IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.7, Special Issue.2, Jan 2019 E-ISSN: 2347-2693

An Analysis the Traffic Accident Using Datamining Technique

T. Shobana[1*], R. Rajakumar[2]

[1]Dept. of Computer Science, MaruthuPandiyar College, Thanjavur, India
[2]Dept. of Computer Science, PG and Research department, Marudupandiyar college, Thanjavur, India

Available online at: www.ijcseonline.org

***Abstract*—**Association rule mining algorithms are generally used to discover all principles in the database fulfilling some base help and least certainty requirements. So as to diminish the quantity of produced rules, the adjustment of the affiliation rule mining algorithm to mine solitary a specific subset of affiliation rules where the characterization class credit is appointed to one side hand-side was examined in past research. In this examination, a dataset about traffic accidents was gathered from Dubai Traffic Department, UAE. After data preprocessing, Apriori and Predictive Apriori affiliation rules algorithms were connected to the dataset so as to investigate the connection between recorded accidents' variables to mishap seriousness in Dubai. Two arrangements of class affiliation rules were created utilizing the two algorithms and condensed to get the most intriguing standards utilizing specialized measures. Exact outcomes demonstrated that the class affiliation rules produced by Apriori algorithm were more compelling than those created by Predictive Apriori algorithm. More relationship between mishap elements and mishap seriousness level were investigated while applying Apriori algorithm.

***Keywords*—**Association Rule mining, Apriori Algorithm, Road Accident

I. INTRODUCTION

Data mining utilizes different strategies and algorithms to find information from tremendous measures of data and distinguish justifiable examples from data. It is considered as a standout amongst the most imperative patterns in data innovation in the former decade. Association rule mining algorithms are utilized to discover all rules in the database fulfilling some base help and least certainty imperatives. Classification rule mining means to locate few rules in the database to shape a precise classifier. A little arrangement of rules are produced by standard classification algorithms to frame a classifier, however these algorithms use area free predispositions and heuristics. The Apriori Algorithm is an amazing algorithm for mining successive itemsets for boolean association rules. Two separate advances are followed in this algorithm to create association rules; applying least help to locate all continuous itemsets in a database, and utilizing these successive itemsets and the base certainty limitation to produce rules. Then again, support and certainty imperatives are consolidated into a solitary measure known as precision which is utilized to produce the Apriori association rules in Predictive Apriori algorithm.
In WEKA, these algorithms are utilized to produce "n" best association rules dependent on the quantity of rules "n" controlled by the client.So as to diminish the quantity of created rules, the adjustment of the association rule mining algorithm to mine solitary a specific subset of association rules known as the class association rules (CARs) was explored in past research. This system is joining between two data mining procedures; association rule mining and classification rule mining. The mix is connected to concentrate on a unique subset of association rules where the classification class credit is doled out to one side hand-side. It comprises of two sections; a standard generator utilizing Apriori algorithm for finding association rules, and a classifier manufacturer.

Experimental outcomes demonstrated that huge reserve funds and increasingly exact classifiers could be manufactured along these lines than those created by the classification algorithm C4.5. Furthermore, this coordination helped in tackling various issues in the classification algorithms. Data mining has generally been actualized in some genuine spaces. Numerous analysts featured the importance of applying data mining procedures in the traffic mishap field so as to separate examples that can lessen the seriousness of the accidents. As indicated by a traffic report created by the World Health Organization (WHO), a huge number of street traffic accidents occur worldwide every year, and the fatalities because of these traffic accidents are additionally in the millions. More than 1.2 million individuals are executed each year because of street traffic accidents, and around fifty million get wounds or inabilities. As indicated by the places for Disease Control and Prevention (CDC), street traffic accidents cost $100 billion in restorative consideration consistently. A pile up unfortunate casualty is treated in a crisis room experiencing mishap wounds like clockwork and very nearly 40,000 passings are slaughtered yearly because of traffic accidents. With the expanded number of vehicles joined by the quick growing street development executed projects, UAE

encounters expanding number of traffic accidents with wounds and fatalities causing a genuine general health issue. In UAE, there are around 600 individuals executed in vehicle accidents every year.

Street traffic accidents are the second real reason for passings in the United Arab Emirates. The expenses of fatalities and wounds because of street traffic accidents greatly affect the general public. Past investigations demonstrated that setback and casualty rates in UAE and in other Gulf nations are a lot higher than in the creating nations with similar vehicle proprietorship levels. Dubai specifically endured lost Dh4.7 billion because of street traffic accidents over the most recent 10 years. Traffic accidents are not just bringing about loss of lives and wounds but on the other hand are gravely influencing the emirate's economy. Traffic accidents in 2007 caused a financial loss of some Dh720 million which is around one percent of the gross domestic product (GDP) of Dubai. Street traffic accidents issue needs more research to find related accidents hazard factors and recognize new techniques to diminish the vast number of accidents and fatalities. Applying data mining into traffic accidents space still needs further research. This exploration work can help with finding intriguing rules from an arrangement of created rules utilizing both association rules algorithms. Experimental outcomes can help police chiefs in the plan of new arrangements and traffic rules from some concealed examples. End and future work are created toward the end.

II. Related Work

Different investigations and numerous bits of research have concentrated on predicting the elements in street traffic mishap seriousness using data mining systems. Past research has been evaluated for the period somewhere in the range of 2001 and 2015. Ossenbruggen and Pendharkar et al. (2001) connected a strategic relapse display for discovering the variables of accidents and mishap related injuries. The models used to make a hazard evaluation of an explicit area included traits, for example, plan of roadside, arrive use movement, traffic control gadgets use, and presentation to traffic. Their examination reasoned that town destinations were less hazardous than private or shopping locales. Sohn and Hyungwon (2001) directed research on street traffic accidents (RTA) seriousness in Korea.

The analysts connected three strategies of data mining; choice tree, neural system and calculated relapse, investigated various noteworthy variables, created grouping models for mishap seriousness, and looked at the exactness of their three orders. Sohn and Lee (2002) investigated how street traffic accidents seriousness and driving condition factors are identified with one another using a few procedures and algorithms.

The specialists connected neural system, choice tree classifiers, classifier combination dependent on the Dempster– Shafer calculation, the Bayesian method, strategic model, and clustering using k-mean calculation. Their trial results demonstrated that clustering strategy was superior to different procedures. Ng, Hung and Wong (2002) combined relapse examination, topographical information framework (GIS), and bunch investigation procedures for grouping homogeneous mishap data, predicting traffic accidents number, and assessing street traffic accidents hazard in Hong Kong. Their calculation performed proficiently for lethal accidents and walker related accidents. Their proposed calculation could enable experts to recognize high hazard regions and could help town organizers to design increasingly safe streets.

III. Significance of the Research

The importance of this examination is in proposing a methodology utilizing association rules mining algorithms to convey fascinating rules from a vast arrangement of found rules extricated from Dubai traffic accidents genuine data. Figure1 portrays the system utilized in the examination.As appeared above, data identified with traffic accidents were gathered from the Dubai police expert. Various meetings with master space were directed.

Gathered data were then cleaned and preprocessed including numerous errands, for example, data total, examining, highlight choice and creation, dimensionality decrease, and treatment of missing qualities. The data mining instrument, WEKA, was utilized to apply association rules data mining procedures on the dataset. The cleaned dataset contained various qualities of three sorts, driver-related traits, street related properties, and mishap related characteristics. Qualities and class marks are made dependent on the genuine data. The class mark ('Accident Severity') had three ostensible qualities: ('Death', 'Extreme', and 'Moderate').

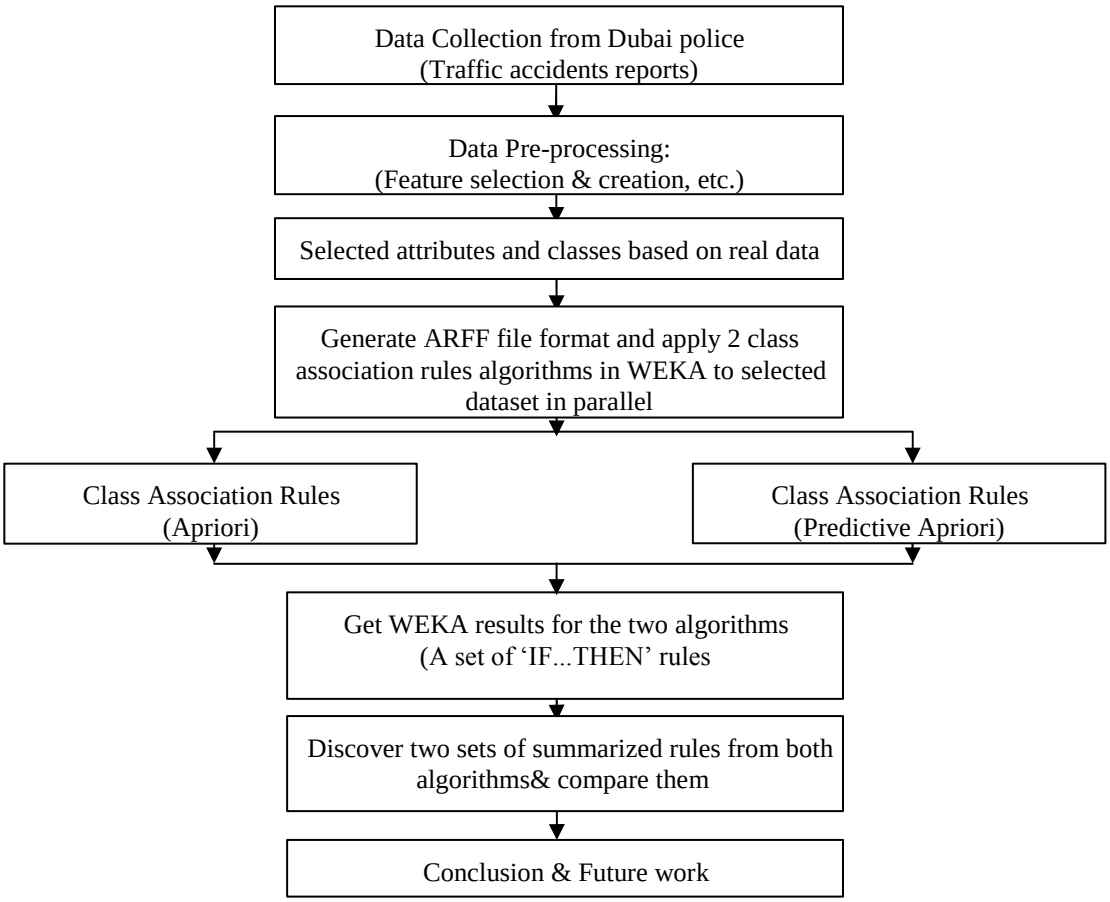

Figure 1. Summary of research methodology

IV. ASSOCIATION RULES DATA MINING

Association rules are utilized to discover the connection between data things in a value-based database. Association rules data mining algorithms are utilized to find visit association or connection in database by creating association rules of the frame: If predecessor then ensuing where precursor and subsequent are itemsets of at least one things. To produce association rules, two stages are included. To begin with, least help is utilized to locate all successive itemsets in a database. Second, these incessant itemsets and the base certainty measure are utilized to create association rules.

There are numerous association rule algorithms, however we concentrated on two association rule algorithms; Apriori and Predictive Apriori in our investigation. Both association rule algorithms are utilized to find the best blend of mishap factors that connects to the mishap seriousness level.

A. Apriori Algorithm

Apriori Association rule algorithm is utilized for mining the incessant examples in database where support and certainty are the two estimates used to quantify association rule quality. Support is the level of exchange in the database that contains XUY for the association rule X->Y. Certainty is the proportion of the quantity of exchange that contains XUY to the quantity of exchange that contains X.

B.Predictive Apriori Algorithm

Predictive APRIORI association rule algorithm is likewise utilized for mining designs in database. It varies than Apriorialgorithm in that both help and certainty measures are consolidated into a solitary measure called predictive exactness.

Table 1. Attributes with their Description

Attribute Name	**Description**
Accid_day	The day in which the occurred.
Accid_month	The month in which the occurred.
Accid_year	The year in which the occurred.
N_of_p_injured	Number of injured persons
Accid_cause	The cause of the accident
Accid_type	The type of the accident
Driv_gen	Driver's gender
Driv_nation	Driver's nationality
Driv_age	Driver's age

Driv_vec_type	Driver's vehicle type causing the accident
Driv_drink	Whether the driver was drunk or not.
Driv_belt	Whether the driver's seat belt was fastened or not
Road_accid_place	The place where the accident occurred
Road_accid_type	Type of the accident road
Road_light_cond	The light condition of the road
Weather_cond	The weather condition when the accident occurred
Road_surface	Whether the surface of the road was dry, wet, sandy, etc.

Table 2. Class Labels

Class Label	Class Description
Death	An accident where one or more persons dies within 30 days of the accident.
Severe	An accident where a person is severely injured and needs intensive care.
Moderate	An accident where one or more persons is injured and kept in hospital for more than 12 hours.

Table 3. Selected Records for 2008-2010 Dataset

Class	Year		
	2008	**2009**	**2010**
Death	17	16	15
Severe	5	17	16
Moderate	29	47	41
Total	51	80	72

C. Summarizing Generated Class Association Rules

Principle Covers strategy is a procedure used to outline the quantity of best rules acquired. It is connected to register a cover for an arrangement of rules with the equivalent resulting. In the wake of producing an arrangement of rules has the equivalent ensuing. This is now done by producing class association rules dependent on the class property. The avaricious algorithm is then being connected to discover a cover. For an arrangement of rules R = {r1: A1-> B, r2= A2-> B, rd= Ad-> B}, the number or rules secured by ri is figured. In the event that Ai contains or equivalent to Aj, ri covers rj. On the off chance that a standard ri has most extreme cover(ri) where it is the quantity of rules secured by ri. On the off chance that ri has most extreme cover(ri), it is added to the standard cover R new subsequent to expelling and disposing of all its secured rules. By applying this algorithm, the standard cover should have at most the first number of rules and each standard is secured. In spite of the fact that this algorithm isn't ideal to the quantity of rules chosen, it is quick to classify and abridge the rules in a post-handling stage. In this examination, we could abridge the rules created by Apriori algorithm utilizing Rule Covers strategy. Applying this technique for the rules created by Predictive Apriori algorithm was not working. No mix of factors was created inside the rules. Along these lines, we utilized another strategy to outline the rules produced by Predictive Apriori algorithm utilizing OR to consolidate the distinctive rules. The last exactness of the condensed principle can be determined as pursues: precision of abridged guideline = Sum (precision of each consolidated standard)/add up to number of joined rules.

V. EXPERIMENTAL RESULT

There are four subsections to concentrate on while talking about the outcomes identified with the investigation; the created rules found from both association rules algorithms for every one of the three classes with outline of these rules and the examination of the rules removed for each class from every algorithm to discover which algorithm is increasingly reasonable for traffic mishap examination. The rules for the years 2008 to 2010 are created for every one of the three classes; demise, extreme, and moderate utilizing Apriori and Predictive Apriori algorithms. Best twenty rules were first produced for death class, serious class, and moderate class by WEKA. The condensed last arrangements of rules for each class were then produced utilizing rule covers technique. A standard spreads strategy is connected to outline the best rules created by Apriori algorithm. It condenses the rules by evacuating shorter rules shrouded in greater rules. For the rules produced by Predictive Apriori algorithm, we consolidated the rules and determined the precision for the each condensed guideline.

A. Analysis of Rules for Death Class

The best rules created for Death class utilizing Apriori algorithm and Predictive Apriori algorithms and the arrangement of abridged last rules are appeared in the accompanying areas.

B. Rules Generated Using Apriori Algorithm

Table 4 exhibits the best rules created for death class utilizing Apriori algorithm and Table 5 demonstrates the last arrangement of abridged rules in the wake of applying rule covers strategy.

Table 4: Best Rules for Death Class Using Apriori Algorithm

Rule#	Best Rule
1	If Driv_gen=M 47 ==> Class=Death 47 conf:(1)
2	If Road_surface=Dry 47 ==> Class=Death 47 conf:(1)
3	If Weather_cond=Clear 46 ==> Class=Death 46 conf:(1)
4	If Driv_gen=M Road_surface=Dry 46 ==> Class=Death 46 conf:(1)
5	If Weather_cond=Clear Road_surface=Dry 46 ==> Class=Death 46 conf:(1)
6	If Driv_gen=M Weather_cond=Clear 45 ==> Class=Death 45 conf:(1)
7	If Driv_gen=M Weather_cond=Clear Road_surface=Dry 45 ==> Class=Death 45 conf:(1)

8	If Driv_drink=Not_checked 42 ==> Class=Death 42 conf:(1)
9	If Driv_drink=Not_checked Road_surface=Dry 42 ==> Class=Death 42 conf:(1)
10	If Driv_gen=M Driv_drink=Not_checked 41 ==> Class=Death 41 conf:(1)
11	If Driv_drink=Not_checked Weather_cond=Clear 41 ==> Class=Death 41 conf:(1)
12	If Driv_gen=M Driv_drink=Not_checkedRoad_surface=Dry 41 ==> Class=Death 41 conf:(1)
13	If Driv_drink=Not_checkedWeather_cond=Clear Road_surface=Dry 41 ==> Class=Death 41 conf:(1)
14	If Driv_gen=M Driv_drink=Not_checkedWeather_cond=Clear 40 ==> Class=Death 40 conf:(1)
15	If Driv_gen=M Driv_drink=Not_checked Weather_cond=Clear Road_surface=Dry 40 ==> Class=Death 40 conf:(1)
16	If N_of_p_injured=1 31 ==> Class=Death 31 conf:(1)
17	If N_of_p_injured=1 Driv_gen=M 31 ==> Class=Death 31 conf:(1)
18	If Driv_vec_type=Private 30 ==> Class=Death 30 conf:(1)
19	If N_of_p_injured=1 Road_surface=Dry 30 ==> Class=Death 30 conf:(1)
20	If Driv_vec_type=Private Weather_cond=Clear 30 ==> Class=Death 30 conf:(1)

Table 5: Summarized rules for Death Class Using Apriori Algorithm

Rule#	Summarized Rule
15	If Driv_gen=M and Driv_drink=Not_checked and Weather_cond=Clear and Road_surface=Dry 40 ==> Class=Death 40 conf:(1)
17	If N_of_p_injured=1 Driv_gen=M 31 ==> Class=Death 31 conf:(1)
19	If N_of_p_injured=1 Road_surface=Dry 30 ==> Class=Death 30 conf:(1)
20	If Driv_vec_type=Private Weather_cond=Clear 30 ==> Class=Death 30 conf:(1)

As expressed above, death accidents happened generally by male drivers driving private vehicles at clear climate condition and on dry streets. It isn't certain whether the driver was flushed or not. One harmed individual was recorded at the season of mishap. While applying rule covers technique on the best rules for death class produced utilizing Apriori algorithm, we can see that among the twenty best rules, four rules just show up in the wake of disposing of shorter rules shrouded in longer ones.

VI. Discussions and Recommendations

In view of this exploration, we can see that while applying rule covers technique on the best rules for the three classes produced (demise, serve, and moderate) utilizing Apriori algorithm, few outlined rules were acquired in the wake of taking out shorter rules canvassed in longer ones. Then again, when we connected guideline covers technique on the best rules for the three classes utilizing Predictive Apriori algorithm, no rules were disposed of. We could just total them together. Moreover, the best twenty rules produced by Apriori algorithm contained mixes of different accidents' variables not normal for the best twenty rules created by Predictive Apriori algorithm where each standard contained a solitary factor at any given moment and did not have the associations among the distinctive mishap factors. Along these lines, exact outcomes demonstrated that class association rules created by Apriori algorithm were more viable than those produced by Predictive Apriori algorithm if associations between the distinctive elements are of high centrality. Increasingly number of rules could be dispensed with and more associations between mishap variables and mishap seriousness were investigated while applying Apriori algorithm.

VII. Conclusion

Our exact investigations demonstrated that while applying rule covers strategy on the created class association rules utilizing Apriori and Predictive Apriori algorithms, many class association rules produced by Apriori algorithm were wiped out and more successful rules than those created by Predictive Apriori algorithm were acquired. What's more, more associations between mishap elements and mishap seriousness were investigated while applying Apriori algorithm. Then again, Predictive Apriori algorithm could infer progressively number of rules that could be valuable when considering the impact of every individual factor to mishap seriousness. The adjustment of the association rule mining algorithm to mine solitary a specific subset of association rules where the classification class credit is doled out to one side handside could effectively create progressively successful rules covering every one of the three classes. It tackled the issues that could happen when customary classification algorithms connected including no fascinating rules could be found if the number of inhabitants in a specific class was so substantial. Apriori algorithm beat Predictive Apriori algorithm while applying rule covers technique for outlining the rules by giving less number of critical fascinating rules. These outcomes may help the chiefs in the traffic mishap department to take activities dependent on some concealed examples from the data. As future work, other data mining procedures and algorithms can be connected to the dataset, for example, grouping and fleeting data mining strategies so as to examine in various ways which mishap variables can influence the event of

accidents. Likewise, space driven data mining approach can likewise be connected to create increasingly significant rules.

REFERENCES

[1]. Abdel-Aty, M., and Abdelwahab, H., Analysis and Prediction of Traffic Fatalities Resulting From Angle Collisions Including the Effect of Vehicles‟ Configuration and Compatibility. Accident Analysis and Prevention, 2003.

[2]. Bedard, M., Guyatt, G. H., Stones, M. J., &Hireds, J. P., The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities. Accident analysis and Prevention, Vol. 34, pp. 717- 727, 2002.

[3]. Domingos, Pedro & Michael Pazzani (1997) "On the optimality of the simple Bayesian classifier under zero-one loss". Machine Learning, 29:103–137.

[4]. Evanco, W. M., The Potential Impact of Rural Mayday Systems on Vehicular Crash Fatalities. Accident Analysis and Prevention, Vol. 31, 1999, pp. 455-462.

[5]. E. Frank and I. H. Witten. Generating accurate rule sets without global optimization. In Proc. of the Int‟l Conf. on Machine Learning, pages 144–151. Morgan Kaufmann Publishers Inc., 1998.

[6].Kweon, Y. J., &Kockelman, D. M., Overall Injury Risk to Different Drivers: Combining Exposure, Frequency, and Severity Models. Accident Analysis and Prevention, Vol. 35, 2003, pp. 441-450.

[7].Martin, P. G., Crandall, J. R., &Pilkey, W. D., Injury Trends of Passenger Car Drivers In the USA. Accident Analysis and Prevention, Vol. 32, 2000, pp. 541-557.

[8].National Highway Traffic Safety Administration, Traffic Safety Facts 2005, 2007, P. 54.

[9].Ossenbruggen, P.J., Pendharkar, J. and Ivan, J. 2001, “Roadway safety in rural and small urbanized areas”. Accidents Analysis and Prevention, 33 (4), pp. 485– 498.

[10].Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

[11].Rish, Irina. (2001). "An empirical study of the naive Bayes classifier". IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence.

