Indian Journal of Science and Technology, *Vol 9(15), DOI: 10.17485/ijst/2016/v9i15/92119, April 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

What You Eat Matters Road Safety: A Data Mining Approach

Preeti Mulay* and Selam Mulatu

Department of Computer Science and Engineering, Symbiosis Institute of Technology, Pune - 412115, Maharashtra, India; preeti.mulay@sitpune.edu.in, selam.mulatu@sitpune.edu.in

Abstract

Background/Objectives: To assess the influence of dietary habit on driving performance of a driver and its impact on road traffic accident to ensure road safety. **Methods/Statistical Analysis**: A total of 573 accidents data from year 2012 is used and a data mining approach, Association rule mining technique is applied to build a pattern to assess the influence of dietary habit on driving performance of a driver and its impact on road traffic accident by integrating traffic accident; Fatality Analysis Reporting System (FARS) and nutritional data; National Health and Nutrition Examination Survey (NHANES) from US. **Findings:** The result indicates that driver's dietary habit can be one of the factors to road traffic accident along with driver's license type and status, gender and type of road. The trained rules shows out of 573 traffic accidents 573(100%) related to drivers who had a habit of having non combination of food, 520(90%) drivers was having non commercial and suspended driving license, 299(52.1%) of drivers were owner of the vehicle, 238(41.5%) has habit of non combination of food, 290(50.6%) were male drivers, 252(43.9%) accidents has relation to non drunk driver, 234(40.8%) occurred in non junction and 270(47.1%) intersection point. **Applications/Improvements:** The results can be used by transportation departments to devise best strategies to ensure road traffic safety and protect the insurance business from losses of great amount of money.

Keywords: Association Rule Mining, Drivers Dietary Habit, Data Mining, Traffic Accident, Weka Tool

1. Introduction

Traffic accident is among the factors which may lead a human being to physical, psychological and economical crisis. Now days every second the number of injured persons and property damage is increasing dramatically due to increase in number of motor vehicles, population, and carelessness of road users and violation of traffic rules, apart from loss of life and property. It leads insurance companies to spend a great amount of money. Due to unbalance between profit and expenditure, so many insurance companies are out of business and the perception of investors to involve in insurance business is declining. If the research study finds out the various causes of road safety, internal to the driver and external which is out of control of the driver then it will be easier to avoid or prevent fatal accidents on road, reducing personal casualty and property loss[1]. "Driver faults" are one of the causality of traffic accident it has strong relationship with change in driving attitudes and behavior. Age is an important factor for driver's change in attitudes and behaviors'[2].

Dozens of researches has been done to find out factors related to road traffic accidents by finding out relationship between factors either using statistical model or by applying different Data Mining techniques. The authors in[3] suggested a new real time crash prediction model using Bayesian Network as a modeling method and clustering method to find out road accidents which share the same crash risk factor by integrating historical data; contains detailed information about accidents such as time and location of accident which occurred for 10 years in the city of Calgary and road geometry information and real time data; traffic update information, weather information and road conditions along with PARAMICS

**Author for correspondence*

simulator simulated data. Authors of[4] also did a quality research on factors affecting crash on road. Their study focuses on roads of Italy, crash data period 2003-2008, standard roundabout geometric design data and traffic data is merged together and Association rule mining is used to analyze interdependencies of factors affecting the crash on road and crash types. Road design according to geography, other designs including Geometry, radius, curvature of road with deviation angle are the important factors causing crash, according to rules generated by Orange Canvas tool.

This study in[2] focused on the influence of age for change and driving attitudes and behaviors and its relation to unsafe behaviors which is the cause of traffic accidents, the authors used data from AAA Foundation Traffic safety culture and analyzed the relationship between age and driving attitudes and behaviors among older Americans using regression model to analyze the degree and nature of the variability in driving behaviors and safety related attitudes among drivers age >=65. The results indicate younger drivers engage in unsafe traffic safety behaviors such as speeding and ignorance of speed cameras compared to older drivers, drivers age range 65-69 most likely read and text messages while driving on the other hand drivers age 75 and above support traffic safety cultures. In this study in[5] the authors developed classification model to construct DSS using adaptive regression trees for injury severity levels caused by road traffic accidents which gives analysis of road traffic accidents for Addis Ababa city. Relation of traffic accident with spatial factors is studied using real accident data and simulated data through Spatial Decision Tree (SDT) and Conventional Decision Tree (CDT) approach. Spatial Decision Tree (SDT) approach is founded better than CDT[6].

Japan institute for traffic accident research data and Statistical model is used to analyze the involvement of drivers to Motor Vehicle Crash (MVC) and severity of injury caused by them to other vehicle occupants based on age. The result indicates age of the driver is one of the factors to cause Motor Vehicle Crash and level of severity differs in different age groups[7]. In this study of[8] General Linear Model (GLM) is used to identify common trends in serious road traffic accident injuries based on different types of road users through period of time using data obtained from hospital admission and police datasets of Great Britain from 1996-2003. The result obtained implicate there is difference in road traffic accident causality practices. The authors in[9] suggested piezoelectric sensors which can help to improve pedestrian's safety which will be embedded on highway roads. The efficient updatable clustering algorithm based on incremental learning approach proposed in[10,11] can be applicable to build patterns in efficient way for large and dynamic learning data such as traffic accident.

Driving a motor vehicle leads a human body to engage in both physical and mental exercise that requires burning a large amount of calories. Food is the source of energy to a human body, having healthy food helps us to obtain balanced nutrients necessary for health[12]. Through food intake human being gains enough amount of energy which has impact on driving activity. The amount of food consumed, the time of food consumed, which category of road the driver is driving with which speed etc., plays vital roles related to accidents these days. In this paper Road Traffic accident involved fatal crashes data is integrated with nutritional health survey data to analysis the association of dietary habit of a motor vehicle driver's to road traffic accident by applying Association rule mining algorithms. Figure 1 show the complete system architecture used in this research and Figure 2 represents the mind map of the system which provides the general overview of this research study[13].

2. Materials and Methods

In this study unsupervised learning technique, the process of finding out knowledge from a specific area without having former knowledge about it, specifically association rule mining method which plays a great role in the process of identifying relationship between instances is used to analyze the impact of dietary habit of a motor vehicle driver's to road traffic accident. Two Association rule mining Algorithms; Apriori and Predictive Apriori Association Rule algorithms are applied on the training

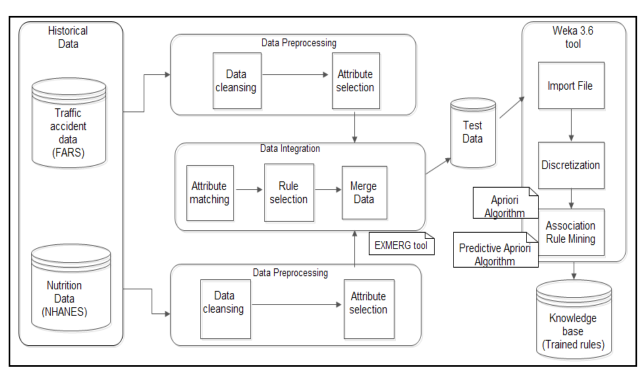

Figure 1. System architecture of this research idea.

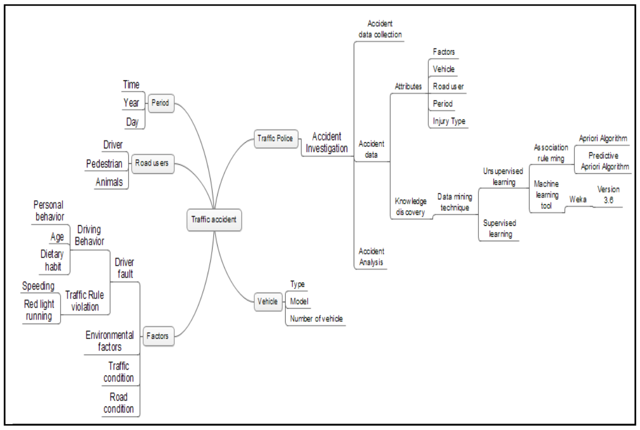

Figure 2. Mind map of the system.

dataset which is obtained by integrating two different publically available data from US, fatal crash accident and nutritional data (FARS and NHANES) using Weka 3.6 to build various association Rules.

2.1 Training Data set

FARS data contains attributes which provide detail information about accident day, year, factors, vehicle involved, persons involved ...etc, Data obtained from NHANES have attributes related to demographic and Dietary, a total of 39 attributes are selected from both datasets and after data cleansing (removing record which contain missing values), a sample of 573 traffic accident data which causes drivers fatality of period 2012 is selected and joined with nutrition data which is obtained from NHANES by matching age and sex attributes. Table 1 shows the details regarding various attributes used in this research.

2.2 Association Rule Mining

Association rules defines correlation among a set of items, the relationship is defined in terms of thefrequency of co-occurrence or appearance of the items together in each transaction process[14]. In Association rule mining various rules are created in the form X=>Y, X is termed as antecedent where as Y is called as consequent. Each of such formed rules shows the probability of occurrence of Y wherein X has already occurred depending on the support and confidence values. These are two strategically measures of importance of the association rule mining process, statistical significance of a rule is termed as support and confidence is degree of certainty of the detective associations, as mentioned by authors in their paper[15].

Table 1. Attributes and description

Attribute Name	Description
DAY	Crash day
MONTH	Crash Month
YEAR	Crash Year
DAY - WEEK	Day of week
HOUR	Crash Time
NHS	National Highway system
ROAD -FUNC	Road functionality
MAN-COLL	Mode of collision
REL-JCT2	Relation to Junction
REL-ROAD	Relation to Traffic way
LGT_COND	Light condition
WEATHER	Weather condition
DRUNK-DR	Drunk Driver
OWNER	Vehicle Owner
MAKE	Vehicle Make
MODEL	Vehicle Model
TRAV-SP	Travel Speed
ROLLOVER	Manner of Rollover
L-STATE	License state
L-STATUS	License Status
CDL-STAT	Commercial Driver License
AGE	Driver Age
SEX	Driver sex
DR1DSTZ	Amount of water taken per day(gm)
DR1CCMNM	Combination of food flag
DR1CCMTX	Combination of food type
DRD040Z	Meal place
DR1XIGRMS	Meal(gm)
DBQ095Z	Type of table salt used
DR1XIKCAL	Amount of energy (kcl)
DR1XIPROT	Protein(gm)
DR1XICARB	Carbohydrate(gm)
DR1XIFIB	Total amount of Fiber
DR1XITFAT	Total Fat(gm)
DR1XISFAT	Total Saturated fatty acid (gm)
DR1XIMFAT	Total Mono saturated Fatty acid
DR1XIPFAT	Total poly saturated fatty acid(gm)
DR1XICHOL	Cholesterol level (mg)
DR1-320Z	Amount of water taken per day(gm)

2.2.1 Apriori Algorithm

Apriori algorithm is an Association rule algorithm discovered by, to determine large item set L the first pass of Apriori algorithm counts item occurrences. In the next pass *k*, there are two phases. The large item set L_{k-1} found in the $(k-1)^{th}$ pass are used to generate the candidate item sets C_k, using the Apriori candidate generation function. In a given transaction **t** in the next phase as mentioned by authors in[15] the database requires scanning and after which support of candidates in C_k is counted.

2.2.2 Predictive Apriori Algorithm

Like Apriori algorithm it starts by building frequent item sets, and then it uses predictive accuracy to build the association rule[16]. The alphabet D is assumed as database whose individual records are shown by variable r. These individual records r are the outcome of P, static process. Association rule is described by X⇒ Y. In this algorithm, the predictive accuracy is given by c(X ⇒Y) = Pr where, (r satisfies Y | r satisfies X) is the conditional probability of $Y \subseteq r$ given that $X \subseteq r$ when the distribution of r is governed by P[17].

2.2.3 Weka Tool

One of the important open source machine learning /data mining tool is Weka[18]. This tool has collection of classification and clusteringalgorithms. It also supports various data pre-processing and visualization techniques for data mining and association rule mining and its features. In this study Weka GUI, version 3.6 is used to implement Apriori algorithm and Predictive Apriori Algorithm on the test dataset.

3. Results and Discussion

The test dataset which contains 39 attributes and 573 instances are imported to Weka tool using explorer feature for pre-processing (to convert attributes value from numeric to nominal), Predictive Apriori algorithm and Apriori algorithm is applied with min support value equal to 0.1 and confidence equal to 0.9.

3.1 Result of Apriori Algorithm

The first experiment is done with default attribute values of Weka explorer (number of rules = 10, min support = 0.1 and confidence = 0.9), the result indicates the main casualties of traffic accidents occurred in the period of 2012 were driver's whose dietary call status meet "minimum criteria", drivers having non-commercial driver license and suspended driver license. Figure 3[page 9] Show the screen shot of first 10 Best rules generated by Apriori algorithm using Weka.

Input:

Instances :573(year 2012)

Attributes : 39

Min support = 0.1

Confidence : 0.9

Number of rules = 10 and 20

Output: Best Rules generated by Apriori Algorithm

Description:

DR1DRSTZ = 1(Dietary call status meet the minimum criteria)

CDL-STAT = 0(Driver has non-commercial driver license)

L-STATE = 1(Suspended driver license)

Main factors of road accident:-

- Driver's whose dietary status meet minimum criteria (Rule1, 2, 5, 6, 7, 8, 9, and 10)
- Non- commercial driver license (Rule 4, 6 and 10)
- Suspended driver license (Rule 3,5,7,8 and 9)

The second experiment is conducted after attribute configuration in Weka explorer is done as follows (number of rules = 20, min support = 0.1 and confidence = 0.9). Apriori Algorithm generated best 20 rules as shown in Figure 4.

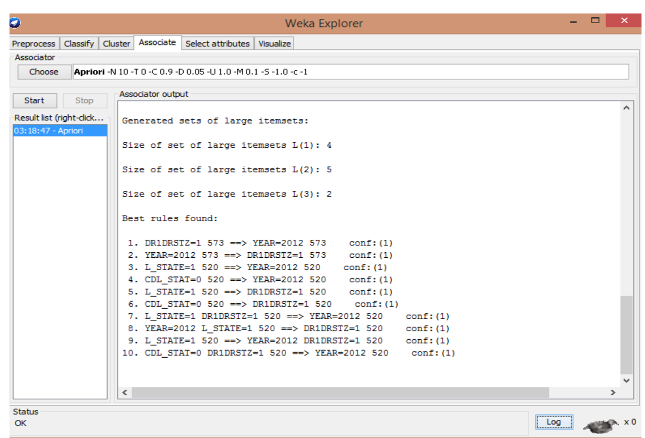

Figure 3. 10 best rules generated by Apriori algorithm.

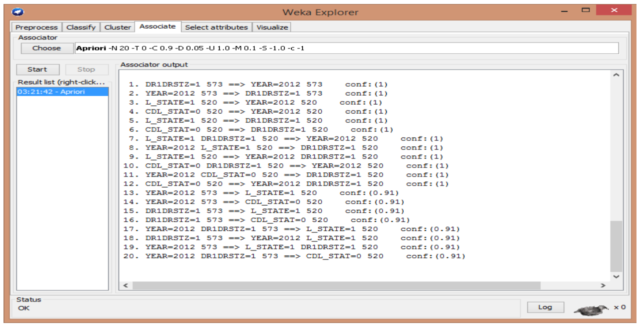

Figure 4. 20 best rules generated by Apriori algorithm.

Description:
DR1DRSTZ=1(Dietary call status meet the minimum criteria)
CDL-STAT=0(Driver has non-commercial driver license)
L-STATE=1(Suspended driver license)
Main factors of road accident:-

- Driver's whose dietary status meet minimum criteria (Except Rule 3, 4, 13 and 14)
- Non- commercial driver license (Rule 4, 6, 10, 11,12,14,16 and 20)
- Suspended driver license (Rule 3,5,7,9,12,14,16 and 20)

3.2 Result of Predictive Apriori Algorithm

The first experiment is done with attribute values of Weka (number of rules = 10) the result indicates the main factors of traffic accident in period of 2012 were driver's whose dietary status meet "minimum criteria", owner of the vehicle, driver had habit of having non combination food, "ordinary salt " used in food preparation and male drivers and accidents were happened in non junction point and at intersection area during day time in Non national highway system and there were no collision. Predictive Apriori Algorithm generated best 10 rules as shown in Figure 5.

Input:
Instances :573(year 2012)
Attributes : 39
Number of rules = 10 and 20
Output: 10 Best Rules generated by Predictive Apriori Algorithm
Description:
OWNER = 1 (The vehicle is owned by the driver)

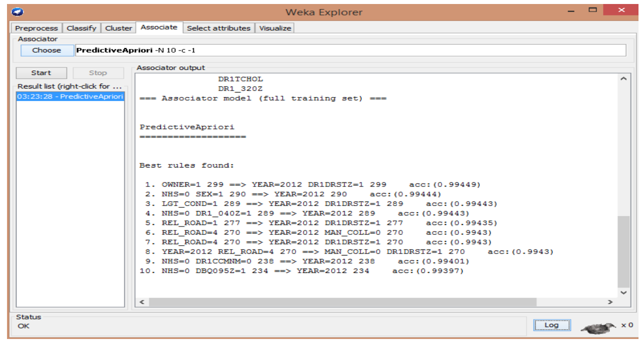

Figure 5. 10 best rules generated by predictive Apriori algorithm.

DR1DRSTZ=1(Dietary call status meet the minimum criteria)
REL-ROAD =1(Accident is happened on non junction point)
REL-ROAD =4 (Accident is happened on Intersection area)
L-CONDTION=1 (Daylight)
MAN-COL =0(No Collision)
SEX=1(Male driver)
NHS=0(Non national highway system)
DR1CCMNM=0 (Combination of food flag)
DBQ095Z=1(Ordinary table salt used in food preparation)
Main factors of road accident: -Accidents happened in:-

- Driver's whose dietary status meet minimum criteria (Rule1, 3, 5, 7 and 8)
- Owner of the Vehicle (Rule 1)
- Driver's had habit of having non combination of food (Rule 9)
- Ordinary Salt used in food preparation (Rule 10)
- Male driver's (Rule 2)

- Non junction point (Rule 5)
- NonIntersection area (Rule 10)
- No collision (Rule 6, 7 and 8)
- Day time (Rule 3)
- Non-national highway (Rule 2, 9 and 10)

The second experiment is conducted after attribute configuration in Weka explorer is done (number of rules = 20), Predictive Apriori Algorithm generated best 20 rules as shown in Figure 6.

Description:

OWNER = 1 (The vehicle is owned by the driver)

DR1DRSTZ = 1(Dietary call status meet the minimum criteria)

REL-ROAD = 1(Accident is happened on non junction point)

REL-ROAD = 4(Accident is happened on Intersection area)

WEATHER = 1 (Clear weather condition)

MAN-COL = 0(No Collision)

ROLLOVER = 0 (No Rollover)

SEX = 1(Male driver)

DRUNK-DR = 0(Non drunk driver)

NHS = 0(Non national highway system)

DR1CCMNM =0 (Combination of food flag)

DR1CCMTX = 0(Combination of food type)

DBQ095Z = 1(Ordinary table salt used infood preparation)

Main factors of road accident: -Accidents happened in:-

- Driver's whose dietary status meet minimum criteria (Rule1, 3, 4, 8, 9, 10, 11, 13 and 14)
- Owner of the Vehicle (Rule 1)
- Driver's had habit of having non combination of food (Rule 16 and 19)
- Ordinary Salt used infood preparation (Rule 18)
- Male driver's (Rule 12)
- Non drunk driver (Rule 14)

- Non junction point (Rule 6 and 9)
- Intersection area (Rule 7, 8 and 9)
- No collision manner (Rule 6, 7 and 8)
- Non national highway (Rule 2, 9, 10)
- Clear weather condition (Rule 11 and 15)
- Non rollover condition (Rule 10)

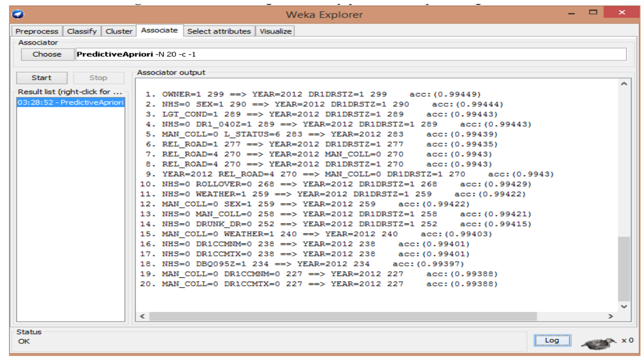

Figure 6. 20 best rules generated by predictive Apriori algorithm.

3.3 Performance Measurement

Performance of Apriori and Predictive Apriori Algorithm is measured by comparing time taken by the algorithm to build the rules, the result shows that Apriori Algorithm is faster than Predictive Apriori Algorithm and time increases in case of Predictive Apriori relative to the number of rules required to generate. Table 2 Show the time taken by Apriori Algorithm and predictive Apriori Algorithm to generate the rules.

3.4 Result Analysis

The model built using Apriori Algorithm shows among total number of 573 accidents happened in year 2012; 573 has relation with driver's whose dietary status meet minimum criteria and 520 has relation with Non-commercial driving license and suspended driving license. Accidents which has relation with drivers' whose dietary status meet minimum criteria, Non-Commercial driver license and suspended driver license built by Apriori Algorithm shown in Figure 7.

Among 573 accidents happened the model built using Predictive Apriori shows 299 has relation with owner of the vehicle, 238 has relation with drivers' having habit of non-combination of food habit, 290 were

Table 2. Execution time for algorithms

No. Instances	Generated rules	Execution Time (sec)	
		Apriori Algorithm	Predictive Apriori Algorithm
573	10	0.05	1.07
573	20	0.05	1.82

caused by male drivers', 252 were non drunk drivers and 234 had relation with "ordinary salt" used in food preparation. Accidents related to owner of vehicle, Non-combination of food habit, male drivers, non drunk drivers and "ordinary salt used in food preparation" is shown in Figure 8.

The model build using predictive Apriori indicate among 573 accidents 290 were occurred on Non national highway roads, 277 were at Non junction point and 270 were at intersection point. Figure 9 shows accidents occurred in non junction point, non national highway system and at intersection area.

Model built using Predictive Apriori shows out of 573 accidents 289 were happened during day time and 259 were happened in clear weather condition, 283 doesn't involve collision and there were no rollover in 268 accidents. Figure 10 shows accidents occurred in clear weather condition, day time and involve neither collision nor rollover.

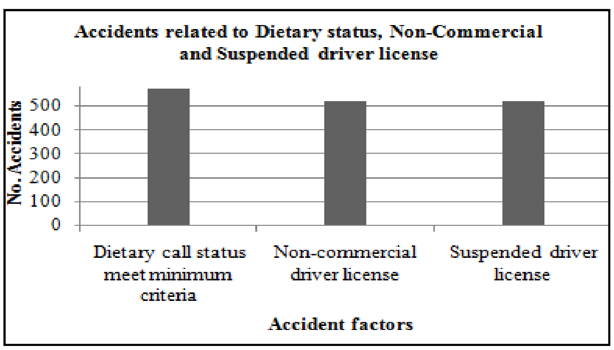

Figure 7. Number of accidents related to dietary status, non commercial and suspended driver license.

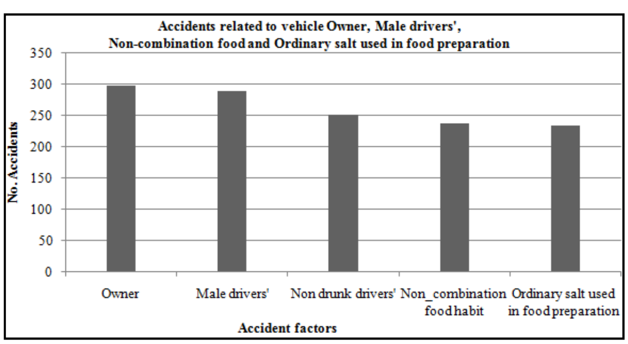

Figure 8. Number of accidents related to owner, male drivers', non drunk drivers', non combination food and ordinary salt.

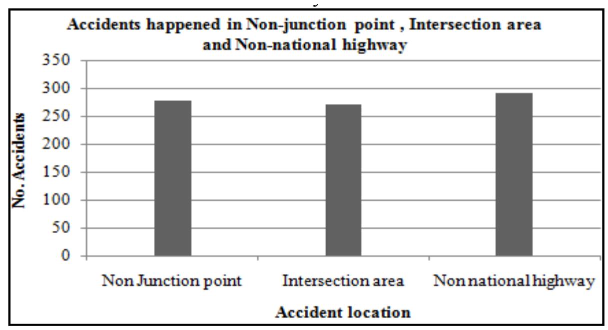

Figure 9. Number of accidents occurred in Non national highway road, at Intersection area and Non Junction point.

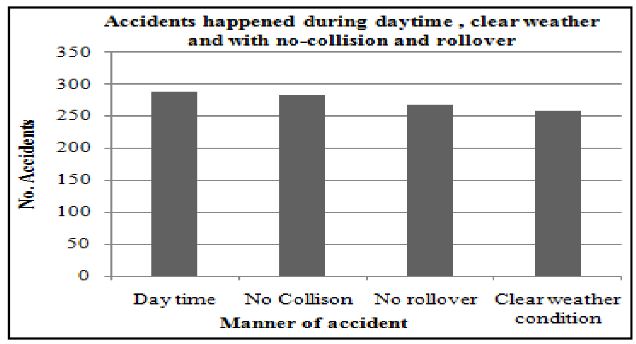

Figure 10. Number of accidents occurred in clear weather condition, day time and doesn't involved collision and Roll over.

4. Conclusion

In this study using Association rule mining technique; Apriori Algorithm and Predictive Apriori algorithm the influence of nutrition for traffic accident is predicted by integrating two datasets namely Traffic accident data and Nutrition examination data. The rules generated by both Apriori Algorithm and Predictive Apriori implicate dietary habit of a driver can be one of the causalities of road traffic accident.

5. Acknowledgement

This research paper is made possible through the support of my project guide Dr. Preeti Mulay; I would like to express my profound gratitude and deep regard for her-exemplary guidance, valuable suggestions, feedback and constant encouragement throughout the duration of the research. Her valuable suggestions were of immense help throughout this research work. I would also like to give my sincere gratitude to the reviewer of the paper for kindly read my paper and offered valuable suggestions. The last but not the least I would like to thank to my parents for their encouragement and support.

6. References

1. Rui T, Yang Z, Zhang M. Method of road traffic accidents causes and analysis based on data mining. 2010 IEEE International Conference on Computational Intelligence and Software Engineering; Wuhan. 2010. p. 1–4.
2. Alexander JM, Brian CT, Lindsay SA, Jurek GG. The relationship between age and driving attitudes and behaviors among older Americans. Injury epidemiology. 2015; 2(1):9.
3. Paikari E, Mohammad M, Ahajj R, Far BH. Data integration and clustering for real time crash prediction. 2014 IEEE 15th International conference on Information Reuse and Integration (IRI); Redwood City, CA. 2014. p. 537–44.
4. Alfonso M. Identifying crash contributing factors at urban roundabouts and using association rules to explore their relationship to different crash types. Accident Analysis and prevention. 2011; 43(4):1451–63.
5. Tebebe BT, Ajith A, Crina G. Rule mining and classification of road traffic accidents using adaptive regression tree. Internatioal Journal of Simulation. 2010; 6(10-11):80–94.
6. Ghimire B, Bhattecharjee S,Ghosh SK. Analysis of spatial correlation for road traffic accident data based on Spatial Decision tree. 2013 IEEE 4th International Conference on Computing for Geospatial Research and Application; 2013. p. 111–5.
7. Ichikawa M, Nakahara S, Tangiguchi A. Older drivers' risks of at fault motor vehicle collision. Accident Analysis and prevention. 2015; 81:120–3.
8. Lyons RA, Ward H, Brunt H, Macey S, et al. Using multiple datasets to understand trends in serious road traffic causalities. Accident Analysis and Prevention. 2008 Jul; 40(4):1406–10.
9. Siva KM, Parthasarathi V, Akshay B, Surya M, Vasudevan SK. A smart/efficient method to facilitate highway pedestrian protection. Indian Journal of Science and Technology. 2015; 8(21):1–4.
10. Mulay P, Kulkarni PA. Knowledge augmentation via incremental clustering: New technology for effective Knowledge management. International Journal of Business Information Systems. 2013; 12(1):68–87.
11. Borhade M, Mulay P. Online interactive data mining tool. Procedia of Computer Science. 2015; 50:335–40.
12. Komaroff AL, editor. Harvard medical school family health guide. New York: Free Press; 2005.
13. Mind mapping. Available from: https://en.wikipedia.org/wiki/Mind_map
14. Solanki SK, Patel JT. A survey on association rule mining. IEEE 5th International Conference on Advanced Computing and Communication Technologies (ACCT); Haryana. 2015. p. 212–6.
15. Agarwal R, Srihant R. Fast algorthim for mining association rules. Proceedings of 20th International Conference on VLDB; Chile. 1994. p. 1–13.
16. Stefan M, Mark H, Frank E. Using classification to evaluate the output of confidence-based association rule mining. In Proceeding of the 5th European Conference on Principles and Practice of Knowledge Discovery in Databases; Springer Berlin Heidelberg. 2004 Dec. p. 1–12.
17. Vanitha K, Santhi R. Evaluating the performance of association rule mining algorithms. Journal of Global Research in computer Science. 2011 Jul, 2(6):101–3.
18. University of Wakito, Downloading and installing Weka. Available from: http://www.cs.waikato.ac.nz/ml/weka/ Wikipedia