IOSR Journal of Computer Engineering (IOSRJCE)
ISSN: 2278-0661, ISBN: 2278-8727 Volume 6, Issue 2 (Sep-Oct. 2012), PP 14-22
www.iosrjournals.org

Traffic Accident Data Profiling and Clusteringwith Data Mining Process

Ogwueleka, Francisca Nonyelum[1]and Ogwueleka, Toochukwu Chibueze[2]
[1]*Department of Computer Science, Federal University of Technology, Minna*
[2]*Department of Civil Engineering, University of Abuja*

Abstract: *Some of this traffic accident crisis in Nigeria is caused by the increase in number of vehicles and inefficient drivers on the road, bad condition and poor maintenance of the roads. The significance of the study lies on the profiling of clusters of traffic roads in terms of accident related data and the degree in which these accident characteristics are perceptive between the different created clusters. Applying data mining process to model traffic accident data records helped in obtaining the characteristics of drivers' behaviour, road condition and weather condition that are connected with different injury severities and death. The traffic roads are divided into a low accident risk and high accident risk traffic roads, determining accidents in different age categories and period of accidents. A design of a data mining model for analysis and prediction of accidents rate in Nigeria was presented. In this study, we profiled traffic roads, differentiated the data set into pre-processing and transforming data set; created the association rules; and post-processed the frequent accident item sets. The data mining function was used and data cleaned using feature selection.*
Keywords: *Data mining, road accident, profiling, vehicles, clusters, traffic road*

I. Introduction

The amount of accident data stored in Nigeria's Federal Road Safety Commission (FRSC) database grows every twelve months at a rate of 100%, which shows that a lot of data are obtained and still more data are still being collected. Data mining is a useful tool to tackle the need for sifting useful information such as hidden patterns from databases; especially now the accumulation of data is increasing in an alarming rate. Knowledge of remarkable patterns of data is important subjects to be considered to regulate strategies, make proper and maximum use of data, and then discover latest opportunities. Data from many sources are gathered and organized in a consistent and useful manner for effective learning and result. The collected data must also be analyzed, understood, and turned into suitable information. This is the stage where the application of data mining is needed.

A road traffic accident is defined as any vehicle accident occurring on a public highway. It includes collisions between vehicles and animals, vehicles and pedestrians, or vehicles and fixed obstacles. Single vehicle accidents, which involve a single vehicle that means without other road user, are also included (Safecarguide, 2004). The cause that affects traffic accidents can be a factor or combination of many factors. The basic factors, which cause or increase the severity of probable accidents, are driver's behaviour, vehicle features, highway characteristics, environmental effects and traffic characteristics (Ozgan, 2003).

Data mining is a process that extracts knowledge by analyzing data to discover hidden patterns and dependencies in the database (Hand et al, 2001; Berthold et al, 2003). Applying data mining techniques to model traffic accident data records can help to understand the characteristics of drivers' behaviour, road condition and weather condition that are connected with different injury severities and death. This can help decision makers to formulate better traffic safety control policies.

Traffic volumes of number of vehicles per day and road lengths in kilometre are the most important explanatory variables in an accident prediction model, both for road sections and intersections. The parameters of the model can vary between road types and countries because road characteristics, user behaviour and vehicle types can differ. An accident prediction model is a mathematical formula describing the relation between the level of existing road safety such as crashes, injured, fatalities etc. and variables that explain this level such as road length, width, and traffic volume.

RTA is endemic in Nigeria, with seasonal epidemics (Asogwa, 1978). The police records show that between 1980 and 1989, a total of 98,168 Nigerians died and 244,864 were injured in RTA (NPFM, 1989). These averaged about 10,000 killed and 25,000 injured annually (NPFM, 1989). A study has shown that 1 out of 3 and 1 out of 9 of the Nigerian population stand the risk of getting injured or killed respectively, on a yearly basis from RTA (Ezenwa, 1986). In spite of the concerted effort at reducing the rate of RTA, Nigeria still ranked amongst the highly affected nations of the world (Agunloye, 1988). It is even more disturbing to note that the young adults group, the economic backbone of a nation, is most affected by this ravaging menace (Oyemade, 1973; Asogwa, 1980).

Some studies done on data mining and its techniques used in modeling or predicting accidents are the used of data mining techniques in predicting driver's behaviour so that unsafe actions can be rectified (Krishnaswamy et al, 2005); Pande and Abdel-Aty (2006) applied data mining techniques to predict rear-end crashes on highways and warn drivers about potential crashes 5-10 minutes prior to the crash; Sohn et al. (2003) applied data fusion, ensemble and clustering to improve the accuracy of individual classifiers for two categories of severity (bodily injury and property damage) of road traffic accident; Cameron (1997) indicates that clustering methods are an important tool when analyzing traffic accidents as these methods are able to identify groups of road users, vehicles and road segments which would be suitable targets for countermeasures; and in Ng et al (2002), a combination of cluster analysis, regression analysis and Geographical Information System (GIS) techniques is used to group homogeneous accident data together, estimate the number of traffic accidents and assess the risk of traffic accidents in a study area.

From a statistical perspective it can be viewed as a computer automated exploratory data analysis of large complex data sets (Friedman, 1997). Data sets can be much larger than in statistics and data analyses are on a correspondingly larger scale (Hosking et al., 1997). Statistical models are preferable when fairly simple models are adequate and the variables can be identified before modeling, but when dealing with a large and complex data set of road accidents, data mining methods are more useful to identify the relevant variables towards a better understanding of accident circumstances.

In this study, data mining process was used to analyze and model traffic accident data records for obtaining the characteristics of drivers' behaviour, road condition and weather condition that are connected with different injury severities and death; and also in profiling of clusters of traffic roads for accident related data and the degree in which these accident characteristics are perceptive between the different created clusters.

II. Analysis Of Road Accident In Nigeria

In the late fifties and early sixties, there was little or no concern about road safety matters in Nigeria. Insignificant awareness and concentration was given to road traffic crashes, prevention strategies and remedies; as economic activities were very low, the incidents of road traffic crashes were invariably low in relation to low vehicle volume plying the equally scanty routes in the country. Mostly, railways or waterways undertook haulage. The dawn of civilization brought about urbanization, education, and increase in infrastructure development. These developments in economic activities gave rise to numerical increase in vehicular statistics to meet up with the sharp demand of a developing economy.

Osime et al. (2006) observed in their study that between 1970 and 1979, there were 285 699 reported cases of RTA, while there were 57,136 deaths, representing 20%. On the other hand, the period between 1990 and 1999 recorded 188,012 cases of RTA with 76,870 deaths, accounting for 41%. There may be several explanations for this. The oil boom in Nigeria was experienced in the 1970s and many people were possibly financially empowered and so were able to buy cars. Also during this period, the value of the Naira was high. Most of the roads were also in good repair. But because not many people had been exposed to the use of cars before then, there were many cases of RTAs. Also, because the experience was relatively new, it could be assumed that there were few cases of speeding. Thus, with the good roads and reasonable speed on the part of the drivers, even though there were many cases of RTAs, the deaths were just about 20%. However, a sharp contrast is observed during the period between 1990 and 1999. Even though the number of RTAs has declined relatively, the mortality rate increased to 41%. This is the period when the country started to observe an economic downturn. Consequently, most people were no longer able to buy new cars, instead there was a high demand for used vehicles with the attendant consequences of increasing fatalities on the roads. Also, because of the poor economy, most roads were now in a state of disrepair becoming more prone to fatal accidents. It was also within this period that the health-care facilities witnessed utter neglect and were barely able to handle cases resulting from RTAs.

Poor road structure and population growth have greatly led to increase in the roads and accident rate. To reduce the accidents, the Federal Road Safety Corps (FRSC) was established by the government of the Federal Republic of Nigeria in 1988 (vides Decree 45 of 1988 as amended by Decree 35 of 1992, with effect from 18th February, 1988). The Commission was charged with responsibilities for, among others, policymaking, organization and administration of road safety in Nigeria. FRSC Corps Marshal and Chief Executive has 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 (Idoko, 2010). Nigeria loses about 3% of GDP from Road Traffic Crashes, that is, about 17% of current National reserves. Income lost from 2009 road traffic collision in Nigeria was more than the GDP of over 20 individual African countries.

In Nigeria, human factor is believed to constitute about 85% of the recorded causes of road traffic accidents. Various researches conducted in Nigeria are of the consensus that accidents through the human factor results from drunk driving, drugs, poor driving skills, health problems psychological problems and temperament. These manifest in different ways among drivers. In a study Ogunsanya and Waziri (1989) identified the human/driver related factors as the single most important contributory factor to the increasing tide of traffic accident in Nigeria. Such behavioural attitude of the drivers that had earned him this reputation include: cutting corners, sleeping on steering and fatigue, faulty trip preparation; ignorance of the highway code, driving under the influence of alcohol; uncorrected bad eye sight; refusal to use seat belts, inability to handle emergency, wrong signalling, overtaking and incompetent manoeuvring.

A study of road accident trends in Nigeria between the period 1960 and 1989 revealed a sharp increase in fatal accident occurrence. Between 1960 and 1969, it was observed that over 18,000 deaths occurred as a result of road accidents. By the third decade (1980–1989), this figure had increased to about five times, that is, more than 92,000 deaths **(**Oluwasanmi, 1993**).** The road crash situation in Nigeria is total of 969,850 reported cases of road crashes from 1960 to 2004, as well as 275,178 person killed with 843,691 people injured during the same period. There is a dramatic reduction in the rate of road crash from 1988. Analysts have ascribed this improvement to the activities of the Federal Road Safety Commission in Nigeria since 18th February 1988. The mandate of the Federal Road Safety Commission is to reduce the rate of traffic crashes to the minimum since at this moment, zero vision has not been contemplated in Nigeria. Enforcement and raising awareness are the main strategies adopted by the Commission as the means of promoting road safety in Nigeria.

The number of wounded and dead people is currently increasing rapidly in Nigeria. If the present trend continues, by the year 2020 road traffic accidents will be the second most noted cause of fatalities in the world. Traffic accidents gave rise to large economic costs, deep human suffering and tragedy. This development is serious since traffic accidents can be prevented. A road traffic system that is sustainable in the long term may be achieved if the traffic safety work is developed and intensified.

According to statistics obtained from Federal Road Safety Corps (FRSC) in Abuja, about 4,120 persons lost their lives while 20,975 others were seriously injured in fatal accidents that involved 11,031 vehicles across the nation in 2009. The FRSC recorded 11,341 accidents with total number of deaths put at 6,661 and 27,980 injured in 2008. Also, between January and June 2010, 5,560 cases of auto accidents were recorded in which 3,183 persons lost their lives and 14,349 others sustained various degrees of injuries.

The road traffic accidents (RTA) data obtained from the Federal Road Safety Commission and Nigerian Police Force was from 1998 to 2010. In 2003, there were 16,795 reported cases of road traffic accidents; 28,215 injured persons and 8672 deaths.Table 1: A graphical representation of the reported number of road traffic accident cases from 1998 to 2010,

Table 1: A graphical representation of the reported number of road traffic accident cases, the injured and deaths from road traffic accidents (RTAs).

Year	Total cases reported	No. of persons killed	No. of persons injured
1998	17,117	6578	17,547
1999	12,503	5953	18,000
2000	12,325	6336	20,555
2001	15,621	7845	26,745
2002	16,452	8452	27,102
2003	16,795	8672	28,215
2004	14,279	5351	16, 897
2005	8962	4519	15,779
2006	9114	4944	17,390
2007	9132	4916	20,944
2008	11,341	6661	27,980
2009	11,031	4120	20,975
Jan-June 2010	5,560	3183	14, 349

III. Methodology And Results

In this study, we profiled traffic roads, differentiated the data set into pre-processing data set and transforming data set; created the association rules; and post-processed the frequent item sets. The data mining function was used. Data was cleaned and using feature selection, we selected the required features necessary for the data set and the categorization task. The data was modified and aggregated.

The data mining approach involves steps, which is illustrated in Figure 1 (Ogwueleka, 2011):

1. Select an appropriate algorithm
2. Implement the algorithm in software
3. Test the algorithm with known data set
4. Evaluate and refine the algorithm as you test with other known data sets
5. Publish the results.

The four basic steps, which were used in order to complete this data mining implementation on road traffic accident. They are data selection, data transformation, applying algorithms and results interpretations. The objective of data selection was to determine the type of information and the way it is organized. Some part of the road accident data available from the source data file was required and it helped in quick collection of relevant data after identification. The required road traffic data was sampled and the sample also mined. After data selection, the data was transformed, and then algorithms were applied in order to extract the required information, which will aid in achieving the required objective. The result of applying data mining algorithms was interpreted and also analyzed using a visualization and decision support tool. The designed model was validated and tested. There was need in some aspect to refine the data and repeat the process sequence again.

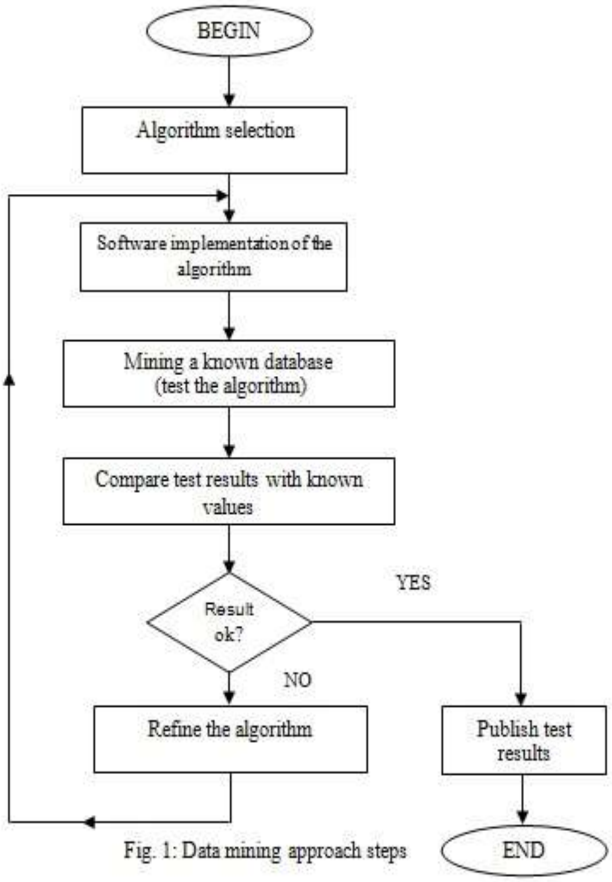

Fig. 1: Data mining approach steps

3.1 Data collection

Types of information that the road safety office records about an accident include driver and vehicle characteristics, road and weather conditions, date and time of the accident, type, injury severity and possible causes of such accidents. The specific attributes by which a given accident can be described are date and time, accident identity, driver's name, vehicle type, driver's age, driver's gender, driver's educational level, driver's license status, relation of the driver and vehicle, driver's experience, possession of the vehicle, 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. In addition to the input variables mentioned above the output variable for this research that is injury severity is also another attribute of a given accident. The target attribute, injury severity, has four classes: fatal injury, property damage (no injury), serious injury and slight injury.

These categorizes of data were collected at the Federal Road Safety Corps (FRSC) and Nigerian Police Force (NPF) offices during data collection and used in the analysis.

3.2 Profiling Traffic Roads

We clustered and profiled each traffic road item. We used the two components covariance model, which groups the traffic roads into two clusters. These clusters shows different results for the general accident risk on the roads, as it is not every accident variable that is important when describing the different set of traffic roads. A comparative analysis between the accident characteristics of the different clusters is conducted, which provides new insights into the complexity and causes of road accidents. The data set was differentiated into pre-processing data set and transforming data set. These two steps used in the data set mining process were for the preparation of the data obtained, for generating frequent item sets and in used in the post-processing step for pattern decoding and evaluation.

3.2.1 Pre-processing and modifying the data set.

The data set of road traffic accidents was divided into two clusters depending on the type of traffic road. In the first cluster, we got fifteen (15) traffic roads with forty (40) records of road traffic accident; and in the second cluster, for the eight (8) traffic roads, one hundred and twenty-five (125) road traffic accident records was obtained. The discriminating character of the accident characteristics for the two clusters was established through result comparison of the two analyses.

3.2.2 Traffic roads profiling.

Each cluster of traffic road was profiled and grouped into two clusters displaying the different results for the whole accident risk on the roads. As all variables in all accidents are not so significant in determining the various groups of traffic roads, we conducted an analysis between the accident characteristics of the different clusters, which provided the possible solution into the complexity and causes of road accidents.

3.2.3 Creating the rules.

In choosing the association rules, a lot of issues were considered, for example, choosing low minimum support rules can lead to poor use of result due to the computer memory margins and if a too high support parameter is chosen, the resulting algorithm will only be capable of generating trivial rules. Therefore, in the analysis, minimum support value thirty-five (35) was used.

The generated algorithm obtained 30520 frequent item sets with a set minimum support of thirty percent (30%) of maximum size of 5 in the first cluster, while in the second cluster, 29055 frequent item sets were obtained with the same maximum size and minimum support.

3.2.4 Post-processing the frequent accident item sets.

The item sets that are frequent for both groups of accidents were chosen and the accident patterns chosen are descriptive for the first cluster and the second cluster, resulting to 25558 accident patterns. The discriminating character of these accident patterns was determined by means of a specific interest measure. The frequent items having positive interest value of "1" was chosen and used. Analysis showed that the second cluster has stronger accident patterns and is also the cluster with the highest accident risk. I > 0 was used for selecting the item sets and 5 resulted in 20 item sets of size 305 item sets of size. For 325, we got item size of 4. The first and second cluster in accidents for frequent item sets is shown in Table 2.

Table 2: First and second cluster in accidents for frequent item sets in town

Item 1	Item 2	Item 3	Item 4	S_1	$Lift_1$	S_2	$Lift_2$	I
Weekday	Town	2 road users		45,16%	0,94	71,96%	1,02	0,37
Weekday	Road user age <18 and >29			38,71%	0, 92	59, 81%	1,02	0,35
Dry road surface	Town	Weekday		38,70%	0,84	57, 94%	0,99	0,33
Straight direction	Town	Normal weather	Weekday	32,25%	0,74	52,33%	0,97	0,38
Side collision	Female road user	Normal condition	Weekday	35,48%	0,88	55,14%	1,02	0,35
Side collision	Town	Weekday	Car	38,70%	1,03	57,94%	1,09	0,33
Straight direction	50km/hr	Daylight		32,25%	1,10	46,72%	1,06	0,30

Results of table 2 show that the accident patterns that occur more frequently in cluster 2 than cluster 1 often occur on a weekday, inside the built up area (town) with 2 road users [N=1], on a dry road surface with the road user's age <18 and >29. Accidents involving a female road user, driving a car is noted to be during the weekday and in normal condition with I = 0,35. Also, a road user frequently driving in a straight direction with a speed limit of 50 kilometers per hour in daylight, during normal weather on a weekday has I = 0,30. It was noted that accidents in sideways collision are during the weekday in normal condition.

The data in the databases are collected from Federal Road Safety Commission, Abuja-Nigeria. The data are pre-processed through duplicate removal and supplying of missing values. Data was transformed to a format suitable for clustering because data elements with high similitude are bunched up within same cluster while data elements with less similarity are crowded up in different clusters through unsupervised classification of patterns into groups. To prepare the data suitable for mining process, data needs to be transformed accordingly. This pre-processed data is refined further to convert it to a format compatible for clustering. Clustering was performed with k-means algorithm.

3.3 Data Mining Function

All accident record in the data set is an input/output pair with an associated output. The output variable, which is the target feature is categorical and has four classes made up of fatal injury (deadly injury), property damage (no injury), serious injury (severe injury) and slight injury (minor injury). A supervised learning algorithm was used to map an input vector to the preferred output class. In this supervised learning algorithm, an input vector is presented at the inputs together with a set of desired responses, one for each variable for the output.

3.4 Data Cleaning

In any large volume of data set, there is always possibility of errors as most collected data are not clean and constitutes different features, some of which are not needed in the data mining process. During data cleaning, some features in the traffic records were removed. These features include license class of the driver; ownership of the vehicle; name and section of the road; category, occupation, accident day, estimate of the damage cost; driver's educational level; health condition, relation of the driver and the vehicle; total number of vehicles involved and their plate number; accident areas and road direction; accident date and time; and vehicle fault.

3.5 Feature Selection

After cleaning the data, some features were removed. In selecting the required features necessary for the data set and the categorization task, an approach was used approach to ensure that all selected feature is significant. The selected features with their data type and description are shown in Table 3. For all selected feature that is numerical, N is used while T is used for text features.

Table 3:Selected features with their data type and description

S/N	Feature Name	Data Type	Description
1	Accident_ID	N	Identification number used for a specific accident
2	Driv_Name	T	Driver's name
3	Driv_Age	N	Driver's age
4	Driv_Exp	N	Driver's years of driving experience
5	Driv_Gender	T	Driver's Gender
6	Vehi_Type	T	Vehicle type
7	Vehi_Age	N	Vehicle service year
8	Road_Cond	T	Condition of road at the time of accident
9	Weather_Cond	T	Weather condition at accident time
10	Acci_Type	T	Type of accident
11	Acci_Cause	T	Cause of accident
12	Inj_Type	T	Injury type class incurred from accident

3.6 Data modification and Aggregation

Data was modified; the features derived were produced and categorized using different feature names. In the driver's experience classification is denoted as Driv_Exp and it is the base feature of driver's experience categorizing the input values into three as no experience, between the range of 1 to 10, and above 10 years.

The driver's age is featured as Driv_Age category. It classified the input values as 18, between 18 and 30, 31 and 50 and above 50 years; while Vehic_Agewas derived from the feature of 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. The data feature modification, aggregation and categorization helped to condense the features to reduced size for accurate and reliable result. The modified features are shown in Table 4.

Table 4: Derived features with their base features

Feature Name	Derived Feature
Driver's experience	Driv_Exp_Cate
Driver's age	Driv_Age_Cate
Vehicle's age	Vehic_Age_Cate

The final dataset used for modeling had 3, 640 records expressed by 15 features of 12 base features and 3 derived features. Using the dependent variable, injury type, the records obtained are fatal injury is 428 (11.7%); serious injury is 400 (11%); slight injury is 695 (19.1%); and no injury is 2110 (58%).

IV. Discussion

Comparing 2004 and 2005 RTA summary from Table 5, there is an observed 37% reduction in total RTA cases; 16% reduction in the number of persons killed; and 7% reduction in the number of persons injured.

Table 5: Summary of 2004 and 2005 RTA

Year	2004	2005	Remarks
Total reported RTA cases	14279	8962	37% reduction in RTA
Persons killed	5351	4519	16% reduction in the number of persons killed
Persons injured	16897	15779	7% reduction in the number of persons injured

Table 6(a): 2005 RTA summary — Table 6(b): 2004 RTA summary

Month	Total cases reported	Persons killed	Persons injured	Month	Total cases reported	Persons killed	Persons injured
Jan	980	447	1760	Jan	1333	544	1550
Feb	1040	358	1267	Feb	1039	397	1572
March	591	359	1178	March	1312	498	1705
April	679	450	1373	April	1498	481	1458
May	1118	556	1567	May	1206	440	1459
June	841	381	1291	June	1144	476	1661
July	438	276	1045	July	487	233	761
Aug	956	399	1518	Aug	1136	365	1354
Sept	630	289	955	Sept	829	963	428
Oct	376	224	1117	Oct	1925	383	1279
Nov	624	325	1128	Nov	1036	455	1526
Dec	689	455	1580	Dec	1104	542	1598
Total	**8962**	**4519**	**15779**	**Total**	**14049**	**5777**	**16351**

From the comparison of 2004 and 2005, it was observed that 2004 has higher reported RTA cases, number of persons killed and also in number of persons injured.

Fig. 2 presents the trend in RTA from 2000 to 2007. The total number of RTA in 2007 was 12,038 representing an increase of 3.2% over that of 2006 and a 2.8% increase over the 2000 figure.

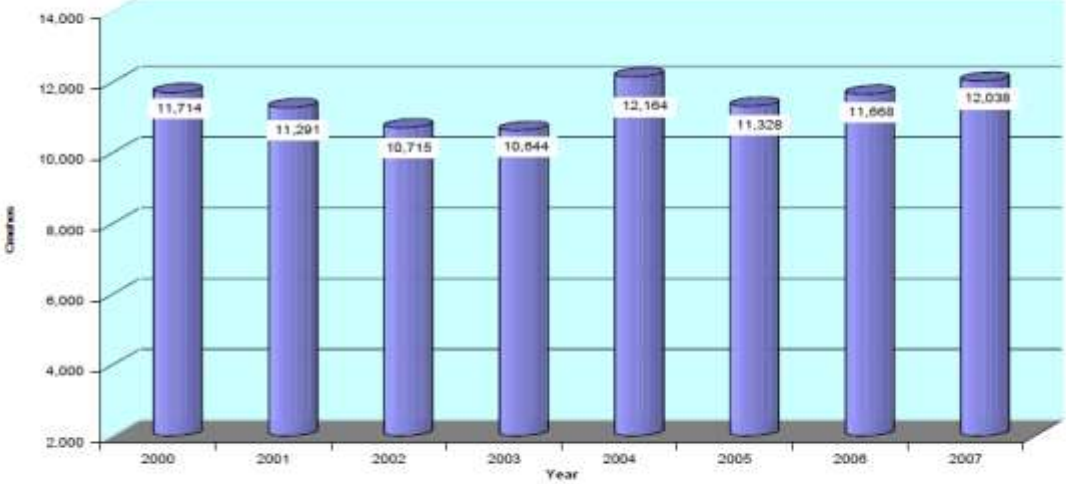

Fig.2: RTA 2000 to 2007

From the analysis of 2004 against 2005 and also from 2000 to 2007, the RTA target verses actual from 2001-2015 were obtained through prediction from collected data. Fig.3 shows the predicted RTA for 2010 to 2015.

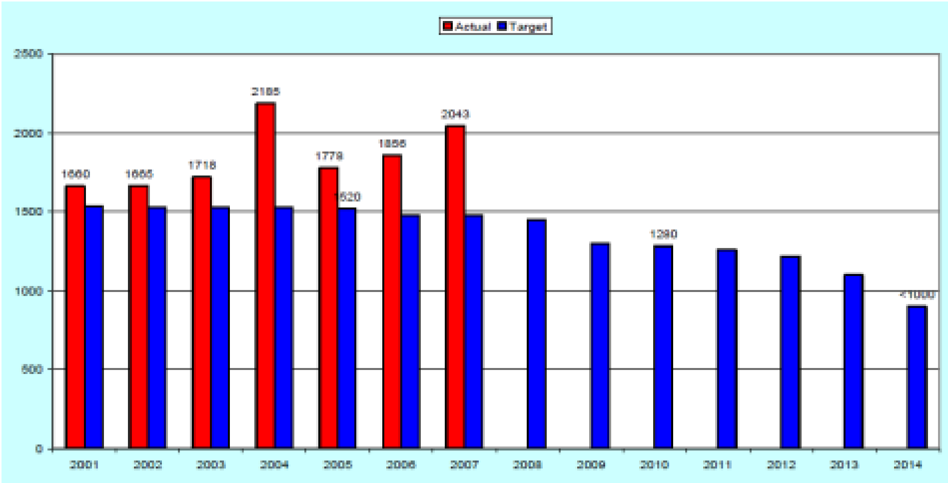

Fig. 3: RTA fatalities target vs. actual 2001 to 2015

Since the dataset of this experiment has an unbalanced dependent class, injury severity, using a random sample will most likely result the minor class belonging in only one of the partitions. In order to overcome this problem the balanced partitioning option was used. To validate model, a dataset is usually divided into two partitions. One is used for training, the learn partition, and the other is held back for testing, the validation partition. The dataset was divided into the training set (75%) and validation set (25%). After partitioning the training dataset contained 3493 (75% from each class) records and the validation set contained 1165 (25% from each class) records.

V. Conclusion

In this study, the association algorithm was used on a data set of traffic accidents to profile the two clusters of traffic roads. The analysis showed that by generating frequent item sets the identification of accident circumstances that frequently occur together is made possible leading to better clarification and accurate analysis of the occurrence of traffic accidents. The used frequent item sets describes the co-occurrence of accident conditions and also gave direction to more information and guidance helping road traffic agencies or governments to adapt their traffic policies to the different kind of accident circumstances. The results of this study indicate that the use of the association algorithm do not only allow a descriptive analysis of accident patterns within one cluster, but also create the possibility to find the accident characteristics that are distinguish the two classified traffic roads. For the traffic roads with a low accident risk, crossroads with priority to the right is one of the accident problems while in the results for the high accident risk traffic roads, these accidents are observed to occur in different age categories and also during the weekends. This research analysis shows that traffic policy towards these clusters should be considered, since each cluster is characterized by specific accident circumstances, which require different measures to enhance the traffic safety.

References

[1] Asogwa, S. E. (1978). Road traffic accidents: A major public health problem. Public
[2] Health 92:237-45.
[3] Asogwa S. E.(1980). Some characteristics of drivers and riders involved in road traffic accidents in Nigeria. East Africa. Med Journal. 57:399-404.
[4] Agunloye O. (1988). Road danger in Nigeria- Truth and consequences. J Fed Road
a. Safety Commission, Nigeria. 1:11-6.
[5] Berthold M. and Hand D. J. (2003) 'Intelligent Data Analysis', Springer, 2nd edition.
[6] Cameron, M. (1997) Accident Data Analysis to Develop Target Groups for Countermeasures. Monash University Accident Research Centre, Reports 46 and 47.
[7] Ezenwa A. O. (1986). Trend and characteristic of RTA in Nigeria. J Roy Soc Health 1:27-9
[8] Friedman, J. H. (1997) Data mining and statistics: What's the connection? Proceedingsof the 29th Symposium on the Interface Between Computer Science and Statistics, Texas.
[9] Hand D. J, Mannila H., and Smth P. (2001) 'Principles of Data mining', The MIT press.
[10] Idoko Clement (2010). FRSC in fresh war against death on Nigerian roads - Begins operation zero tolerance. Nigerian Tribune.
[11] Krishnaswamy S., Loke S. W, Rakotonirainy A., Horovitz O., and Gaber M. M. (2005) 'Towards situationawareness and ubiquitous data mining for road safety: rationale and architecture for a compelling Application', Proceedings of Conference on Intelligent Vehicles and Road Infrastructure.
[12] Ng, K-S, Hung, W-T and Wong W-G (2002) An algorithm for assessing the risk of traffic accidents, Journal of Safety Research, 33 387-410.

[13] Oyemade A. (1973). Epidemiology of road traffic accidents in Ibadan and its environs.
[14] Niger Med J. 3:174-7.
[15] Oluwasanmi Adeolu J. (1993). Road accident trends in Nigeria. Accident analysis and prevention. Vol. 25, Issue 4. Pages 485-487.
[16] Ogwueleka F. N. (2011) Data mining application in credit card fraud detection system. Journal of Engineering, Science and Technology. Vol. 6. No. 3, pp 311-322.
[17] Osime, O. C., Ehikhamenor, E., Oludiran, O. O., Iribhogbe, P. E., Ighedosa, S. E. and Elusoji, S. O. (2006). Road traffic accident trends in Nigeria, International Journal of Injury Control and Safety Promotion, 13:4, 251 – 253.
[18] Pande, A., and Abdel-Aty, M. (2006). Assessment of freeway traffic parameters leading to lane-change related collisions. Accident Analysis and Prevention. 38, 936–948
[19] 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, Vol. 4, Issue1, pp. 1-14.
[20] SafeCarGuide, 2004, International Injury & Fatality Statistics: http://www.safecarguide.com/exp/statistics/statistics.htm
[21] The Nigeria Police Force Management services and Research department. (1989). Data on establishment of force. Force Headquarters, Lagos. pp. 1-19