An Investigative Model On Road Accident Injury Analysis Using Data Mining Techniques – A Case Study of Lagos State, Nigeria

[1]Akinrotimi Akinyemi Omololu & [2]Mabayoje Modinat Abolore
Department of Computer Science
Faculty of Communication and Information Sciences
University of Ilorin
Ilorin, Kwara State, Nigeria
E-mail: [1]akinrotimiakinyemi@ieee.org, [2]mabayoje.ma@unilorin.edu.ng

ABSTRACT

One of the leading causes of fatal injuries and death in most countries of the world is road traffic accidents. Nigeria as a country has experienced a high rate of road accidents in the past couple of years with Lagos State having the highest share. These accidents have resulted in a lot of causalities with varying degrees of injuries and fatalities. The socio-economic development of the country has also been negatively affected. In this paper, a decision support model for traffic control within the Lagos State Metropolis is presented. In carrying out this study, traffic data was collected from the LASTMA (Lagos State Traffic Management Authority), Lagos headquarters and analyzed with a focus on the severity levels of injuries sustained by several victims of road accidents. The analysis revealed that the developed model is capable of providing a satisfactory precision in classifying different types of accidents and also important information that can be used in taking preemptive and preventive measures.

Keywords: Data mining; Accident; Injury Severity; Regression Trees; Generic Rules.

iSTEAMS Proceedings Reference Format
Akinrotimi, A.O. & Mabayoje, M.A. (2019): An Investigative Model On Road Accident Injury Analysis Using Data Mining Techniques – A Case Study of Lagos State, Nigeria. Proceedings of the 18th iSTEAMS Multidisciplinary Cross-Border Conference, University of Ghana, Legon, Accra, Ghana. 28th – 30th July, 2019. Pp 11-20. www.isteams.net - DOI Affix - https://doi.org/ 10.22624/AIMS/iSTEAMS-2019/V18N1P2

1. INTRODUCTION

Roadway traffic safety is a major concern for transportation governing agencies as well as ordinary citizens. In order to give safe driving suggestions, careful analysis of roadway traffic. datasets is critical to find out variables that are closely related to fatal accidents (Liling et al, 2017). Having a good knowledge of how information churned out of interesting patterns found in these data sets, is essential in not only making the best use of the information, but also making new discoveries in a world were data is growing at a very high rate. Preceding mining of data however, is the collection and arrangement of data in a consistent and profitable manner, as merely gathering of data is not sufficient for obtaining facts from it. Eka et al., (2018) describes data mining a term used to describe the discovery of knowledge in the database, data mining is a process that uses statistical, mathematical, artificial and machine learning technique to extract and to identify pertinent information and related knowledge from a large data base. Data mining can also be said to be a technique involving a combination of other techniques, tools and procedures used for discovering knowledge. Some of the tools employed in data mining include statistical methods and Artificial Neural Network (ANN). These tools are mainly used for improving the database and optimizing its processes. Data mining is classified into directed and undirected data mining.

In directed data mining, techniques such as virtualization, description, clustering and affinity grouping (association rules) are used to establish relationships amongst the variables in a data, while in undirected data mining, techniques such as estimation, classification and prediction are used to develop models that describe a certain variable of interest in a data, in terms of the rest of the data in the database.Furthermore, data mining has also proved to be one of the most viable tools used in detecting several useful patterns (in databases) which are capable of aiding organizations in engineering data resources for carrying out tactical planning and decision-making in their various areas of endeavor. This paper has its focus on road traffic control. Interestingly, this is one of the major areas where data collection and analysis is very crucial in not only attempting to maintain decorum and safety, but also the prevention of loss of life and properties. In this research work, adaptive regression trees have been applied in developing a model that is capable of identifying and analyzing patterns of interest in traffic data, based on the severity of injuries sustained by the victims of the accident, in order to help provide timely and accurate information which can be used to forestall further occurrences.

2. LITERATURE REVIEW

Road traffic accidents are often identified as vehicle accidents which occur on public highways. It often includes collisions between animals and vehicles, pedestrians and vehicles or fixed/moving objects and vehicles, including single vehicle accidents, that is, accidents involving only one vehicle, without any other road user (SafeCarGuide, 2014). Road traffic accidents have emerged as both a national and international problem in most countries of the world. The injury severities resulting from road traffic accidents is one of the major areas of concern, to the Lagos Sate government due to the volume of a traffic and vehicle accidents, recorded at different levels. Out of all accidents registered in Nigeria, Lagos holds about 45%.This is partly because the city encounters a very high traffic volume, every day. In addition to this, the city takes about 70 % of the traffic of registered motor vehicles in Nigeria. All these facts reveal that Lagos, having a great deal of concentration of vehicles and traffic, takes the lion's share in car accidents also. Statistical data from the traffic office shows that Lagos is experiences around 275 accidents per month. The cost of fatalities and injuries due to traffic accidents, have a great impact on various aspects of the society. It is estimated that Nigeria currently loses three billion naira every year to road crashes. Road crashes cost Nigeria 13% of her gross national product (GNP), which inhibits economic and social development (Ogwueleka and Ogwueleka, 2012).

The application of data mining in modeling records of traffic accidents can help to provide in-depth understanding of, roadway conditions, drivers' behavior and weather conditions, which are often closely associated with the different injury severities sustained by accident victims. This can assist road traffic officials in formulating better traffic safety control policies. Ramya et al, (2019) used the Random forest algorithm in comparison with an ANN and the J48 Decision Tree Classifier to identify relevant patterns and for classifying the type of accident severity of various traffic accidents. Their results indicated that Random Forest performed better than the other classification algorithms in identifying the required patterns. Kim et al., (2004) developed and used a log-linear model to clarify the role of driver characteristics and behaviors in the causal sequence leading to more severe injuries. They found that alcohol or drug use and lack of seat belt use greatly increases the odds of more severe crashes and injuries. Ossenbruggen et al., (2010) utilized a logistic regression model in identifying statistically significant factors that predict the probabilities of accidents and injuries sustained in these accidents, with the aim of using these models to perform a risk assessment of a given region. The models developed were functions of factors that describe a site by its land use activity, road side design, use of traffic control devices and traffic exposure. Their study indicated that village sites are less hazardous than residential and shopping sites.

The rate of road accidents in Lagos, Nigeria, increases with the increase of motor vehicles and the size of the population. The increase in automobile ownership, coupled with the poor condition of the roads have led to a high level of road traffic accident occurrences and occasionally results in traffic congestion problems. According to The FRSC reports, about 5117 people died, while more than 35,899 were seriously injured in 2017, as a result of road traffic accidents. In addition, the death rate is 136 per 11,031 vehicles around the country and Nigeria is losing over 3.1 billion naira yearly as a result of road traffic accident. The share of Lagos State in the total number of accidents was 41 percent in 1989, with an annual average traffic accident growth rate of 29.8 percent [FRSC Statistical Digest, 2017]. Olutayo & Eludare, 2014 used ANN and decision tree data analysis techniques in analyzing data from road accidents that occurred between the Lagos-Ibadan express road. Particularly, they used the Multilayer Perception, the Radial Bias Function (RBF) the ID3 and Function Tree algorithms. The results revealed that: of all the techniques considered, the ID3 tree algorithm performed better, producing a higher accuracy. The major causes of road traffic accidents in Lagos has been identified by the FRSC as vehicle defects, not obeying traffic laws and traffic lights, not putting on seatbelts and driver characteristics. The main objective of this research work is to study the application of data mining technology in developing a model that can support road traffic accident severity analysis, for preventing and controlling vehicle accidents in Lagos State, Nigeria.

2.1 Data mining and decision Trees

Data mining is a technique of discovering, comprehending and interpreting important information. Often, tasks are either categorized into prediction data or description data. When a particular result is expected from using a data miming technique on a set of data, the top-down approach is used. This approach is known as predictive modeling. However, when one is attempting to find certain patterns of data within a given data set, the undirected data mining is used. Undirected data mining is often carried out, when the user is interested in finding useful information from a set of patterns identified in a given set of data. This approach is known as the bottom-up approach. There are diverse problem types in a data mining process. When used to solve problems, those problems are often regarded to as data mining tasks. One of the most frequently used data mining tasks include: association, dependency analysis, classification, clustering, description, prediction and segmentation. Decision tree are versatile tools for mining data. It is often used to mine certain sets of data out of a large pool of data. A decision tree is thus made up of many nodes, networked by several branches. When a decision tree is used in predictive modeling, each leaf node in the tree specifies a test for some attributes of the instance and each branch descending from that node corresponds to one of the possible values for this specific attribute. Classification and regression problems are often solved using models, built with decision trees.

2.2 Classification and Adaptive Regression Trees (CART)

The Classification and Adaptive Regression Tree (CART) in other words known as binary recursive partitioning, is a robust data-analysis tool. It has a tremendous ability of searching through highly complex data, fro important patterns, relationships and hidden structures. It is a tree-based methodology. In decision trees, parent nodes bring about child nodes. As such, the CART technique is binary in nature as these parent nodes always produce precisely two child nodes. This process can be repeated several by treating each child node as a parent. The main set of rules used in the CART analysis as follows:

(i) Detecting when a tree is complete
(ii) Allotting a terminal node to the outcome of a class.

3. METHODOLOGY

3.1 Data collection and variable (Attribute) Selection

Collection of details about the accidents analyzed in this research work was obtained from the daily accident report form usually filled by the Lagos State Traffic Management Authority (LASTMA). It contains a detailed report of every single accident reported in it and the understanding of the content and structure of the data it renders, is crucial to carrying out a successful study. The LASTMA office performs series of updates on road accident records from time to time in order to maintain a current report. This is often done through the use of relevant software and operating systems installed on computer units connected to the internet. The data for this study was obtained from the LASTMA office, Lagos, Nigeria.

The data obtained contained accident records from July 2010 to September 2018. The attributes by which an accident can be described and as indicated in the data received from the Traffic office includes date and time, accident id, driver's name, vehicle type, driver's age, driver's gender, driver's educational level, driver's license status, vehicle defect, vehicle age, accident area, accident road name, road segment separation, road direction, road surface type, roadway surface condition, light condition, weather condition, vehicle maneuver, accident type, total vehicles involved, total number of victims, accident victims category, victims profession, victims, health condition, pedestrian maneuver, vehicle plate number, cost estimate of the damage and cause for accident.

However for the purpose of this research work, which focuses on identification of fatal and serious injuries, the following variables have been selected: driver's age, driver's sex, driver's experience, vehicle age, vehicle type, road surface type, road condition, light condition, weather condition, accident type, accident cause and injury severity. These variables are used in obtaining the specific variables used in the construction of the final dataset. Undesirable variables, which were not relevant to the purpose of the research work, were eliminated. The variables used for the classification task were selected using a heuristic approach. Their description, names and types are hereby presented in Table 1 below:

Table 1: List of variables used for classification of accidents

Attribute Name	Type	Description
Accident_ID	N	A number to identify a given accident uniquely
Driv_Age	N	Age of the driver
Driv_Sex	T	Gender of the driver
Driv_Exp	N	Driving experience of the drive
Vehic_Age	N	Service year of the vehicle
Vehic_Type	T	The type of the vehicle
Road_Surf_Type	T	The surface type of the road
Road_Cond	T	Road surface condition at the time of accident
Light_Cond	T	Light condition at the time of Accident
Weather_Cond	T	The weather condition at the time of accident
Aci_Type	T	The type of the accident
Acc_Cause	T	Causes for the accident
Injury_Severity	T	The injury severity level due to an accident

The attributes used to categorize the vehicle's driver are as follows: (i) Driv_Exp_cal (ii) Driv_Age_ Cat (iii) Vehic_Age_Cat. The Driv_Exp_Cat is derived from the base attribute "driver experience" to categorize input values between 1 and 2, 3 and 5, 6 and 10, and above 10 years as 'no experience' from driver's age attribute to class between 18 and 30, 31 and 50 are also derived from the attrib. Driv_Age_ Cat is also derived from driver's age attribute to classify the input values as less than 18, between 18 and 30, 31 and 50 and above 50 years. Similarly Vehic_Age_Cat is also derived from the attribute vehicle age to classify the input values as less than 1 between 1 and 2, 3 and 5, 6 and 10 and above 10 years. This helped to reduce the cardinality of each attribute to a manageable size so as to make the result easily interpretable. Table 2, summarizes the transformed attributes. When the pre-processing was completed, the final dataset used for modeling had 4,658 records described by 16 attributes (13 base and 3 derived). And with respect to the dependent variable, injury severity, there were 341 (7.32%) records with fatal injury, 402 (8.6303134%) records with serious injury, 709 (15.22%) records with slight injury and 3206 (68.82%) records with property damage or no injury. As to the cause of the accidents, the statistics shows that, denying pedestrian's priority, not keeping appropriate distance while driving, driving on the left side and over speeding, ranks from first to fourth respectively.

For instance from existing data set of 4,658 accident records, the above four causes comprised 21%, 19.13% , 19.02%, 16.70% percent respectively.

Table 2: Derived attributes with their base attributes and values

Attribute name	Derived attributes	Values
Driver experience	Driv_Exp_Cat	A-F
Driver age	Driv_Age_Cat	A-D
Vehicle	Vehic_Age_Cat	A-F

4. EXPERIMENTS AND RESULTS

To validate our model, the dataset was divided into two sets, which include: the training set and the validation set. 75% of the data was used for the training set, while 25% of the data was used for the validation of the set. The training dataset consists of 3492 records i.e. 75% from each class, while the validation set consists of 1165 records i.e. 25% from each class.

4.1 Experiment One

Based on a training dataset which includes 75% the data, the first task was to build a decision tree model. The variable called "Accident_Severity" is used as a dependent variable while other variables are used as independent variables. Also, feature selection was carried out, based on the contribution of the input variables in the building of the decision tree. The outcome of the first experiment revealed that the decision tree constructed included 10 attributes even though each of the records in the dataset has 13 attributes including the dependent variable; "Accident Severity". The attributes Driv_Exp_Cat and Vehic_Age_Cat are deemed as statistically insignificant. Also, has shown in Table 3 below, approximately 87% of the records were classified correctly.

Table 3: Prediction statistics of the first experiment

Results	
Total records	1165
Correctly predicted	1035
Percentage	88.46%

4.2 Experiment Two

Given the importance of each of the variables used for classification of the severity of the accident, the next crucial step involved, identifying the variables that were used for dividing (splitting) the tree. The second variable, that is, accident _id was found to have been used to split the tree at the node next to the root but since it was not deemed to be the most important variable when considering the level of injury severity, the next but most statistically significant variable was chosen. By excluding the variable accident_id, the number of variables used in the experiment reduced to nine (9). The variable was jettisoned considering its insignificance to this study. The decision tree was first splitted into 5 individual values, using the variable "accident type', out of these values, "vehicle ped" and "turn over" influenced the overall outcome of the experiment. The next most crucial variable that was used for splitting the decision tree was "accident cause". Also the decision tree nodes indicated that the values, "denying pedestrians priority" and "over speeding" were the most significant reasons behind such levels of fatality and serious injuries.

4.3 Results Assessment and Validation

An evaluation of the results was carried out on both the training and testing set and it was observed that out of the total records for testing (i.e. 1165), 42, 11, 152 and 754 records were classified correctly in the class of fatal injury, property, damage, serious injury and slight injury respectively.

On the other hand, 54 records were incorrectly classified as: serious injury (1), slight injury (22) or property damage (31) while actually they were supposed to be in the fatal injury class and 44 records were classified incorrectly as fatal injury, serious injury, and slight injury while actually they were in the property damage class. This result reveals that from the total records (i.e. 1165), 930 were classified correctly while the remaining 235 records were classified incorrectly (Table 5). Hence this indicates that records whose class is property damage were classified with minimum error as compared with the records in the class "fatal and serious injury". The result of the validation of the second decision tree and its presentation as a confusion matrix is given in tables 4 and 5 respectively, as shown below:

Table 4: Confusion matrix of injury severity based on validation results from the second decision tree

	Predicted				
Actual		Fatal injury	Partial injury	Serious injury	Slight injury
	Fatal injury	42	33	1	23
	Partial damage	7	754	3	36
	Serious injury	9	11	11	67
	Slight Injury	13	38	1	152

Table 5: Validation results from the second decision tree

Statistics	
Total Records	1165
Correctly predicted	970
Percentage	83.26%

Though the performance of the training scheme was satisfactory in order to obtain a model with a higher level of accuracy, further tests were conducted as several other decision trees were built using various numbers and combination of the selected variables. The results of the tests are displayed in Table 6 below:

Table 6: Results of the tests conducted

Number of variables used	**Accuracy**
10 (ten)	89.19%
9 (nine)	83.26%
8 (eight)	88.46%
7 (seven)	87.47%
6 (six)	85.37%

As shown in Table 6, the "best" decision tree was derived by making use of 7 variables out of the 13 variables that were used in building the model. Also, it was found out that two accident types namely vehicle_ ped and Turn_over are largely responsible for a lot of fatalities and serious injuries. The next attribute crucial to the splitting of the tree was Acc_cause. Considering the two important accident types, denying pedestrian priority, driving while under the influence of alcohol and over speeding are major causes of turnover and vehicle_ ped accident types. In depicting this, 208(53%) of 391 vehicle_ ped accidents (caused by denying pedestrians priority) resulted in either serious or fatal injury. Also 45 (85%) out of 53 accidents caused by driving while under the influence of alcohol also resulted in either fatal or serious injury. The result of the validation of the 'best' decision tree and its representation as a confusion matrix is given in tables, 7 and 8 respectively as shown below:

Table 7: Validation results from the "best" decision tree

Statistics	
Total Records	1165
Correctly predicted	1039
Percentage	89.19%

Table 8: Confusion matrix of injury severity based on validation results from the "best" decision tree.

		Predicted			
Actual		Fatal injury	Partial injury	Serious injury	Slight injury
	Fatal injury	32	31	13	17
	Partial damage	8	776	5	0
	Serious injury	5	9	39	49
	Slight Injury	3	9	8	175

The best decision tree (as shown in table 7) was able to achieve an accuracy of 89%.

4.4 Generic Rule generation from Decision Tree

The Generic rules were generated by tracing through the branches up to the leafs. The rules are derived from correlations found between the main variable (or dependent variable) and the other variables (or independent variables). Some of the rules generated from the decision tree are shown below:

Generic rules

RULE # 1
(Whole Tree)
Injury_severity = F-injury
0.0821115879828
Injury_severity = P_ damage
0.0676394849785
Injury_severity = Serious_Injury
0.0725450643777
Injury_severity = Sli_injury
0.184243497854

RULE # 2
If
Acci_Type =Turn Over
Acci_Cause = Denying pedestrians, Priority, Driving with alcohol, inappropriate
Preceding and Turning or Over Speeding

Then
Injury_severity = F-injury
0.535714285714
Injury_severity = P_ damage
0.321428571429
Injury_severity = Serious_Injury

0.0357142857143
Injury_severity = Sli_injury
0.107142857143

RULE # 3
If
Acci_Type = vehicle_Peds
Acci_Cause = Denying pedestrians priority or not keeping appropriate distance
Then
Injury_severity = F-injury
0.155405405405405
Injury_severity = P_ damage
0.00654275675676
Injury_severity = Serious_Injury
0.371621621622
Injury_severity = Sli_injury
0.466216216216216

RULE # 4
If
Acci_Type = vehicle_peds
Acci_Cause = Driving with alcohol
Then
Injury_severity = F-injury
0.333333333333333
Injury_severity = P_ damage
0.083333333333333
Injury_severity = Serious_Injury
0.208333333333333
Injury_severity = Sli_injury

RULE # 5
If
Acci_Type = vehicle_peds
Acci_Cause = Over speeding
Then
Injury_severity = F-injury
Injury_severity = P_ damage
Injury_severity = Serious_Injury
Injury_severity = Sli_injury

RULE #6
If
Acci_Type = vehicle_Peds
Acci_Cause = Denying pedestrians priority or Not keeping appropriate distance
Road_Cond = Dry
then
Injury_Severity = F-Injury

0.112781954887
Injury_Severity = P-damage
0.00751879699248
Injury_Severity = Serious_Injury
0.383458646617
Injury_Severity = Sli-Injury
0.496240601504

RULE #7
If
Acci_Type = vehicle_Peds
Acci_Cause = Denying pedestrians priority
or Not keeping appropriate distance
Road_Cond = slippery
then
Injury_Severity = F-Injury
0.533333333333
Injury_Severity = P-damage 0
Injury_Severity = Serious_Injury
0.266666666667
Injury_Severity = Sli-Injury 0.2

From the rules, it can be seen that, generally, there is a probability of 82% that the Injury_severity classification, will be fatal injury, 65.3% that it will be an injury resulting in partial damage, 72.5% that it will be serious injury and 18.42% that it will be slight injury. The result shows that most of the accidents that involve pedestrians and single vehicle turn over often leads to either fatal or serious injuries. However, it also shows that generally, most road accidents result in fatal injury. It is also evident from rule 4 that there is a high probability of an accident leading to fatalities or injuries, if the accident cause is, “driving with alcohol” and the accident type is vehicle_peds. Another important factor resulting in fatalities or injuries is “overspeeding”, especially if it is associated with vehicle_peds type of accidents.

Generally, the rules generated, reveal the probable conditions in which an accident could result in either of the injury severity classes. The rules have also indicated that the following variables are crucial in the classification of accident severity. They are, ‘accident cause’,’ accident type’, ‘driver age’, ‘road surface type’, road conditions’, vehicle type’, and ‘light condition’. The variables are placed at the higher level of the tree, indicating the significant role they play in all the experiments and their statistical significance, than other variables such as ‘sex’, ’weather condition’ and ‘accident_id’. In the analysis, it was revealed that the variables used were viable in classifying injury severity into the predefined classes namely; fatal, serious, slight and property damage. The decision of choosing the best decision tree is based on the reliability of the generated rules in addition to the number of misclassified records of different levels of injury severity.

5. CONCLUSION

The findings generated from the records and the results of the analysis carried out, can help road traffic decision makers to develop more efficient traffic safety control rules and regulations, as they provide a better understanding of a driver's behavior, the condition of the road and weather conditions, which are often related to road traffic accidents with different levels of injury severity. In this research work, different decision trees were built for analyzing road traffic accidents and the best performing decision tree (which performs the best classification),was selected based on the soundness of the rules generated, and its ability to perform accurate classification of an accident's severity.
Although an optimal accuracy of 89% was obtained for the classification accuracy, more advance techniques can be used, in obtaining better results. The results arrived at, can help provide feasible recommendations for traffic officials in making accurate decisions and strategies that are capable of reducing road traffic accidents.

REFERENCES

[1]. Liling Li, Sharad Shrestha, Gongzhu Hu, 2017, Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques. 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA) Publication Year: 2017, pp: 11 – 14.
[2]. Eka S., Kamarul J., Bushrah B., Miftachul H.,Shankar K., Andino M., 2018, International Journal of Pure and Applied Mathematics, Volume 119 No. 15 2018, 2239-2249.
[3]. SafeCarGuide, 2014, International Injury & Fatality Statistics: http://www.safecarguide.com/exp/statistics/statistics.htm
[4]. Ogwueleka F.N and Ogwueleka, T.C, 2012, Traffic Accident Data Profiling and Clustering with Data Mining Process. IOSR Journal of Computer Engineering (IOSRJCE), Vol. 6, Issue 2, 2012, pp. 14-22.
[5]. Kweon, Y. J., and Kockelman, D. M. 2003, "Overall Injury Risk to Different Drivers: Combining Exposure, Frequency, and Severity Models". *Accident Analysis and Prevention*, 35, pp. 441-450.
[6]. Martin, P. G., Crandall, J. R., and Pilkey, W. D., 2002, "Injury Trends of Passenger Car Drivers In the USA". *Accident Analysis and Prevention*, 32, pp. 541-557.
[7]. Tavris, D. R., Kuhn, E. M, and Layde, P. M. 2001, "Age and Gender Patterns in Motor Vehicle Crash injuries: Importance of Type of Crash and Occupant Role". *Accident Analysis and Prevention*, 33, pp.
[8]. Ramya S., Reshma S., Dhatri V., Saroja Y., Gandhi G., 2019. Accident Severity Prediction Using Data Mining Methods, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Volume 5, Issue 2, ISSN: 2456-3307.
[9]. Bedard M., Gordon G., Micheal S., Hirdes J., 2002,The independent contribution of driver, crash, and vehicle characteristics to driver fatalities, Volume 34, Issue 6, November 2002, Pages 717-727.
[10].Sohn, S. Y., and Lee, S. H. 2003, "Data Fusion, Ensemble and Clustering to Improve the Classification Accuracy for the Severity of Road Traffic Accidents in Korea". *Safety Science*, 4(1), pp. 1-14.
[11].Kim, K., Nitz, L., Richardson, J., and Li, L. 2004, "Personal and Behavioral Predictors of Automobile Crash and Injury Severity". *Accident Analysis and Prevention*, 27(4), pp. 469-481.
[12].Tesema T., Ajith A. and Crina G. 2005, Rule Mining and Classification of Road Traffic Accidents Using Adaptive Regression Trees, International Journal of Simulation: Systems, Science & Technology 6(10) · September 2005.
[13].Ossenbruggen, P.J., Pendharkar, J. and Ivan, J. 2010, "Roadway safety in rural and small urbanized areas". *Accidents Analysis and Prevention,* 33 (4), pp.485–498.
[14].Olutayo A., Eludare A. 2014,Traffic Accident Analysis Using Decision Tree and Neural Networks, I.J. Information Technology and Computer Science, 2014, 02, 22-28 Published Online January 2014 in MECS (http://www.mecs-press.org/)
[15].FRSC Statistical Digest, 2017, 2nd Quarter Edition, https://frsc.gov.ng/fdg.pdf.