Using Decision Tree Data Mining Algorithm to Predict Causes of Road Traffic Accidents, its Prone Locations and Time along Kano –Wudil Highway

L. J. Muhammad[1], Sani Salisu[2], Atomsa Yakubu[3], Yusuf Musa Malgwi[4], Elrufai Tijjani Abdullahi[5], I .A. Mohammed[6] and Nuhu Abdul'alim Muhammad[7]

[1, 3,7]Mathematics & Computer Science Department, Federal University, Kashere,
[2]Computer Science Department, Federal University, Dutse, Computer Science
[4,5]Department, Modibbo Adama University of Technology, Yola Department of
[6]Public Administration, Hassan Usman Katshina Polytechnic, Katsina, Nigeria,
[7]Department of Computer Science, Yobe State University, Damaturu, Yobe State, Nigeria
mljtech@gmail.com[1], sani.salisu@fud.edu.ng[2], au.nlaro@gmail.com[3]
yumalgwi@yahoo.com[4] tijjayrufai@gmail.com[5],ibrahimsallau@gmail.com[6]
,elgurama1989@gmail.com[7]

Abstract

Road traffic accidents, the inadvertent crash involving at least one motor vehicle, occurring on a road open to public circulation, in which at least one person is injured or killed; intentional acts (murder, suicide) and natural disasters excluded, is indisputably one of the most frequent and most damaging calamities bedeviling human societies, in particular, Nigeria, today. It is therefore, of paramount importance to seek to identify the root causes of road traffic accidents in order to proffer mitigating solutions to address the menace. This research, aimed at predicting the likely causes of road accidents, its prone locations and time along Kano– Wudil highway in order to take all necessary counter measures is a step forward in this direction. In this study data mining decision tree algorithm was used to predict the causes of the accidents, its prone locations and time along Kano – Wudil Highway that links Kano State to Wudil Local Government Area Kano State for effective decision making.

Keywords: Accident, Data mining, Decision tree, Id3 tree, Algorithm

1. Introduction

Road Traffic Accidents killed more than 1.2 million people, and injured between 20 and 50 million others in 2004, thereby becoming the ninth most common cause of death in that year. Road traffic accidents remains among the most central public health problems in the world. A tragic fact is that among the young people aged between 15 and 29 years, road traffic accident is one of the most common causes of death worldwide [3].

The incidence of fatal road accidents in Nigeria is phenomenal. Trend analysis of fatal road accidents between June 2006 and May 2014 using Nigeria Watch database shows that 15,090 lives were lost to fatal road accidents in 3,075 events. The highest fatality occurred in 2013 (2,061 deaths), a 2.8% increase from the 2012 record of 1,652 deaths. However, the figures were rising again in 2014, with fatality records of 964 deaths between January and May 2014 [2].

Nigeria is ranked second-highest in the rate of road accidents among 193 countries of the world. Aside from the Boko Haram crisis, accidents are currently by far the main most violent cause of death in Nigeria. The World Health Organization (WHO) adjudged Nigeria the most dangerous country in Africa with 33.7 deaths per 100,000 population

every year. According to their report, one in every four road accident deaths in Africa occurs in Nigeria [3].

The intensity of fatal road accidents varies from state to state. States are ranked by a low, medium, or high severity index. Ebonyi and Lagos states ranked low (0.00 –2.99%) in the severity index of fatal road accidents that occurred between June 2006 and May 2014. Ebonyi had a severity index of 2.4 from 51 deaths in 21 crashes, while Lagos had a low severity index of 2.5 from 1,590 deaths in 620 crashes. States that are ranked medium (3.00 –5.99%) include FCT (Abuja), Ekiti, Delta, Akwa Ibom,Plateau, Bauchi, Bayelsa, Kwara, Osun, Cross River, Taraba, Ogun, Abia, Nasarawa, Oyo, Anambra, and Sokoto. Those ranked with a high index (6.00 – 8.99%) are Imo, Jigawa, Benue, Niger, Edo, Gombe, Borno, Ondo, Enugu, Kano, Kaduna, Rivers, Zamfara, Kogi, Katsina, Kebbi, Adamawa, and Yobe. Highest in this category is Yobe State, with a severity index of 11.4. [2].

In Nigeria today, the issue of road accidents has become a teething problem and there has been a major problem of the cardinal causes responsible for these accidents and their prone locations along high ways in Nigeria. This research harnessed data mining algorithms to predict the causes of accident and accident prone locations along Kano-Wudil Highway for counter-accidents decision making by relevant stakeholders.

2. Related Works

Many works have been carried out by different researchers on the prediction of the causes of accidents, their prone locations along various roads around the globe using data mining algorithms.

[5] conducted data mining research focusing on building tree-based models to analyze freeway accident frequency. Using the 2001-2002 accident data of National Freeway 1 in Taiwan, the authors developed classification and regression tree (CART) and negative binomial regression models to establish the empirical relationship between traffic accidents and highway geometric variables, traffic characteristics, and environmental factors. The authors found that the average daily traffic volume and precipitation variables were the key determinants of freeway accident frequency.

[6] investigated the potential application of data mining algorithms to develop models supporting the identification and prediction of major driver and vehicle risk factors that cause road traffic accidents. The research used the WEKA version 3-5-8 tool to build the decision tree (using the J48 algorithm) and rule induction (using PART algorithm) techniques. Performance of the J48 algorithm was slightly better than that of the PART algorithm. The license grade, vehicle service year, vehicle type, and experience were identified as the most important variables for predicting accident severity.

[8] investigated the cause of accident and accident prone locations on Highways in Nigeria, Lagos- Ibadan highway as a case study using decision tree data mining algorithm. WEKA software was used to analyse accident data gathered along this road. The results showed that causes of accidents, specific time/condition that could trigger accident and accident prone areas could be effectively identified. There were 50 rules generated from this tree. Rule 1- 18 indicate the occurrence of accident in Location 3 and rule 19-50 also shows the occurrence of accident in location 2. This indicates that, location 2 (Above 10km – 20km, from Lagos to Ibadan) has the highest number of road accident occurrence with Heavy-vehicle in the after-noon and during the dry season. Rule 41 is the best one that can be used for prediction. The rule says that, Tyre bust is the cause of road accident with heavy vehicle within location 2 in the day time and during the dry season.

This research harnessed the decision tree data mining algorithm to predict the causes of accident and accident prone locations along Kano- Wudil highway. The reason of using decision tree data mining algorithm for the research could be seen in the research of [8], where a comparison of different Decision Tree algorithms and Artificial Neural Networks

performance were analyzed using road accidents data set. The location is between the first 40 kilometers along the Ibadan-Lagos Express road. The work used Multilayer Perceptron as well as Radial Basis Function (RBF) Neural Networks, Id3 and Function Tree algorithms. Results shows that the Id3 tree algorithm performed better with higher accuracy rate, while Radial basis function performed better than multilayer perceptron in terms of time used in the building of the model and number of correctly classified instances. The result showed that, Decision Tree techniques outperformed Artificial Neural Networks with a lower error report and with a higher number of correctly classified instances and better accuracy rate generated. Tyre burst, broken shaft and loss of control variables were the three major causes of accidents where tyre burst represents the major cause of accidents.

3. Methodology

3.1 Data Mining

Is defined as extracting information from huge sets of data. In other words, we can say that data mining is the procedure of mining knowledge from data. It can also be defined as an interactive process of discovering valid and novel, useful and understandable patterns or models in large database. Data Mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions. Data mining uses advances in the field of Artificial Intelligence (AI) and Statistics.

3.2 Decision Tree

Decision tree is a "divide-and-conquer" approach to the problem of learning from a set of independent instances, which leads naturally to a tree-like style of representation called a decision tree. A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node.

This research harnessed decision tree data mining algorithm to predict the causes of the accident, its prone locations and time along Wudil – Kano Highway.

3.3. Accidents along Wudil- Kano Highway

Kano to Wudil Highway is one of the busiest roads in Kano. It is a route that links Kano to Jigawa, Gombe, Bauchi, Adamawa, Taraba, Maiduguri and Yobe States and it is 44km away from Kano City, along Maiduguri Road. According to FRSC Traffic Digest, January 2014, total Road Traffic Crashes 164 cases increased by 24% in 2011 compared to 2010 figure of 132. Thereafter, a downward trend is observed in total Road Traffic Crashes Fatality (119) increased by 53% in 2013 over 2012 figure of 78 deaths. There was consistent increase in Traffic volume per hour from 2010 with 392 vehicles per hour to 542 vehicles per hour in 2013 representing 38% increase along Kano – Wudil Route. In January, 2014 Wudil - Kano Highway is among Top 20 Routes with highest no. of person killed and Top 20 Routes with Highest Number of Persons Injured. Chart 1.1 and Chart 1.2 depicts the top 20 routes with highest of person killed and that of highest number of persons injured in January, 2014.

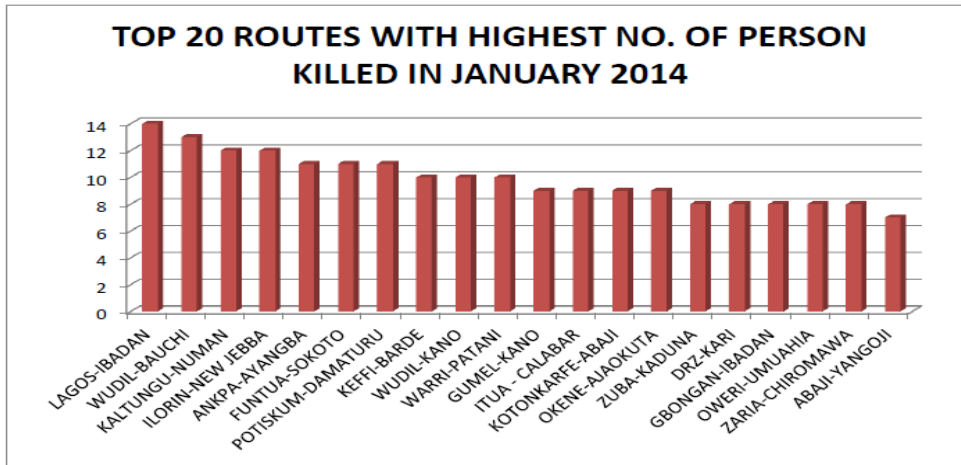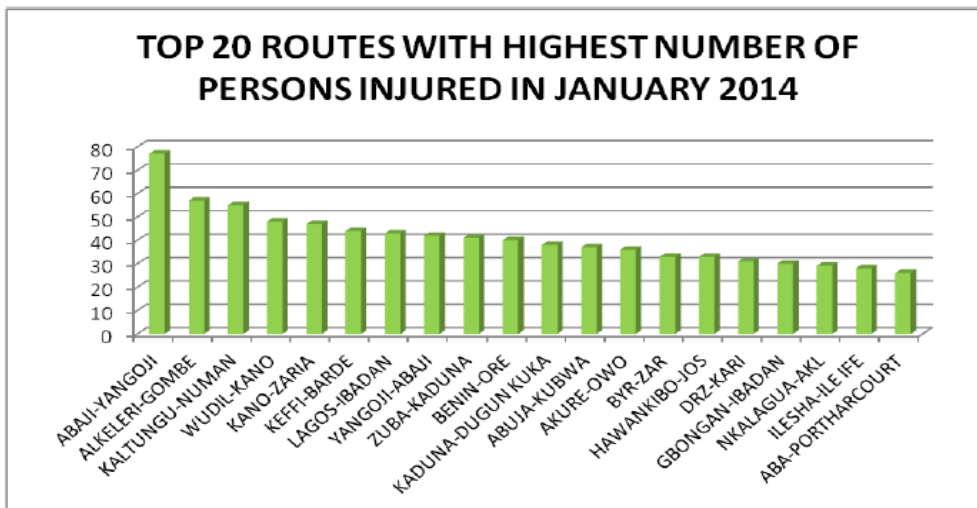**Chart 3.1. Source: Federal Road Safety Corp (2014)**

Chart 3.1. Source: Federal Road Safety Corp (2014)

The above statistics depicts that predicting the causes of accident, its prone locations and time along Kano – Wudil Highway is very important input for counter–accidents decision making of the high way. Table 3.1 depicts the statistics of the accident along the high way between January, 2014 to June, 2016.

Table 3.1

SN	Month of the Accident	No. of Accidents
1	January, 2014	8
2	February, 2014	6
3	March , 2014	5
4	April , 2014	6
5	May , 2014	7
6	June , 2014	1
7	July , 2014	4
8	August , 2014	7
9	September , 2014	5
10	October , 2014	4
11	November, 2014	8
12	December , 2014	5
13	January, 2015	6
14	February, 2015	7
15	March , 2015	5
16	April , 2015	7
17	May , 2015	6
18	June , 2015	10
19	July , 2015	5
20	August , 2015	7
21	September , 2015	9
22	October , 2015	8
23	November, 2015	2
24	December , 2015	2
25	January, 2016	8
26	February, 2016	2
27	March , 2016	11
28	April , 2016	3
29	May , 2016	4
31	June , 2016	5
Total Number of Accidents		**165**

4.5 Data Modeling

The research considered the data of accident record between 44 km Kano to Wudil L.G.A. The data were organized into a relation. The sample data used covered the period of 30 Months, from January 2014 to June, 2016 as indicated in Table 3.1 The output variable is the location and the locations can be divided into four distinct locations tagged location A, B, C and D. Location 1 – 11km is LocationA, above 11 km – 22km LocationB, above 22km – 33km LocationC and above 33 km – 44km LocationD. Table 3.2 Showing variables given both continuous and categorical value.

Table 3.2

SN	Variables	Description	Value	Type
1	Type of vehicle	Small Car	A	Categorical
		Heavy Car	B	Categorical
2	Time of accident	Morning	A	Categorical
		Afternoon	B	Categorical
		Evening	C	Categorical
3	Causes of accident	OverSpeed	A	Categorical
		LossOfControl	B	Categorical
		WrongOvertaking	C	Categorical
		TyreBlowouts	D	Categorical
		PoorLights	E	Categorical
		UncertainCause	F	Categorical
		BrakeFailure	G	Categorical
4	Location of Accident	LocationA	A	Categorical
		LocationB	B	Categorical
		LocationC	C	Categorical
		LocationD	D	Categorical

4. Results

Weka data mining software was used to mine the data using Id3 decision tree data mining algorithm. The algorithm is one of the most widely used and practical methods for inductive inference over supervised data. It represents a procedure for classifying and categorical data based on their attributes. It is also efficient for processing large amount of data, so is often used in data mining application.

4.1 Causes of accident

Below are results obtained using Id3 decision tree for the causes of accident along Kano - Wudil Highway.

4.1.1 Id3 tree

AccidentLocation = LocationA: UncertainCause
AccidentLocation = LocationB
| VehicleType = SmallCar: PoorLights
| VehicleType = HeavyCar: LossOfControl
AccidentLocation = LocationC
| AccidentTime = Morning: WrongOvertaking
| AccidentTime = Afternoon: null
| AccidentTime = Evening: WrongOvertaking
AccidentLocation = LocationD
| AccidentTime = Morning: OverSpeed
| AccidentTime = Afternoon: WrongOvertaking
| AccidentTime = Evening
| | VehicleType = SmallCar: WrongOvertaking
| | VehicleType = HeavyCar: BrakeFailure

4.1.2 Summary of the Result

Correctly Classified Instances   120   72.7273 %
Incorrectly Classified Instances   45   27.2727 %
Kappa statistic   0.5803
Mean absolute error   0.0989
Root mean squared error   0.2235
Relative absolute error   49.0673 %
Root relative squared error   70.7409 %
Total Number of Instances   165

4.1.3 Detailed Accuracy by Class

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.636	0.098	0.618	0.636	0.627	OverSpeed
0.375	0	1	0.375	0.545	LossOfControl
0.84	0.381	0.68	0.84	0.751	WrongOvertaking
0	0	0	0	0	TyreBlowouts
1	0	1	1	1	PoorLights
1	0	1	1	1	UncertainCause
1	0	1	1	1	BrakeFailure

4.1.4 Confusion Matrix

a	b	c	d	e	f	g	<-- classified as
21	0	12	0	0	0	0	a = OverSpeed
0	6	10	0	0	0	0	b = LossOfControl
13	0	68	0	0	0	0	c = WrongOvertaking
0	0	10	0	0	0	0	d = TyreBlowouts
0	0	0	0 1	0	0	0	e = PoorLights
0	0	0	0	0	6	0	f = UncertainCause
0	0	0	0	0	0	9	g = BrakeFailure

4.2. Locations of Accident

Below are results obtained using Id3 decision tree for the locations of accident along Kano - Wudil Highway.

4.2.1. Id3 Tree

AccidentCause = OverSpeed
| AccidentTime = Morning: LocationD
| AccidentTime = Afternoon: null
| AccidentTime = Evening: LocationC
AccidentCause = LossOfControl
| VehicleType = SmallCar: LocationC
| VehicleType = HeavyCar: LocationB
AccidentCause = WrongOvertaking
| AccidentTime = Morning: LocationD
| AccidentTime = Afternoon: LocationD
| AccidentTime = Evening: LocationC
AccidentCause = TyreBlowouts: LocationD

AccidentCause = PoorLights: LocationB
AccidentCause = UncertainCause: LocationA
AccidentCause = BrakeFailure: LocationD

4.2.2. Summary of the Result

Correctly Classified Instances   133   80.6061 %
Incorrectly Classified Instances   32   19.3939 %
Kappa statistic   0.6838
Mean absolute error   0.0951
Root mean squared error   0.2231
Relative absolute error   31.3524 %
Root relative squared error   57.4578 %
Total Number of Instances   165

4.2.3. Detailed Accuracy by Class

TP Rate	FP Rate	Precision		Recall	F-Measure	Class
1	0	1		1	1	LocationA
1	0	1		1	1	LocationB
0.831	0.208	0.69	0.831		0.754	LocationC
0.738	0.123	0.861		0.738	0.795	LocationD

4.2.4 Confusion Matrix

a   b   c   d   <-- classified as
6   0   0   0 | a = LocationA
0   16   0   0 | b = LocationB
0   0   49   10 | c = LocationC
0   0   22   62 | d = LocationD

4.3. Time of Accidents

   Below are results obtained using Id3 decision tree for the time of accident along Kano - Wudil Highway.

4.3.1. Id3 Tree

AccidentCause = OverSpeed
| AccidentLocation = LocationA: null
| AccidentLocation = LocationB: null
| AccidentLocation = LocationC: Evening
| AccidentLocation = LocationD: Morning
AccidentCause = LossOfControl: Evening
AccidentCause = WrongOvertaking
| AccidentLocation = LocationA: null
| AccidentLocation = LocationB: null
| AccidentLocation = LocationC: Evening
| AccidentLocation = LocationD: Afternoon
AccidentCause = TyreBlowouts: Evening
AccidentCause = PoorLights: Evening
AccidentCause = UncertainCause: Morning
AccidentCause = BrakeFailure: Evening

4.3.2. Summary of the Result

Correctly Classified Instances	127	76.9697 %
Incorrectly Classified Instances	38	23.0303 %
Kappa statistic	0.6077	
Mean absolute error	0.1821	
Root mean squared error	0.3036	
Relative absolute error	49.5963 %	
Root relative squared error	70.9891 %	
Total Number of Instances	165	

4.3.3. Detailed Accuracy by Class

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.529	0	1	0.529	0.692	Morning
1	0.182	0.386	1	0.557	Afternoon
0.856	0.162	0.883	0.856	0.869	Evening

4.3.4. Confusion Matrix

```
a     b     c   <-- classified as
27    13    11 |  a = Morning
0     17    0  |  b = Afternoon
0     14    83 |  c = Evening
```

5. Discussion

There are 7 identified causes of accidents along the Kano-Wudil Highway which include; over speed, loss of control, wrong overtaking, tyre blowouts, poor lights, uncertain causes and brake failure. The result showed that out of the 165 instances of the accident, between January, 2014 to June, 2016, 81 instances of the accident occurred as a result of wrong overtaking, 33 as a result of over speed, 16 instances as a result of loss of control, 10 instances as a result of tyre blowout, 10 instance also as a result of poor light, 9 instances as a result of brake failure and 6 instances of the accident was uncertain. The best decision tree result was obtained with Id3 with 165 instances, 120 instances were correctly classified and 45 instances were incorrectly classified, which represent 72.7273%, 27.2727% respectively. The mean absolute error is 0.0989, root mean squared error is 0.2235, relative absolute error is 49.0673 % and root relative squared error is 70.7409 %.

For the prone locations of the accident, the result indicates that, out of 165 instances of accident 84 occurred at location D, 59 accidents occurred at Location C, 16 accidents occurred at Location B and 6 accidents occurred at Location A. . The best decision tree result was obtained with Id3 with 165 instances, 133 instances were correctly classified and 32 instances were incorrectly classified, which represent 80.60661%, 19.3939% respectively. The mean absolute error was 0.0951, root mean squared error was 0.2231, relative absolute error was 31.3524 % and root relative squared error was 57.4578 %. For the prone times of the accident, the result also indicates that, out of 165 instances of accident 97 occurred in the evening, 51 accidents occurred in morning and 17 accidents occurred in afternoon. However, the best decision tree result was obtained with Id3 with 163 instances, where 127 instances were correctly classified and 38 instances were incorrectly classified, which represent 76.9697% and 23.0303 respectively. The mean absolute error is 0.1821, root mean squared error is 0.3036, relative absolute error is 49.5963 % and root relative squared error is 70.9891 %.

6. Conclusion

The historical data collected for the accidents occurred between January, 2014 to June, 2016 along Kano–Wudil Highway was analyzed using WEKA data mining software using Id3 decision tree and predicted the causes of the accident, its prone location and time. The result showed that mostly the cause of the accident is wrong overtaking, followed by loss of control, then tyre blowout, poor lights, uncertain causes and brake failure. The result indicated that, accident mostly occurred in location D, followed by Location C, then B and the mostly at evening time.

References

[1]   Data Mining - Decision Tree Induction (2016) accessed date: 8th July, 2016 http://www.tutorialspoint.com/data_mining/dm_dti.htm
[2]   Federal Road Safety Corp (2014). Traffic Digest Report on Road Traffic Crashes (RTC) Involving Buses on Nigerian Roads (2014)
[3]   Global status report on road safety: time for action, WHO, 2009.
[4]   J. Han & M. Kambe (2001) Data mining Concepts and Techniques Morgan Kaufmam, Academic Press
[5]   L. Chang & W. Chen (2005). Data mining of tree based models to analyze freeway accident frequency. Journal of Safety Research 36: 365- 375
[6]   M. Getnet (2009). Applying data mining with decision tree and rule induction techniques to identify determinant factors of drivers and vehicles in support of reducing and controlling road traffic accidents: the case of Addis Ababa city. Addis Ababa Addis Ababa University
[7]   N. U. Vitus (2014) Trends and patterns of fatal road accidents in Nigeria (2006-2014) IFRA-Nigeria working papers series
[8]   T. A. Dipo & O. Akinbola (2012). Using Data Mining Technique to Predict Cause of Accident and Accident Prone Locations on Highways. American Journal of Database Theory and Application 2012, 1(3): 26-38
[9]   V.A. Olutayo & A. A. Eludire (2014). Traffic Accident Analysis Using Decision Trees and Neural Networks. I.J. Information Technology and Computer Science, 2014, 02, 22-28