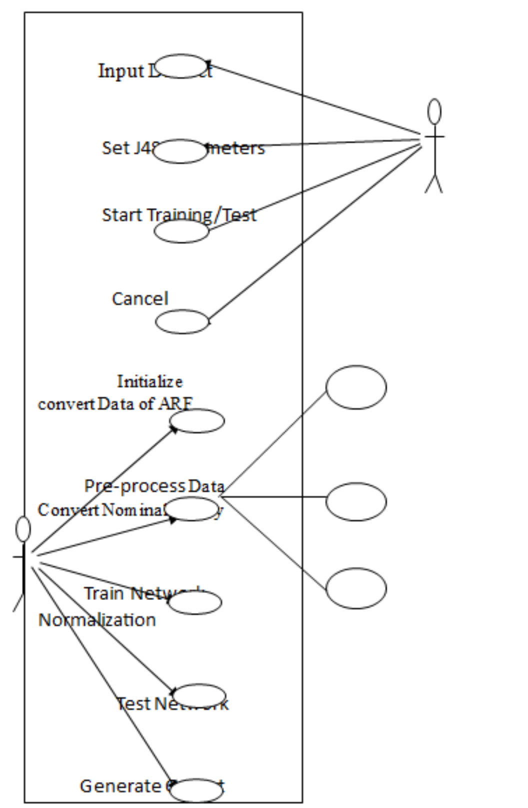

Figure 2: Use Case Diagram of the System

2. The Use Case Diagram of the system

The Use Case Diagram provides the functional description of the decision tree accident Predictive Model and its major processes. It provides a graphical description of the users of the model and what kinds of interactions to expect within the model. The Use Case Diagram for this model is as shown in Figure 2.0. It consists of two actors and their respective use Cases. These are the User (researcher) and the System.

3. Program Algorithm of the model

The detailed algorithm of the system is as shown figure 3.0. The system accepts data in the form of CSV format. The dataset is processed and normalized using WEKA to remove unwanted and noisy instances. The preprocessed data is then fed into the J48 decision tree algorithm for classification. The J48 classification parameters were turned several times to obtain the desired output with higher classification accuracy. The system generates the actual output of leave nodes and the error after each iteration. If the classification accuracy of the system is not accepted, the J48 algorithm is tuned and the classification restarted after reaching maximum iteration.

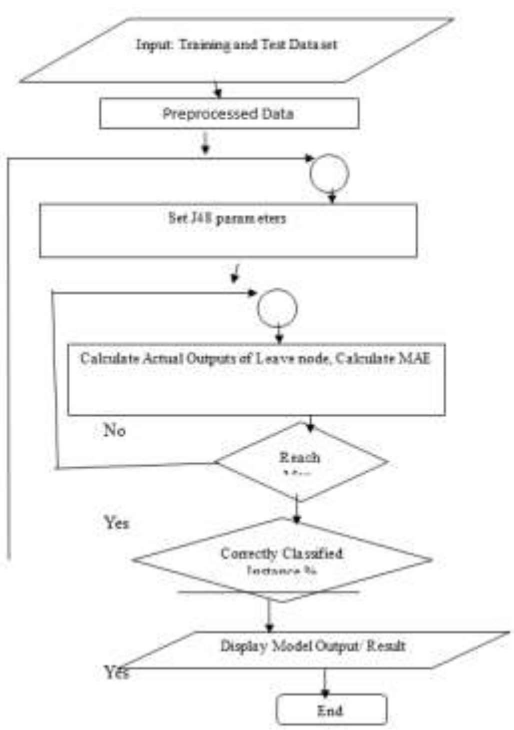

Figure 3: The Model Flow Chart

4. Assumptions of the Model

- J48 algorithm works by creating a tree to evaluate an instance of data
- While building a tree, J48 ignores the missing values that is the value for that item can be predicted based on what is known about the attribute values for the other records
- When classifying in a decision tree, there are two possible outcomes on a tree node: positive and negative
- The predictions are the leaves of the tree.
- The paths from root to leaf represent classification rules.

IV. SYSTEM IMPLEMENTATION

The software tools used for the development and implementation of this Road Accident Predictive Model for Benue State are WEKA and MS Excel
WEKA 3.8.1: WEKA was used to design the process architecture of the model, preprocessed the dataset and classify the dataset to obtain the classification accuracy of the model. The WEKA's time series forecast feature was also used to forecast the values for each of the classes from year 2019 to 2028 based on the predicted values obtain from the WEKA classifier.

Microsoft excel package was used to extract the dataset and convert it to comma delimited file format (CSV).

1. Data Extraction and Preparation

The 499 previous road accidents data used in the development of the model were obtained from the Benue State's Federal Road Safety Corps (FRSC) Sector Command in Makurdi. These were reports of road accidents that occurred between the year 2015 and 2018 in Benue State. The accident reports were produced in hard copy and stored as classified documents in the Command. The reports contain the cause(s) of each of the road accidents, the severity of the accident, date and time of the accident, the name and type of vehicle, motorcycle or tri-cycle that was involved in the crash. These files were obtained and the required information extracted into Microsoft Excel file. The causes of the accident, the year of the accident and the nature or severity of the accidents formed the data input to the model. The dataset was saved in CSV (comma delimited) file format using MS-Excel. The six major causes of the road accidents identified in the dataset are Human Factor, Road Condition, Head-on Collision, Environmental Factor, Mechanical Failure and Improper Maintenance which served as the input variables while Severity of the road accidents served as the output variable. The road accidents severities were classified into Fatal Accidents, Serious Accidents and Minor Accidents. The fatal accidents were those that involves loss of lives and vehicle(s) damaged beyond repair, the serious accidents were the ones that the motorists sustained severe injuries and the vehicle substantially damaged and the minor accidents were the ones the motorists sustained minor injuries and the vehicles slightly damaged. Table 1 shows the actual values for the three classes; Fatal, Serious and Minor accidents and the total road accidents for each of the years from 2015 to 2018.

Table 1: Actual Values of the accident's datasets from 2015 to 2018.

Year	Fatal	Serious	Minor	Total
2015	16	109	1	126
2016	18	99	0	117
2017	35	100	2	137
2018	19	91	9	119
TOTAL	88	399	12	499

2. Implementation of the Model

After the extraction and preparation of the dataset, the dataset was fed into the WEKA J48 decision algorithm (Classifier), the confidence factor used for pruning the tree was set to 0.25, the minimum number of instances per leaf was set to 2, and the seed used for randomizing the data when reduce error pruning was set to 1, the WEKA's percentage split was used to split the dataset in appendix 3 into 75% training dataset and 25% test dataset. The J48 decision tree algorithm was used to classify the dataset and make prediction. The J48 parameters were tuned several times to improve the prediction accuracy of the model. The prediction accuracy obtained after training and testing was 82.4%. The loss or mean absolute error was 0.2135. The prediction of the model was based on the classification accuracy of the model result. Higher classification accuracy means good prediction while lower classification accuracy means poor prediction.

The predicted values for the Fatal class, Serious class and Minor class were obtained using the model prediction accuracy as shown in table 2.

Table 2: Predicted Value from the J48 Decision Tree Classifier

Year	Fatal	Serious	Minor	Total
2015	13	89	1	103
2016	15	82	0	96
2017	29	82	2	113
2018	16	75	7	98
TOTAL	73	328	10	410

In order to obtained the future values of the Fatal class, Serious class and Minor class, the WEKA's time series forecast feature was used to forecast the values for each of the classes from the year 2019 to 2028 based on the predicted values obtain from the WEKA J48 classifier.

The WEKA's time series forecast parameters were tuned several times to obtain predictions in the range of the actual value. The learning rate was set to 0.3, the momentum was set to 0.2, the number of iterations was set to 100 and the seed was set to 1. The predicted results obtained for the years 2019 to 2028 is as shown in table 3.0.

Table 3: Predicted Result from the WEKA Times Series Feature

Year	Fatal	Serious	Minor	Total
2019	6	68	10	84
2020	32	71	7	110
2021	34	79	3	116
2022	16	68	10	94
2023	9	67	10	86
2024	33	68	9	110
2025	34	70	9	113
2026	22	68	10	100
2027	22	67	10	99
2028	33	68	10	111
TOTAL	241	694	88	1023

3. Results of the Model

The road accident predictive model is a decision tree model developed for the purpose of predicting the number of road accidents that will occur in Benue State in a particular year based on the previous road accident records. The model combines different data mining techniques such as data extraction and conversion, Decision Tree J48 classification algorithm, and Time Series Forecasting to perform the task. These data mining techniques deal mainly with data extraction, data preprocessing, data classification and forecasting.

Figure 4.1 shows the results of the classifier which contains the classification accuracy, the mean absolute error or loss based on the dataset in appendix 3. In this result, the classification accuracy after the final epoch was 82.4% for training dataset and test dataset. The error was 0.2135 for training dataset and test dataset. The full result is as shown in appendix 1 and the HTML code of the result generated by WEKA is as shown in appendix.

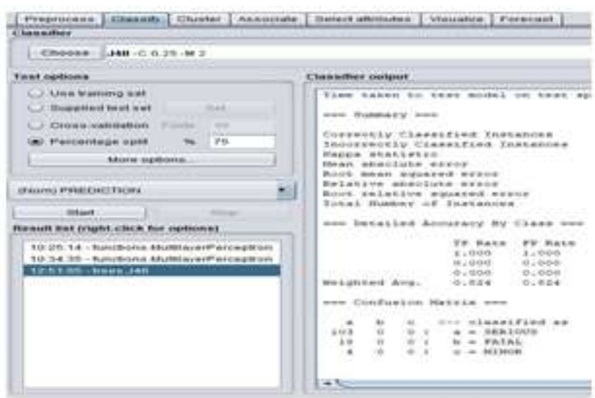

Figure 4: Result of the J48 Decision Tree Model

Having obtained the predicted values for the years 2015 to 2018 using the J48 decision tree classifier's classification accuracy, the output is used for each of the predicted classes (Fatal, Serious and Minor accidents) in table 4.2 to predict the future values for the years 2019 to 2028. In order to obtain this, the dataset in table 4.2 was loaded as input into the WEKA time series forecasting algorithm to obtain the future values of each of the classes. The future forecast of each of the output classes Fatal, Serious and Minor accidents are then obtained using the WEKA time series forecast feature as shown in figure 4.2

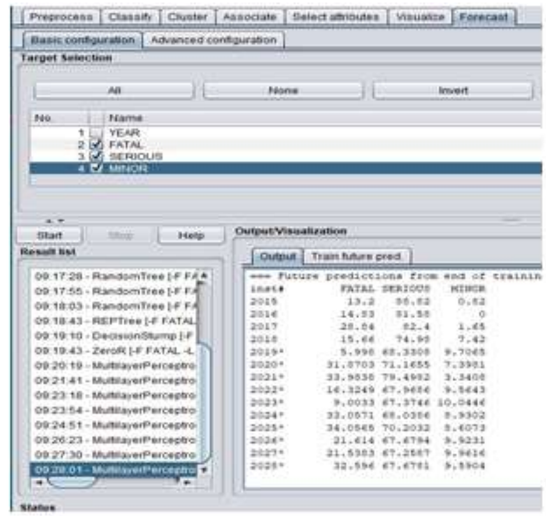

Figure 5: Predicted Values for the Fatal, Serious, and Minor Accidents for the Years 2019 to 2028.

Figure 4.3 shows the graph for the predicted values of Fatal, Serious and Minor accidents for each year which was plotted using the WEKA time series forecast features. This graph was created using the data in table 4.2 above. The squares in the graph show the predicted values obtained using the J48 decision tree model (classifier) based on the model classification accuracy while the oval shapes represent the values obtained from the WEKA time series forecast technique. The square represents predictions from 2015-2018 while the oval represents predictions from 2019-2028. The red graph represents the fatal accident class, the blue graph represents the serious accident class and the green graph represent the minor accident class. Hence, the model has predicted future values for fatal, serious and minor accidents in Benue State as shown in figure 4.3

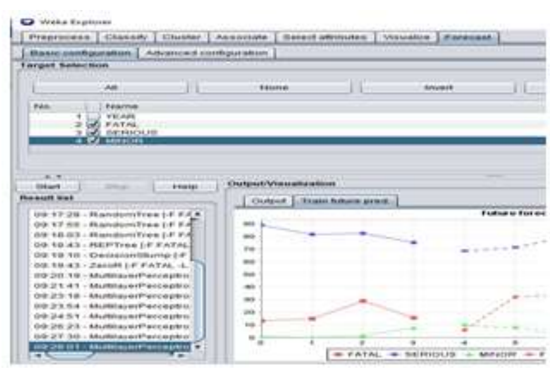

Figure 6: Graph of the Predicted Values from 2019 to 2028 using the WEKA Time Series Forecast

Model Result Validation

In order to validate the accuracy of the J48 decision tree algorithm used to develop the model, two other machine learning algorithms namely Logistics Model Trees (LMT) and Random Forest were used to classify the dataset and compare their results with the result of the J48 decision tree algorithm. The Logistics Models Trees' classification accuracy was 82.4% and the loss or mean absolute error was 0.4444, the classification accuracy of the Random Forest was also 82.4% and the loss or mean absolute error was 0.2154. Table 4.0 below shows the summary of the results of the three machine learning algorithms.

Table 4: Results of the Three Algorithms used to Classify the Dataset

S/N	Machine Learning Algorithm	Classification Accuracy	Loss or Mean Absolute Error
1.	J48 Decision Tree	82.4%	0.2135
2.	Logistics Model Trees	82.4%	0.4444
3.	Random Forest	82.4%	.02154

As depicted in table 4.0 the three algorithms have the same classification accuracy of 82.4%, however the J48 decision tree algorithm has the least loss or absolute error. This is an indication that the J48 algorithm is better than the other algorithm in the classification of the accident dataset.

4. Testing the Model

For the purpose of this study, black box testing technique was used. This involves testing the functionality of the model without knowledge of the design or structure of the code. The various components of the model were tested to ensure that they are free from defects. The testing process was carried out in two phases namely; integrated testing and system testing. Functional testing was done to ensure that the behavior of the model adheres to the requirements specification of the system. The model has several functional components with specific functional requirements specification. Some of these components are as shown in table 5.0.

Table 5: Functional Components of the Model

ID	Name of Component	Functional Description
1	Load and preprocess data component	Loads the dataset into the WEKA and cleans the dataset
2	Split data components	Split the dataset into 75% training dataset and 25% testing dataset.
3	Classify the model component	Classify the dataset of the model.
4	Generate the result component	Generate the result of the model
5	WEKA time series forecast interface and graph	Makes future forecast based on the data in table 2, provides time series forecast algorithm, display results of the forecast, display graph of forecast etc.

The functionality of the various components that made up the model was tested individually before they were combined to form subsystem. These subsystems were also tested to see if these components can be integrated without defects. The following test case was used to test whether the components were properly integrated.

The functional test case for the model is as shown in figure 4.6.

Table 6: Functional Test Case

Test ID	Objective	Test Case Description	Valid Result	Invalid Result
FT1	Load dataset in appendix 3	Read the dataset in appendix 4 to the model, preprocessed and split the dataset into training and test sets	Dataset read and preprocessed and splited into training and test sets	Unable to read, preprocessed and split dataset.
FT2	Classify dataset in appendix 4	The user is expected to use the model to classify the dataset, tune the J48 Decision Tree parameters to achieve the best result	Data classified with high classification accuracy.	Unable to classify dataset or incorrectly classified result.
FT3	Forecast future data and plot Graph	The user is expected to load data in table 2 into the WEKA, select forecast algorithm and click apply	Forecast results display with the time series graph and the result	Unable to display result

Table 7: Integrated Test Case

Test Case ID	Test Case Objective	Test Description	Expected Result
TC1	To check the WEKA read data command line and see if it can read the CSV dataset	Load dataset in appendix 3 to the model	Reads the dataset in appendix 3 into the model.
TC2	To check the split dataset into training and test sets	Splits the dataset in Appendix 3 into 75% training and 25% test sets	Splits the dataset in appendix 3 into training and test sets
TC3	To check the WEKA forecast features values and graph	Load predicted data in table 2 into the WEKA, click on Forecast button and click Start	Show the future predictions in both graph and values.
TC4	To check the predicted values and the graph	View Values per year and check whether they are in the range of the expected value	Show values for each year in the timestamp.

System testing was carried out to ensure that the model works as expected. This is the testing of the fully integrated software product or complete system to ensure that the software or the product meets the required specifications. System testing evaluates functional, behavioral and quality requirements of a system. The focus in this approach is to test the functionality of the complete system.

Table 8: System Test Case

Test ID	Test Case Objectives	Test Case Description	Expected Results
1	Check if the classification accuracy of the model is acceptable	Compute the result and classification accuracy of the model	The accuracy of the model is relatively high.
2	Check if the loss or error is acceptable	Compute the loss or the error of the model	The loss or error of the model is significantly low.
3	Check whether the time series predicted result is within the range of the actual data	Use WEKA time series to forecast values for the classes of accident	The predicted values are in the range of the actual values

V. RESULTS DISCUSSION

The main aim of this research was to develop a road accident predictive model for Benue State that will predict the number and severity of road accidents that will occur in a particle year in Benue State using J48 decision tree algorithm. The model was developed using the dataset in appendix 3 for training, testing and validating the model. The classification accuracy recorded was 82.4% and the absolute error or loss was 0.2135 as shown in figure

The classification accuracy of the model was used to compute the predicted values of each of the three classes (Fatal, Serious and Minor) for the years 2015-2018. The WEKA time series forecast feature was used to predict future values for Fatal, Serious and Minor classes. In this model, the metric used to measure the performance is the classification accuracy. The classification accuracy of 82.4% is appreciably high which indicates excellent performance of the model. The second metric used to measure the performance of the model is the loss or the error. The loss or the error obtained in our model was 0.2135 which is relatively small, hence the model performance is excellent. In order to valid the result of the model, two other machine learning algorithms namely Logistics Model Trees (LMT) and Random Forest were used to classify the accident datasets and the performance of the J48 decision tree algorithm was better because it has the least loss or mean absolute error

.

The model showed efficiency in predicting the number of road accidents for each year. Out of the total of 499 road accident records used for the development of the predictive model, the model predicted 410.2 road accidents which is proportional to an accuracy of 82.4% as shown in table 4.2 This shows that the J48 Decision Tree predictive model is an efficient tool for predicting road accidents in Benue State. Out of the total actual road accident records of 499, 88 of them were fatal accidents and the model was able to predict 72.53 fatal accidents which is about 82.4%. The serious accidents in the actual road accident records was 399 and the model was able to predict 327.78 as serious accidents which represented 82.2% and the minor accidents in the actual road accident records was 12 and the model was able to predict 9.89 which is proportional 82.42%. The analysis of the model's results on yearly bases indicated that the actual values of the road accidents and the predicted values were within the same range. For instance, the actual values of the fatal accidents for the years 2015-2018 as shown in table 1.0 were in the range of 16 to 35 and the predicted values of fatal accidents for the years 2015-2018 in table 2.0 were in the range of 13 to 28. The actual values and the predicted values of serious accidents are also in the same range. Similarly, the actual values of the minor accidents and the predicted values are in the same range. This shows that both the predicted results and the actual results were in the same range which implies the accuracy of the model. The predicted values for the years 2019 to 2028 indicated that the fatal accidents will be in the range of 6 to 34, the serious will be in the range of 67 to 70 and the minor accidents will be in the range of 3 to 10 as shown in table 3.0. These results were found to be in the same range with the actual values in table 1.0. The model was found to have predicted values that are almost the same as the actual values. The J48 decision tree model developed using WEKA was found to have the capacity to predict road accidents based on the input dataset. The model was able to accept the road accidents dataset as input and classified it.

The use of the time series algorithm to make future prediction based on the data in table 3.0 was significant because the results of the output classes predicted by this algorithm were in the same range with the actual values. This has given the model the capability to predict the number of road accidents that will occur in the next ten years and beyond in Benue State. The results from the WEKA time series algorithm as shown in the graph indicated a slight increment in the number of road accidents in Benue State. This is an indication that in the future, the number of road clashes per year will slightly increase.

VI. CONCLUSION

The model was designed using UML and flowchart and implemented using WEKA. The model provides a visual and graphical framework of the predicted results to intending users. It provides a robust platform for users to view the number of road accidents that will occur in the future without rigorous efforts. The model has identified the pattern of the future road accidents based on the existing conditions of roads and other factors in Benue State. The previous road accidents records were collated and processed using data mining techniques and the number of future road accidents were predicted on yearly bases. The accuracy of 82.4% and the error of 0.2135 recorded indicated high performance of the model. However, when compared with Logistics Model Trees (LMT) and Random Forest using the same dataset, we obtained an error of 0.4444 and 0.2154 respectively which implies that J48 decision tree algorithm performs better. The comparison of the predicted values with the actual values shows that the predictions of the model are accurate because they are in the same range. The trend of the graphs also shows that the number of both fatal, serious and minor accidents will increase slightly in the future. The integrated and system tests carried out on the data mining or machine learning processes showed that the functional modules of the model produced the expected result.

Contribution of the Study to Knowledge

Some of the key contributions of the study to knowledge are as follows:

- The developed model provides a machine learning approach to forecasting road accidents in Benue State using human, mechanical, and environmental factors, equipment failure, improper maintenance and head on collision with other vehicles as input variables.
- The ability of the model to predict number of road accidents that will occur in Benue State on yearly bases greatly provides an opportunity to prevent future accidents.

Recommendations

The new model will undoubtedly improve safety in the road transport sector in Benue State by providing the stakeholders information regarding the occurrence of future accidents. We therefore recommend that:

- The road transport stakeholders in Benue State should use the model to generate future road accidents predictions so as to identify useful and accurate pattern and trends of the road accidents
- The pattern or trend of the future accidents generated by the model should be used as basis for road safety plan in the State.
- Adequate precautionary measures should be put in place to mitigate the major road accidents contributory factors identified in this research.

REFERENCES

1. Aci, C. and Ozden, C. (2018). Predicting the Severity of Motor Vehicle Accident Injuries in Adana-Turkey Using Machine Learning Methods and Detailed Meteorological Data.

International Journal of Intelligent Systems and Application in Engineering, 6(1):72-79.

2. Agbonkhese, O. ; Yisa, G. L.; Agbonkhese, E. G.; Akanbi, D. O.; Aka, E. O. and Mondigha, E. B. (2013). Road Traffic Accidents in Nigeria: Causes and Preventive Measures. Journal of Civil and Environmental Research. 3(13).
3. Ahmed, M. (2018).5 Reasons Why Machine Learning Forecasting Is Better Than Traditional Forecasting Techniques.
4. Ahmed, S., Hossaina, M. A., Rayb, S. K., Bhuiyanc, I. M. M. and Sabuj, S. R. (2023). A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance. Journal of Transportation Research Interdisciplinary Perspectives, 19(2023).
5. Ali, P.A. and Erdem, D. (2009). An application of modified Smeed, adapted Andreassen and artificial neural network accident models to three metropolitan cities of Turkey. Journal of Scientific Research and Essay, 4 (9) :906-913.
6. Amorim, B. S. P., Firmino, A. A., Baptista, C. S., Braz Jr, G., de Paiva, A. C. and Almeida Jr, F. E. (2023). A Machine Learning Approach for Classifying Road Accident Hotspots. International journal of Geo-informatic, 12(227).
7. Amuche, A., Esiefarientrhe,B.M. and Agaji, I. (2017). Accident predictive system in Benue State using Ariticial Neural Network. Journal of information and Computing Science, 12(1):033-040.
8. Arafah, M.; Rakha, H.; Abdel-Salam, A. G.; Guo, F. And Medina, A. (2010). Linear Regression Crash Prediction Models: Issues And Proposed Solutions. Retrieved From www.Researchgate.Net/Publication/266140149_Linear_Regression_Crash_PredictionModels_Issues_And_Proposed_Solutions on the 12/07/2020
9. Bzdok, D.; Altman, N. and Krzywinski, M. (2018). Points of Significance: Statistics versus machine learning. 1(15): 233–234.
10. Etim, J. (2018). Road Transport: Definition, Types, Advantages and Disadvantages of Road Transport. Retrieved from http://www.jotscroll.com/forums/3/posts/194/road-transport-typesadvantagesanddisadvantages-of-road-transport.html on the 12th July, 2020.
11. Federal Road Safety Corps (2016). Benue records 183 road accidents in 6 months. Retrieved from https://www.vanguardngr.com/2016/09/benuerecords-183-road-accidents-6- months/ on the 12/07/2020.
12. Gupta, D. L.; Malviya, A. K. and Singh, S. (2012). Performance Analysis of Classification Tree Learning Algorithms. International Journal of Computer Applications, 55(6) :39-44.
13. Gururaj, H. L., Janhavi, V., Tanuja, U., Flamini, F., Soundarya, B. C. and Ravi, V. (2022). Predicting Traffic Accidents and Their Injury Severities Using Machine Learning Techniques. International Journal of Transport Development Integration, 6(4) :363 377.s
14. Kaur, G. and Chhabra, A. (2014). Improved J48 Classification Algorithm for the Prediction of Diabetes. International Journal of Computer Applications, 98(22) :13-17.
15. Khurana, S. (2017). Machine Learning: Decision Tree using Spark For Layman.
16. Littlefield, A. (2016). The Beginner's Guide To Scrum And Agile Project Management. Retrieved from https://blog.trello.com/beginners-guide-scrum-and-agile project-management on the 12/07/2020.
17. Marker, A. (2017). What is prediction in data mining?
18. Mathew, T.V. and Rao, K. V. K. (2007). Introduction to Transportation Engineering: Role of transportation in society.
19. Nagesh, U. B., Rekha, H., Methish, R., Roopashree, J. Nagashree, S. (2021). Analysis and Prediction of Road Accident using machine learning techniques. International journal of creative research thought, 9(8).
20. National Population Commission (2006). 2006 Population and Housing Census: Population Distribution.
21. New Zealand Ministry of Transport (2014). Contribution of Transport to Economic Development: International Literature Review with New Zealand Perspective.

22. Ohakwe, J., Iwueze, I. S. and Chikezie, D. C. (2011). Analysis of Road Traffic Accidents in Nigeria: A Case Study of Obinze/Nekede/Iheagwa Road in Imo State, Southeastern, Nigeria. Asian Journal of Applied Sciences 4(2) :166-175.
23. Rodrigue, J. and Notteboom, T. (2018). Transportation and Economic Development: The Economic Importance of Transportation.
24. Sayad, S. (2018). Decision Tree – Classification.
25. Schwaber, K. and Sutherland, J (2017). Scrum Theory and Values. Retrieved from https://www.scrumalliance.org/learn-about-scrum/scrum-elearning-series/scrum theory-and-values on 12/07/2020.
26. Srija, A. (2011). Confederation of Indian Industry: Employment Potential of the Road Transport Sector. Retrieved from www.cci.gov.in/presentation_document/pdf on the 27th July, 2020.
27. Sulaiman, A.; Babatunde, A.; Babatunde, R. and Hambali, M. (2015). Comparative Analysis of Decision Tree Algorithms for Predicting Undergraduate Students' Performance in Computer Programming. Journal of Advances in Scientific Research & Its Application (JASRA), 2(1) :79-92.
28. The Economic Times (2018). Definition of 'Road Transport. Retrieved from https://economictimes.indiatimes.com/definition/Road-transport on 12/07/2020
29. Wilson, G. and Felix, J. M. (2015). Transportation and National Development. Journal of Economics and Sustainable Development, 6(9) :300-307.