

## Prediction of Road Traffic Accident in Nigeria Using Naive Baye's Approach

**Oyetunji, M.O.**

Department of Computer Science & Mathematics  
Mountain Top University  
Km 12 Lagos-Ibadan Express Way  
Prayer City, Ogun State, Nigeria  
mooyetunji@mtu.edu.ng

**Oladeji, F.A.**

Department of Computer Science  
University of Lagos  
Akoka, Lagos, Nigeria  
foladeji@unilag.edu.ng

**Falana, O.J.**

Department of Computer Science & Mathematics  
Mountain Top University  
Km 12 Lagos-Ibadan Express Way  
Prayer City, Ogun State, Nigeria  
ojfalana@mtu.edu.ng

**Idowu, P.A.**

Department of Computer Science and Engineering  
Obafemi Awolowo University  
Ile-Ife, Nigeria.  
paidowu@oauife.edu.ng

### ABSTRACT

Road Traffic Accidents (RTAs) are increasing with rapid pace and presently these are one of the leading causes of death in developing countries. Available data of auto road accident indicate that at least, 162 persons out of 100,000 Nigerians are regular victims of road accidents. This paper presents a predictive model for forecasting road traffic accident in Nigeria using Naïve Bayes'. Naive Bayes' classifier features of Waikato Environment for Knowledge Analysis software was used to formulate the model. The data used consists of 600 road traffic accident data. The result of the prediction shows the system is reliable with 89.83% accuracy using selected dependent variables like the road condition, road dimension, human factor and the vehicular factors. In conclusion, this research presents a road traffic accident predictive model for forecasting road traffic accidents in Nigeria in order to prevent or reduce the occurrence of road traffic accidents using naive bayes' model.

**Keywords:** Road Traffic Accident, Predictive Model, Road condition, Naive Bayes, Road Dimension, Human Factor and Vehicular Factor

---

### Aims Research Journal Reference Format:

Oyetunji, M.O, Oladeji, F.A., Falana, P.A & Idowu, P.A. (2017): Prediction of Road Traffic Accident in Nigeria Using Naive Baye's Approach. *Advances in Multidisciplinary & Scientific Research Journal*. Vol. 3. No.1, Pp 23-30

### 1. INTRODUCTION

Road transport is the commonest means of transportation in Nigeria following the Transportation is an essential part of modern existence, linking the various activities which people participate especially at home, at school, at work, and go to shopping also traveling. Accident is commonly occurring involving one or more transportation vehicle in collision that result in property damages, injury or death. Public always expected that transportation system safe and efficient for all user (Binti et al., 2011) According to Nigerian Federal Road Safety Commission, the country has the highest rate of death from motor accidents in Africa; leading 43 other nations in the number of deaths per 10,000 vehicle road traffic accident (Atubi, 2010).

Nigeria is followed by Ethiopia, Malawi and Ghana with 219, 183 and 178 deaths per 10,000 vehicles, respectively (Daramola, 2004). It is also evident that Nigeria is worse than most other countries in terms of road accidents, in spite of her relatively good road network. Nigeria, a slight heavily motorized country with poor road conditions and transport systems has a high rate of Road Traffic Accidents and the tendency is on the increase (Ohakwe *et al.* 2011). According to the Nigerian Federal Road Safety Corps (FRSC), between 1970 and 2001, Nigeria recorded a total of 726,383 road traffic accidents resulting in the death of 208,665 persons and 596,425 injuries. In that period, each succeeding year recorded more accidents, deaths and injuries. Also between 1997 and 2002, Lagos State alone recorded a total of 39,141 road accidents resulting in the death of 10,132 persons and 18,972 injuries (Atubi, 2006).

More than half of all global road traffic deaths occur among young adults between 15 and 44 years of age and the position of road traffic injuries as a contributor to the global burden of disease is predicted to rise from tenth place in 2002 to eighth place by 2030. Road traffic injuries put significant financial strain on families. 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 (Osoro, 2011). Faulty design, multiple bends, but especially poor or outright lack of maintenance has rendered most of our over 194,000 kilometers of roads in Nigeria as death traps (Osita, 2011) and there is a need for a with a predictive model which allows road officials and road managers to understand the state of the roads and the distribution of road traffic accidents in order to predict future road traffic accidents.

There is a need for a model that can be used to store and compare the distribution of accidents over the years in addition to better understanding the conditions of the road and also using it to find out the factors affecting road traffic accidents thereby identifying the hazardous locations or blackspots in each route thereby also serving as a decision support model for road managers and road administrations so as to prevent or decrease future road traffic accidents. Presence of potholes and the surface conditions on the road contribute to occurrence of road traffic accidents. However, condition of the road needs to be taken into consideration before embarking on a journey. This paper presents the distribution of accidents over the years in addition to better understanding the conditions of the road in South West Nigeria; while, the primary objective of the study is to identify factors that contribute to the cause of accident thereby identifying the hazardous locations or the blackspots location in each route and develop an accident prediction model for the road traffic accident using naive bayes' technique.

## 2. RELATED WORKS

Many Researchers had worked on the prediction of road traffic accident using different types of variables like traffic flow, human factor and vehicular factor but in this paper, road condition, road dimension, human factor and vehicular factor was considered. Some of the researcher and the result of their finding are discussed next.

Zheng *et.al.*, (2011) measure the predictive ability of two models using Data mining predictive techniques (fuzzy logic and neural network techniques) to determine road accident frequencies. It is established based on a data set of 133 segments from urban arterials in Harbin city of China, which takes annual average daily traffic (AADT), lane width (LW), speed limit (SL) and traffic load (TL, calculated by volume/capacity) as input variables and accidents per kilometer per year (AF) as output variable and discovered that both fuzzy logic and neural network techniques were able to predict road accident frequencies with an accuracy of 60% to 100% based on selected dependent variables. Road condition and the vehicular factor were not considered for the prediction.

Abojarudeh (2013) established Traffic Accidents prediction models to improve traffic safety in greater Amman area with the use of regression models. It was developed to predict the city artery traffic accidents, and find the prominent influence factors of high risk artery. The factors used are the dependent variable accident frequency, injured and fatalities while independent variables are driver behaviour mistake that cause traffic accident. Using the SPSS software to analyze the data provided on variables along the different streets of the study location, a number of regression models were developed using a combination of certain driver dependent factors which were responsible for causing all sorts of accidents on each street, signalized intersections and unsignalized intersections. In conclusion, it was discovered that certain number of driver behaviour determine accident frequencies, injuries and fatalities on the streets and accident frequencies at signalized and unsignalized intersections. The limitation of the model was that only the driver behavior was used and other factors like the vehicular factor, the road condition and the road characteristics were not considered for the prediction.

Thomas (2005) developed Predictive models for accidents (involving bicyclist, pedestrian and vehicles) using generalised linear modelling with a Quasi-Poisson distribution to take into account that the accident data is over dispersed compared to a true Poisson distribution to improve accident models for urban links. The strength of the study with the factors considered were the no of road users, vehicle speeds, separation of vehicle accident into single and multiple vehicle accidents. Other factors like the road conditions, road characteristics were not considered for the study. The study shows that the inclusion of exposure data for vulnerable road users' accidents greatly improves the predictive ability of the models. Models including exposure data predicted 71% - 81% of the systematic variation in vulnerable road users (VRU) accident. Models excluding VRU exposure variables predicted 54 -55% of the systematic variation and models including only the variables used in the existing Swedish accident models predicted just 37% of the systematic variation.

Fajaruddin (2011) developed two Accident Prediction Models for rural roadway at Parit Raja area from kilometer 19 to kilometer 23 and used multiple linear regression models to relate the discrete accident data with the road and traffic flow explanatory variable. The models were calibrated based on data collected over a 4 year period (2004 - 2007) using variables like approach speed, Annual Average Daily Traffic (AADT), number of access points per kilometer, traffic light and time gap. The two predictive models developed have a coefficient of determination, R<sup>2</sup> of 0.9973 and 0.9979 respectively. It was also discovered that the factors that contributed to accidents at four lanes two-way undivided rural roadways are, number of access points, vehicle speed, Annual Average Daily Traffic (AADT), motorcycle, motorcar and gap. The strength, the model developed for Malaysian rural roadway in this paper appear to be useful for many application such as the existing number of access point, increasing number of motorcycle and motorcar, rise in speed and shorted in gap are among the potential contribute of increment in accident rate. Other factors like the road conditions, the vehicular factor and the road characteristics were not considered for the prediction.

### 3. RESEARCH METHODOLOGY

Extensive study on related existing body of knowledge on road accident, prediction of road traffic accident, naive baye's and prediction of road traffic accident were carried out. The study adopted the descriptive and exploratory designs that allow the collection of data from some routes. Personal interview (both structured and semi-structured) was also used for the collection of data from the Federal Road Safety Corps (FRSC) in South Western Nigeria. Most routes are mostly in Osun State axis with a view to formulate a road traffic accident predictive model using data mining approach.

The classification task in naive Bayesian classifier is performed by evaluating the posterior probability. Given a set of variables  $X = \{x_1, x_2... x_n\}$ , the algorithm attempts to determine the posterior probability of an event  $C_i$  among a set of possible outcomes  $C = \{c_1, c_2... c_j\}$ . Simplified formula for calculating posteriors: Using this modified Bayes' rule formula, a new case  $X$  is labeled with a class level  $C_i$  that achieves the highest posterior probability.

The predictive model were developed using supervised learning techniques (naive baye's classifier in WEKA's environment. The performance of this model was tested and we found out that it can forecast road traffic accident with an accuracy of 89.83% based on selected dependent variables.

#### 4. RESULT OF THE ROAD TRAFFIC ACCIDENT PREDICTION MODEL

##### Naive Bayes' classifier

Mathematically, the Naive Bayes' Classification is expressed as:

$$P(X|C_i) = P(x_1|C_i) * P(x_2|C_i) * P(x_3|C_i) * \dots * P(x_k|C_i) \quad (1)$$

Hence; the Naive Bayes Classification technique for road traffic accident

For any given input; X defines as:

The inputs are

X = (x<sub>1</sub> = "Route Name = value", x<sub>2</sub> = "Date = value", x<sub>3</sub> = "Time of accident = value", x<sub>4</sub> = "Total Casualty = value", x<sub>5</sub> = "Vehicles involved = value", x<sub>6</sub> = "Human Factors = value", x<sub>7</sub> = "Vehicular factors = value", x<sub>8</sub> = "Route Length = value", x<sub>9</sub> = "Carriage width", x<sub>10</sub> = "Surface = value", x<sub>11</sub> = "Pot-Holes = value", x<sub>12</sub> = "Failed Segments = value", x<sub>13</sub> = "Junctions = value", x<sub>14</sub> = "Bends = value")

Is calculated for each likely output as:

$$\begin{aligned} P(X|Fatal) &= (P(x_1|Fatal) * P(x_2|Fatal) * P(x_3|Fatal) * P(x_4|Fatal) * P(x_5|Fatal) * P(x_6|Fatal) * P(x_7|Fatal) * P(x_8|Fatal) \\ &* P(x_9|Fatal) * P(x_{10}|Fatal) * P(x_{11}|Fatal) * P(x_{12}|Fatal) * P(x_{13}|Fatal) * P(x_{14}|Fatal) \\ P(X|Minor) &= (P(x_1|Minor) * P(x_2|Minor) * P(x_3|Minor) * P(x_4|Minor) * P(x_5|Minor) * P(x_6|Minor) * P(x_7|Minor) * \\ &* P(x_8|Minor) * P(x_9|Minor) * P(x_{10}|Minor) * P(x_{11}|Minor) * P(x_{12}|Minor) * P(x_{13}|Minor) * P(x_{14}|Minor) \\ P(X|Serious) &= (P(x_1|Serious) * P(x_2|Serious) * P(x_3|Serious) * P(x_4|Serious) * \\ &* P(x_5|Serious) * P(x_6|Serious) * P(x_7|Serious) * P(x_8|Serious) * P(x_9|Serious) * \\ &* P(x_{10}|Serious) * P(x_{11}|Serious) * P(x_{12}|Serious) * P(x_{13}|Serious) * P(x_{14}|Serious) \end{aligned}$$

We determine the probability of accident output as:

$$\begin{aligned} P(Minor|X) &= P(X|Minor) * P(Minor) \\ P(Fatal|X) &= P(X|Fatal) * P(Fatal) \\ P(Serious|X) &= P(X|Serious) * P(Serious) \end{aligned}$$

The output is

$$Y = \text{MAX} [P(Minor|X), P(Fatal|X), P(Serious|X)]$$

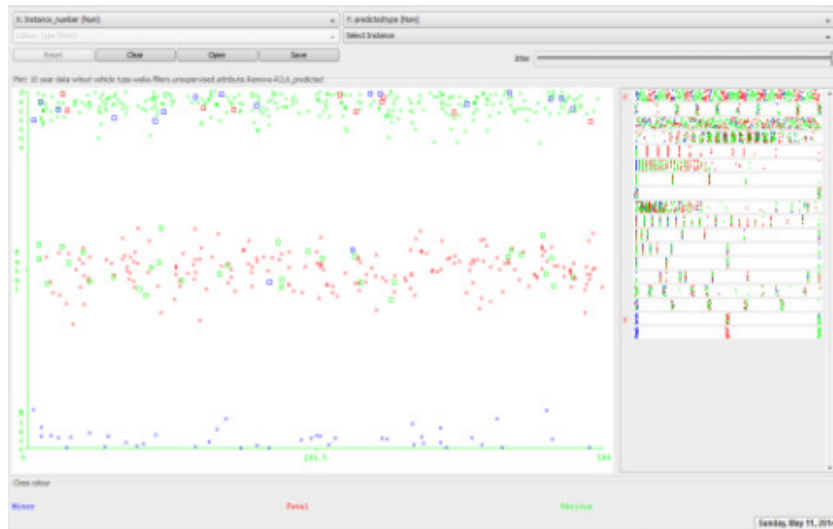


Fig 1: Naive Bayes Graph

A	B	C ←	Classified as
43	2	16	A = Minor
0	196	10	B = Fatal
0	33	300	C = Serious

Fig 2: Confusion Matrix of the Results of Naive Bayes' Classification

Table 1: Summary of the Results of Naive baye's Model

Naïve Baye's Model	
Data Classified	600
Correct Classification	539
Incorrect Classification	61
Accuracy (%)	89.3333
Mean Absolute Error (MAE)	0.0962
Root Mean Square Error (RMSE)	0.2249
Relative Absolute Error (RAE) (%)	25.5798

Table 2: Summary of Naïve Baye's Model's Performance

Naïve Baye's Model			
	<i>Minor</i>	<i>Fatal</i>	<i>Serious</i>
True Positive Rate (recall)	0.705	0.951	0.901
False positive Rate	0	0.089	0.097
ROC Area	0.996	0.965	0.965
Precision	1	0.92	0.92
F-Measure	0.827	0.91	0.91

## 5. DISCUSSION

The result of the classification of the 10 year data supplied to the classifier showed that: out of the total 600 accident data provided there were 539 correct classifications and 61 misclassifications of the road accident status. Out of the 600, 61 were minor accidents, 206 were fatal accidents and 333 were serious accidents. Out of the 61 minor cases – 43 were correctly classified while 2 and 16 were misclassified as fatal and serious cases respectively. Out of the 206 fatal cases – 196 were correctly classified and 110 were misclassified as serious with no minor misclassification. Out of the 333 serious cases – 300 were correctly classified with the remaining 33 misclassified as fatal accidents. The result of this classification showed a prediction accuracy of 89.33% with a mean absolute error (MAE) of 0.0962, root mean square error of 0.2249 and 25.5798%. These values were used to plot the TP rate (recall), FP rate, ROC area, precision and F-measure. The TP rate which had a value of 0.705, 0.951 and 0.901 for minor, fatal and serious respectively showed that the naïve bayes' model was able to predict 70.5% of the actual minor data but 95.1% and 90.1% of the actual serious and fatal cases.

The FP rate which had a value of 0.827, 0.91 and 0.91 for minor, fatal and serious respectively showed that the naïve bayes' model had 30%, 10.1% and 10.8% of actual cases that were misclassified. The precision which had a value of 1, 0.92 and 0.92 for minor, fatal and serious cases respectively showed that 70.5% of the predicted cases were actually minor and 95.1% and 90.1% of the predicted cases were actually fatal and serious from the actual data. This results show that the naïve bayes' model has the tendency of predicting serious cases correctly followed by fatal and minor. Also, it has the tendency of misclassifying about 30% of the fatal and serious data. The area under the ROC graph which has a value of 0.996 shows that the naïve bayes' model is not just randomly guessing the minor cases. It also has about 10% likelihood of randomly guessing the fatal and serious cases judging from the value of 0.695 and 0.965.

**TABLE 3: VALIDATION OF RESULT**

Route name	Accident date	time	Cas.	Total vehicle	Surface	Pot holes	Failed segment	junctions	Bends	Carriage width	Route length	human factor	vehicular factor	actual	Naive bayes
Osogbo - Ejigbo	10-01-01	600	0	2	Earth	12	5	11	7	7.3	31	DGD	Nil	Minor	Minor
Osogbo - Ejigbo	21-01-01	1000	9	1	Earth	12	5	11	7	7.3	31	Nil	BFL	Fatal	Serious
Gbongan - Ife	23-02-01	920	6	1	Paved	9	5	5	3	7.5	24.14	SLV	Nil	Fatal	Fatal
Osogbo - Gbongan	06-03-01	2300	19	1	Paved	16	2	35	15	7.5	41.02	Nil	TBT	Fatal	Fatal
Ikirun - Osogbo	09-03-01	2130	7	2	Earth	4	1	3	3	7.3	22	SLV	Nil	Fatal	Serious
Osogbo - Gbongan	12-03-01	1325	1	2	Paved	16	2	35	15	7.5	41.02	Nil	TBT	Serious	Serious
Osogbo - Gbongan	12-03-01	1325	1	2	Paved	16	2	35	15	7.5	41.02	Nil	TBT	Serious	Serious
Osogbo - Gbongan	12-03-01	45	1	1	Paved	16	2	35	15	7.5	41.02	Nil	TBT	Serious	Serious
Osogbo - Gbongan	05-04-01	1500	16	2	Paved	16	2	35	15	7.5	41.02	SLV	Nil	Fatal	Serious
Ilesa - Akure	11-04-01	1315	2	1	Earth	2	1	4	4	7.2	76	Nil	TBT	Serious	Serious
Gbongan - Ibadan	12-04-01	1200	5	1	Paved	6	2	15	5	7.5	54	SLV	Nil	Fatal	Serious
Gbongan - Ibadan	13-04-01	1122	0	1	Paved	6	2	15	5	7.5	54	Nil	TBT	Minor	Minor
Gbongan - Ife	14-04-01	1500	5	1	Paved	9	5	5	3	7.5	24.14	Nil	TBT	Serious	Serious
Gbongan - Ibadan	17-04-01	1200	2	1	Paved	6	2	15	5	7.5	54	DGD	Nil	Serious	Serious
Gbongan - Ibadan	18-04-01	1630	5	1	Paved	6	2	15	5	7.5	54	SLV	Nil	Serious	Serious
Gbongan - Ibadan	24-04-01	1130	5	2	Paved	6	2	15	5	7.5	54	SLV	Nil	Serious	Serious

## 5. CONCLUSION

The naive bayes's predictive model serves as an effective model from the analysis above and will be recommended for use to predict future road traffic accidents in order to justify results collected on the field in the future and also in projecting the possibility of road traffic accidents. Road traffic accident predictive model for predicting road traffic accidents depends on some major factors as shown in the result of this paper and the factors serves as contributing factor to the occurrence of road accidents. Road officials and road managers can understand the state of the roads and the distribution of road traffic accidents over the years in order to predict future road traffic accidents.

## REFERENCES

1. Abojaradeh, M. (2013). Traffic Accidents Prediction Models to Improve Traffic Safety in Greater Amman Area.
2. Atubi, A. O. (2009) Modelling Road Traffic accidents in Lagos State, South Western Nigeria. *Journal of Society and State*, i(i and Z): 57-74.
3. Atubi, A. O. (2010). Road traffic accident variations in lagos state, nigeria: A synopsis of variance spectra. *African Research Review*, 4(2)
4. Daramola, A. Y., (2004) "Innovative options for Financing Transport Infrastructure in Nigeria" in Niserel, the Magazine of the Nigeria Institute of Social and Economic Resource, 4 and 5, Ibadan.
5. Jonsson, T. (2005). Predictive models for accidents on urban links-A focus on vulnerable road users. *Bulletin/Lund Institute of Technology, Department of Technology and Society*, 226.
6. MANAN, W. N. B. W. (2011). *Accident prediction model at un-signalized intersections using multiple regression method* (Doctoral dissertation, Universiti Tun Hussein Onn Malaysia).
7. Mustakim, F., & Fujita, M. (2011). Development of accident predictive model for rural roadway. *World Academy of Science, Engineering and Technology*, 58(2011), 126-131.
8. Obinna, C. (2007) Road Traffic crashes kill 0.4million youths every year. *Vanguard*, April, 24
9. Ohakwe, J., Iwueze, I. S., & 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, 166-175.
10. Osita, C. (2011) "Nigeria is currently 191st out of 192 countries of the world's 'unsafe' roads". The Nation newspaper Available at ([www.thenationonlineng.net](http://www.thenationonlineng.net)) Accessed on 2nd Marchm 2013.
11. Osoro, A. A. (2011). *An analysis of the incidences and causes of road traffic accidents in kisii central district-Kenya* (Doctoral dissertation).
12. Zheng. L. and Meng. X. (2011). An approach to predict road accident frequencies application of fuzzy neural network. *Journal of Transportation Systems and Information Technology* 9(2), 87-92.