CAUSE ANALYSIS OF ACCIDENTS USING RULE MINING ALGORITHM

[1] Dr. S Manjula,[2] Hitha M V,[3] Swastik K, [4]Varun C U,[5]Akash Anilkumar Rampur

[1]Assistant Professor,[2]Undergraduate, [3]Undergraduate, [4]Undergraduate, [5]Undergraduate
[1]Computer Science Department,
[1] JSS Science and Technology University, Mysuru, India.

Abstract *: Traffic accidents on city streets are the consequence of coordinated activities by various variables, including humans, vehicles, roadways, and the environment. In order to find the key causes of these accidents, it is necessary to mine the relationship rules between significant risk factors from the statistics on these incidences..*
This research advances the Apriori algorithm for mining risk factor connection rules and delves into the causes of traffic accidents on city streets, taking into consideration the many layers and dimensions of accident data. To develop qualifying association rules between risk variables, metrics such as support, confidence, and lift were adjusted depending to the layer and dimension of individual attributes. After then, the data was filtered to create a set of useful association rules.. The study's findings help the traffic department to establish effective accident prevention strategies and improve traffic safety on city streets.

IndexTerms - Urban roads, causes of traffic accidents, data mining, association rules and Apriori algorithm.

1. INTRODUCTION

City street traffic accidents are an inescapable element of urban mobility, resulting in a significant number of fatalities and property damage. China has achieved great results in terms of traffic safety and has avoided further deterioration in the future. However, in terms of traffic safety, the situation in China's cities is not yet encouraging. China is ranked 98th out of 178 countries in terms of road traffic fatalities, according to WHO statistics (10.45 deaths per 10,000 vehicles). According to China's 2015 traffic accident data, there are 23 accidents per 100 kilometres on urban roads in China, significantly greater than on expressways, first-class motorways, and second-class highways..

The traditional Apriori method was enhanced from the perspectives of data structure, measuring indices, and subjective limitations to address these inadequacies in light of the statistical characteristics of traffic accidents on urban highways. The redesigned algorithm mined the association rules in four steps: data processing, multidimensional data modelling, algorithm mining, and rule interpretation.. The enhanced Apriori algorithm was then utilised to identify the key causes of traffic accidents on metropolitan roadways utilising a multidimensional and multilevel mining of accident data. The improved algorithm proposed in this study improves the analysis technique system for causes of urban road traffic accidents, which is quite beneficial.

2. LITERATURE SURVEY

For the idea we have referred several papers. Literature survey is as follows:

Prediction of the cause of accident and accident prone location on roads by web application Published in: 2017 8th International Conference on Computing, Communication and ICCNT.
Authors: Gagandeep Kaur Punjabi University, Patiala, Punjab, India. Harpreet Kaur Punjabi University, Patiala, Punjab, India

Big Data Technology and Applications in Intelligent Transportation
Published in: IEEE Access (Volume: 8). Publisher :IEEE. Page(s): 75607 – 75615. Date of Publication: 16 April 2020

Because of the seriousness of traffic accidents on city roads, Chinese and foreign researchers and professionals have delved into the influencing elements and examined the causes of such incidents. In terms of influencing elements, human factors such as the driver account for 80-90 percent of traffic accidents [1], [2]. After looking into certain sorts of incidents, Larsen and Kines [3] came to the following conclusions: Excessive speed, drunk driving, and driving under the influence of illegal drugs are the most common causes of head-on crashes; all of the drivers engaged in these incidents are male. The most common causes of left-turn accidents are attention mistakes and senior age.. McGwin and Brown [4] examined the types and causes of traffic accidents among drivers of various ages, finding that elderly drivers are more prone to turning and changing lanes due to slow reaction times and lack of observation, whereas young drivers are more likely to be involved in accidents due to their risk-taking and lack of skill. Ballesterors et al. [5] discovered that under safe driving conditions, the youngest drivers have the highest rate of traffic accidents.

In light of the statistical aspects of traffic accidents on urban highways, the traditional Apriori algorithm was upgraded from the perspectives of data structure, measuring indices, and subjective constraints to address these flaws. The R programming language was used to optimise the algorithm flow.. The association rules were mined in four steps in the modified algorithm: data processing, multidimensional data modelling, algorithm mining, and rule interpretation. Then, using a multidimensional and multilayered mining of accident data, the upgraded Apriori algorithm was used to identify the primary causes of traffic accidents on metropolitan roadways.. The enhanced algorithm suggested in this research enhances the analysis technique system of causes of urban road traffic accidents, which is really useful..

3. METHODOLOGY

3.1 DATA COLLECTION AND PROCESSING

The layout of data samples on traffic accidents on city roads has a significant impact on the mining of association rules. If the data is organised, the mining efficiency can be improved. However, there are numerous items missing from the traffic accident records. The original records should be transformed into structured data before using association rule mining. Structured data is frequently created in three processes (FIG -1) based on the original records: data cleansing, data transformation, and data reduction

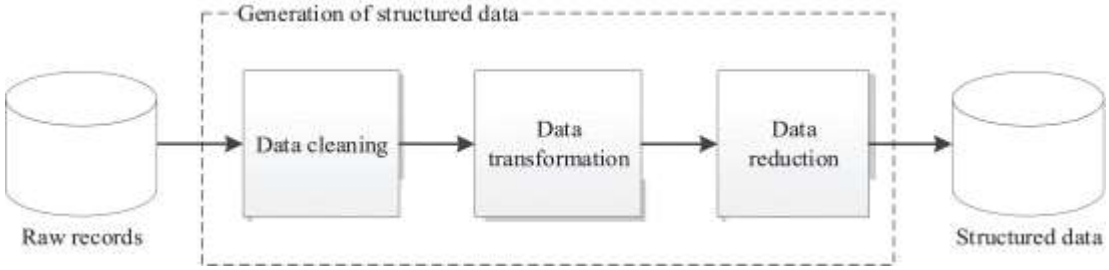

FIG -1: Structured data generation for association rule mining

Initially we thought of collecting accident attributes and accident details from traffic departments in our city. But unfortunately because of COVID – 19 strict lockdown is imposed in our city, so we made a list of common accident attributes and created accident list by our own based on accidents that we have seen. The list of common accident attributes is shown in FIG -2 and for each attribute there will be two values one 'yes' and another 'no'

FIG -2: Attribute list that should be added for each accident

We have created Web Application as a interface which have 3 roles Admin, Inspector and Writer. Only admin has access to add attribute to attribute list mentioned in FIG -2

3.2 . IMPROVED APRIORI ALGORITHM

Before using the Apriori algorithm to mine the association rules between risk variables of traffic accidents on urban roads, all of the attributes must be translated into the same dimension, according to the preceding analysis on association rule mining. This normalisation process may lower the database's scanning performance in the search for frequent itemsets, resulting in nonsensical search results. To put it another way, the Apriori technique isn't immediately relevant to the multilayer, multidimensional dataset of traffic incidents on city highways

As a result, the Apriori method was tweaked to better suit association rule mining of multilayer, multidimensional data from traffic incidents on city streets. Each accident attribute was treated as a predicate in this study. The levels of characteristics related to mining demand were chosen for association analysis because the accident data is multilayered. The multilayer mining problem is thereby reduced to a single-layer mining problem, revealing the relationships between distinct levels of attributes. This strategy can reduce the number of nonsensical errors in standard association rule algorithm output results caused by too many variables, as well as increase the efficiency of data mining algorithms to some level. The following is the flow of the updated Apriori algorithm:

Step 1. Create the frequent itemset.
Input: an r-dimensional database; the minimum threshold support, min-sup.
Output: an r-dimensional frequent itemset.
(1) Initialize the parameters: d=1, G=∅ and P1=∅;
(2) Create the 1st r-dimensional candidate itemset P1;
(3) Create the 1st frequent itemset G1;
(4) Create the d-th candidate itemset pd based on the (d1)-th frequent itemset Gd-1;
(5) Create the d-th frequent itemset Gd based on the d-th candidate itemset Pd: Calculate the sup of each item in Pd by formula (1), and add all the items whose sup>min-sup into G;
(6) Obtain the r-dimensional frequent itemset G, and go to

Step 2. Create the set of multidimensional association rules I
Input: Frequent itemset G; the minimum threshold confidence, min-conf.
Output: A set of association rules I.
(7) Initialize I=∅;
(8) Find the conf and lift between the nonempty subsets of each frequent itemset by formulas (2) and (3), respectively;
(9) Sum up all the association rules whose conf>min-conf and lift>1 into I.
(10) Stop the algorithm

3.3 Interface

As mentioned above we have created web application that will be used by traffic department. This application has 3 main roles Admin, Inspector and Writer. In the following sub-sections functionality of each role is explained

3.3.1 ADMIN

Login
Admin will login to our application to perform this desired task
Manage attributes
Here admin will be managing the attributes which comes under him. (FIG -4)
Manage attribute values
Admin will also be managing the attribute values (FIG -3)
Manage Inspector
Admin will be managing the inspector (FIG -3)
Logout
After performing all the activities, admin will be logging out of our application.

3.3.2. INSPECTOR

Login
Inspector will be logging in to our application
Manage Accidents
Inspector will be managing the accidents which occurs (FIG -6)
Manage accident details
Inspector will be managing the accident details too (FIG -2)
View Prediction
Inspector will be viewing the prediction. (FIG -7)
Logout
After performing this task, Inspector will be logging out of this application.

3.3.2. WRITER

Login
Writer will be logging in to our application
Manage Accidents
Writer will be managing the accidents which occurs (FIG -6)
Manage accident details
Writer will be managing the accident details too (FIG -2)
View Prediction
Writer will be viewing the prediction. (FIG -7)
Logout
After performing this task, Writer will be logging out of this application.

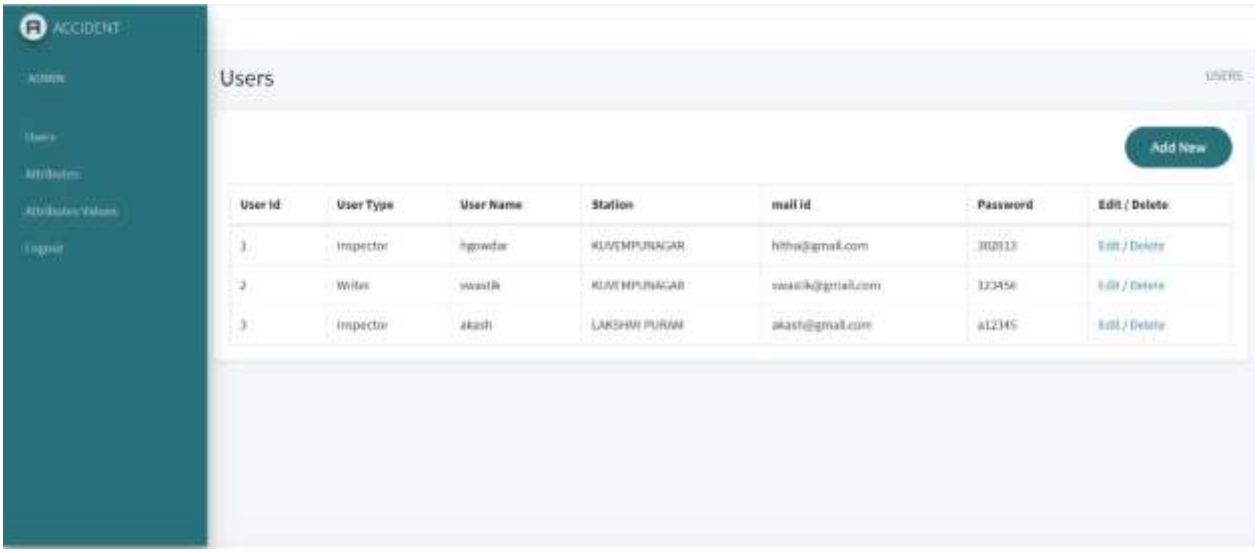

FIG -3: List of Inspector and Writer added by admin

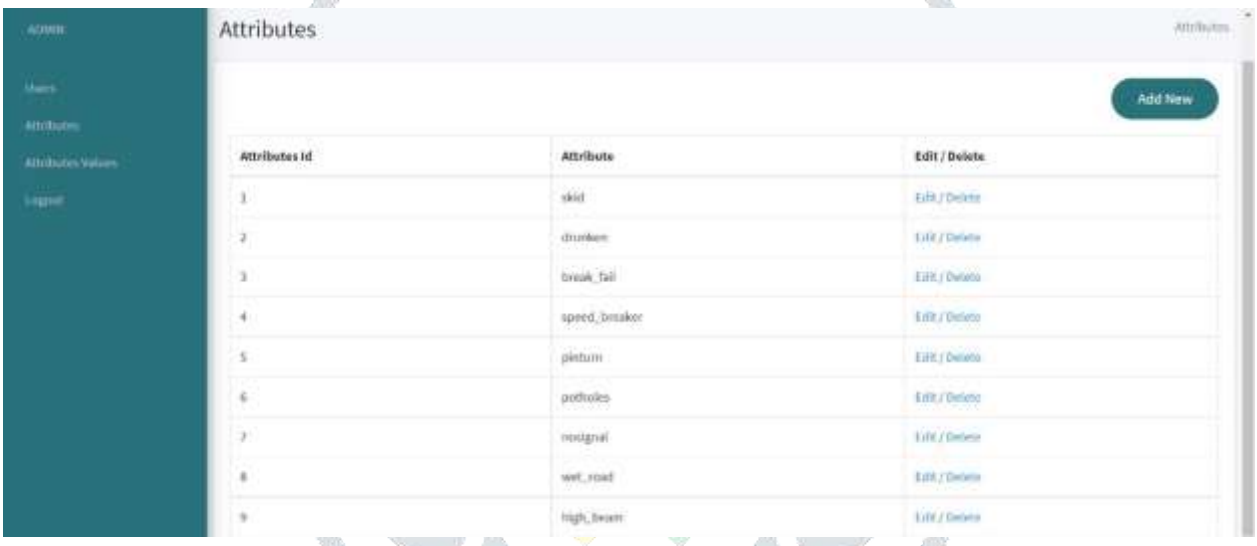

FIG -4: List of attributes added by admin

Only admin has power to add Inspector and Writer to this application as shown in FIG -3 and also as mentioned in previous section only admin can add attributes to application which is shown in FIG -4 and only Inspector or Writer can use algorithm to predict cause

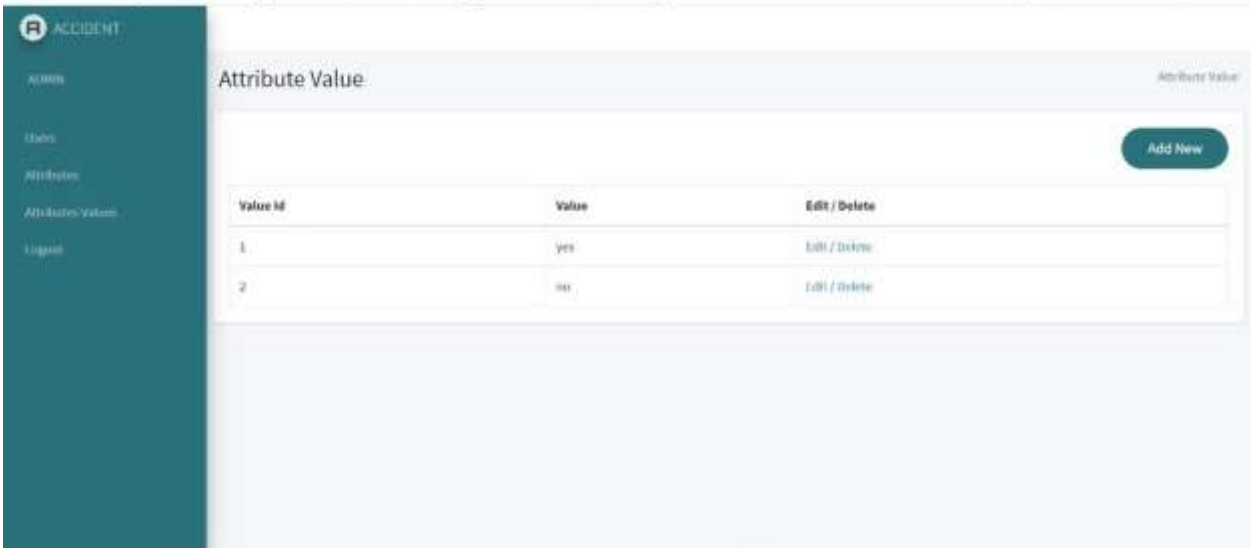

FIG -5: Attribute values

4. RESULTS AND DISCUSSION

Whenever an accident occurs Inspector/Writer will collect the accident details and then they will add accident and its attributes to the database through the application. As shown in FIG -2 the Inspector/Writer will have to set listed attribute values to "yes" or "no".

Fig -6 shows the accidents added by Inspector/Writer by specifying the location which comes under their station limits

Initially the database will be loaded with known accident details. But every time an accident occurs Inspector/Writer will add accidents to database as mentioned above

Once enough accident details are collected Inspector/Writer can use this application to predict accident cause. Inspector/Writer login to the system and according to their station limits accident cause is displayed in the view prediction page

FIG -7 in this document shows the prediction viewed by Kuvempu Nagara station Inspector/Writer. Currently we have designed our system to display individual reason and multiple reasons that might have been the reasons for accidents occurring in that station limits.

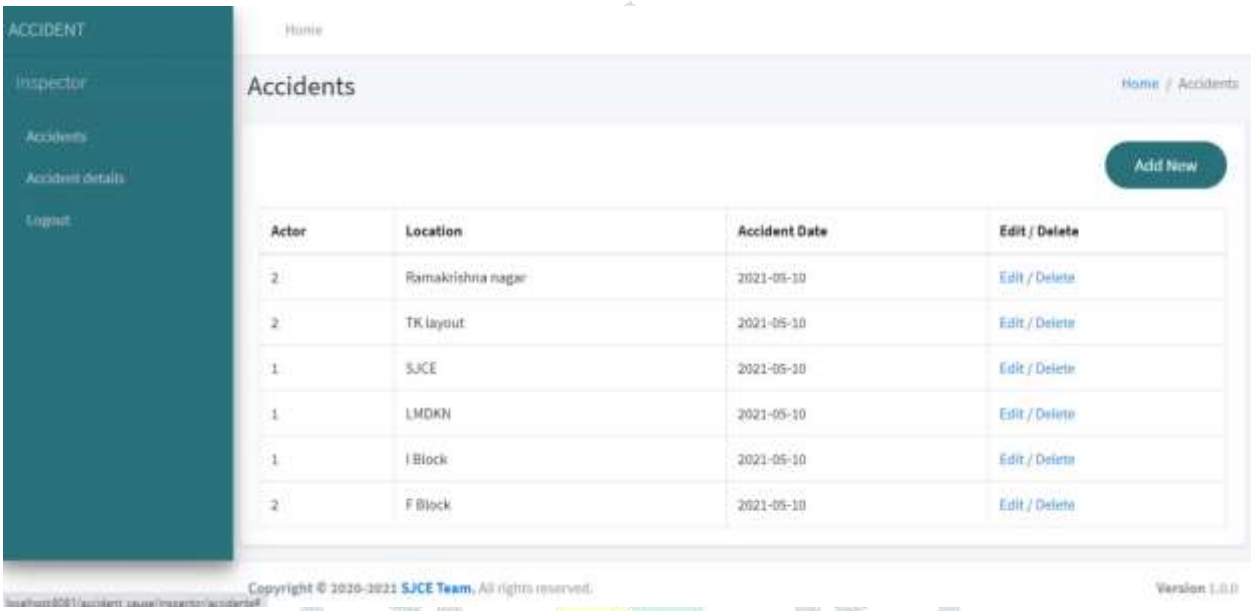

FIG -6: Accidents added by Inspector/Writer

FIG -7: Cause of accidents predicted by our application

5. CONCLUSION

The cause analysis of traffic accidents aids the traffic department in identifying the key influencing variables of accidents and demonstrating the relationships between these factors, allowing risk control measures to be devised. Based on association rule mining, this paper proposes a cause analysis technique for traffic accidents on urban motorways.. The authors tweaked the flow of

the apriori algorithm to find correlation rules between risk variables in traffic accidents on city streets, taking into consideration the many layers and dimensions of data on such incidents.. To develop qualifying association rules between risk variables, metrics such as support, confidence, and lift were adjusted depending to the layer and dimension of individual attributes. After then, the data was filtered to create a set of useful association rules.. Furthermore, while this work uses sample data from urban road traffic accidents to conduct an example analysis, the enhanced methodology described in this study may be applied to realistic application problems with numerous layers and dimensions sample data features.

6. ACKNOWLEDGEMENT

We would like to convey our gratitude to Dr M P Pushpalatha, Head of Department of Computer Science and Engineering, for giving us the opportunity to enter into this journey of knowledge. We are grateful to Dr. S Manjula, Assistant Professor, Department of Computer Science and Engineering and Prof. Parashivamurthy B M, Assistant Professor, Department of Computer Science and Engineering for their guidance, assistance, support and criticism for the improvement of the project. We would also like to thank our friends who supported us directly or indirectly towards the progress of our project.

7. REFERENCES

[1] Prajakta S.Kasbe, Apeksha V sakhare," Review on road accidents data analysis using data minig techniques", 2017 IEEE International conference on innovations in information embedded and communication systems (ICIIECS).

[2] E.Suganya, S. Vijaya Rani "ANALYSIS OF ROAD ACCIDENTS IN INDIA USING DATA MINING CLASSIFICATION ALGORITHMS" 2017 IEEE International Conference on inventive computing and informatics (ICICI).

[3] Liling Li, Sharad Shrestha, Gongzhu Hu, "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques", 2017 IEEE 15th International Conference on Software Engineering Research Management and Applications (SERA)

[4] Alyssa Ditcharoen, Bunna Chhour, Tunyarat Traikunwaranon, "Road Traffic Accidents Severity Factors", 2018 5th International Conference on Bussiness and Industrial Research (ICBIR).

[5] Abhirup Das, abhisek Ray , Abhishek ghosh, "Vehicle accident prevent cum location monitoring system", 2017 8 TH Annual Industrial Automatation on Electromechanical Engineering Conference (IEMECON)

[6] A. Papanikolaou and E. Eliopoulou, ``Impact of ship age on tanker accidents,"in *Proc. Greek Sect. Soc. Nav. Architects Mar. Eng.*, Athens, Greece, 2008, pp. 103_110

[7] O. Agbonkhese, G. L. Yisa, E. G. Agbonkhese, D. O. Akanbi, E. O. Aka, and E. B. Mondigha, ``Road traf_c accidents in Nigeria: Causes and preventive measures," *Civil Environ. Res.*, vol. 3, no. 13, pp. 90_99, 2013

[8] L. Larsen and P. Kines, ``Multidisciplinary in-depth investigations of head on and left-turn road collisions," *Accident Anal. Prevention*, vol. 34, no. 3,pp. 367_380, May 2002

[9] J. G. Mc Gwin and D. B. Brown, ``Characteristics of traf_c crashes among young, middle- aged, and older drivers," Accident Anal. Prevention, vol. 31, no. 3, pp. 181_198, May 1999.