International Journal on Applications in Engineering and Technology
Volume 8 : Issue 2 : May 2022, pp 117 – 121 www.aetsjournal.com ISSN (Online) : 2455 - 0523

ROAD ACCIDENT PREDICTION USING DATA MINING TECHNIQUES

SATHISH KUMAR.G, NITHEESH.R, PRAKASH. M, DIVYA.K

[1] *Assistant Professor, Department of Computer Science and Engineering,*
Paavai College of Engineering

[2,3,4] *Undergraduate student, Department of Computer Science and Engineering,*
Paavai College of Engineering

Abstract: - – Road traffic accident are considered as major problem for public health concern. In order to give safe driving suggestions, make clearance of public view, careful analysis of road traffic data is critical to find out variables that are closely related to fatal accidents. In this paper we apply Probability analysis and data mining algorithms on FARS Fatal Accident dataset as an attempt to address this problem. Classification model was built by Naive Bayes classifier, clusters were formed by simple K-means clustering algorithm. Certain safety driving suggestions were made based on probability, classification model, and clusters obtained.

Key words: Accident prediction, Data mining, Apriori algorithm, Rule mining, Classification

1. INTRODUCTION

There are is lot of vehicles driving on the road every day, and traffic accidents could happen at any time. As human being, we all want to avoid accident and stay safe. To find out how to drive safer, data mining technique could be applied on the traffic accident dataset to find out some valuable information, thus give driving suggestion, Association rule mining algorithm is a popular methodology to identify the significant relations between the data stored in large database and also plays a very important role in frequent item set mining [1]. A classical association rule mining method is the Apriori algorithm who main task is to find frequent item sets, In order to which is the method we use to analyze the road traffic data. Classification in data mining aims at constructing a model (classifier) from a training data set that can be used to classify records of unknown class labels. The Na¨ıve Bayes technique is one of the very basic probability-based methods for classification that is based on the Bayes' hypothesis with the presumption of independence between each pair of variables.

1. REVIEW LITRETURE

Jayasudha [4] analyzed the traffic accident using data mining technique that could possibly reduce the fatality rate.

Using a road safety database enables to reduce the fatality by implementing road safety programs at local and national levels. Those database scheme which describes the road accident via road condition, person involved and other data would be useful for case evaluation,

Collecting additional evidences, settlement decision and subrogation. The International Road Traffic and Accident Database (IRTAD), GLOBESAFE, website for ARC networks are the best resources to collect accident data. It could classify information and provide warning an audio or video. It was also identified that accident rates highest in intersections thenother portion of road [4].

Solaiman et al. [8] Describes various ways accident data could be collected, placed in a centralized database server and visualized the accident. Data could be collected via different sources and the more the number of sources the better the result. This is because the data could be validate with respect to one another few could be discarded thus helping to clean up the data. Different parameters such as junction type, collision type, location, month, time of occurrence, vehicle type could be visualized in a certain time strap tosee the how those.

Parameter change and behave with respect to time, it couldfind the safest and dangerous roads [8].

Partition base clustering and density based clustering were performed by Kumar [6] to group similar accidents together. It's based on a categorical nature of most of the data K-modes algorithm was used. To find the correlation among various sets of attributes association rule mining was performed. Among the various rules that are generated those which seemed interesting were considered based on support count and confidence. The experiments showed that the accidents were dependent of location and most of the accident occurred in populated areas such as markets, hospitals, local colonies. Blind turn on road was the most crucial action responsible of those accidents and main duration of accidents were on morning time a.m. to 6 a.m. on hills and 8 p.m. to 4 a.m. on other roads [6]. Krishnaven and Hemalatha [5]

Some Classification models to predict a severity of injury that occurred during traffic accidents. Naive Bayes Bayesian classifier, AdaBoostM1 Meta classifier, PART Rule classifier, J48 Decision Tree classifier, and Random Forest Tree classifier are compared for classifying the type of injury severity of various traffic accidents. The final result shows that the Random Forest out performs the other four algorithms [5].

Amira, Paree and Araar [2] applied association rules mining algorithm on the dataset about traffic accidents which was gathered from Dubai Traffic Office, UAE. After information preprocessing, Apriori and Predictive Apriori association rules algorithms were applied to a dataset to investigate the connection between recorded accidents and factors to accident severity in Dubai. Two sets of class association rules were generated using the two algorithms and summarized to get the most interesting rules using technical measures. Exact results demonstrated that the class association rules created by Apriori algorithm were more viable than those created by Predictive Apriori algorithm. More relationship between accident factors and accident severity level were investigated while applying Apriori algorithm [2].

1. DRAWBACK OF EXISTING

- In the Previous system data preprocessing clustering and classification is done differently indifferent application
- In Previous system Accident conditions will not be considered for a fatality .(Collision Type , Speed Limit , Light Condition , Weather Condition , Roadway Surface Condition)

Accuracy of the model is low and analysis is not clear.Accuracy of the model is 88.6%.

2. SYSTEM OVERVIEW

5.1 Stage

5.1.1 Data Preparation

Data preparation was performed before each model construction. All records with missing value in the chosen attributes were removed. Fatal rate were calculated and binned to two categories: high and low.

5.1.2 Modeling

Firstly we calculated several Statistics from the dataset to show the basic characteristics of the fatal accidents. Then we applied L, clustering, and Naive Bayse classification to find relationshipsamong the attributes and the patterns.

5.1.3 Result

The results of our analysis include association rules among the variables, clustering of city, states in the country on their populations and number of fatal accidents, and classification of the regions as being high or low risk of fatal accident. We used the data analytic tool Weka to perform these analysis.

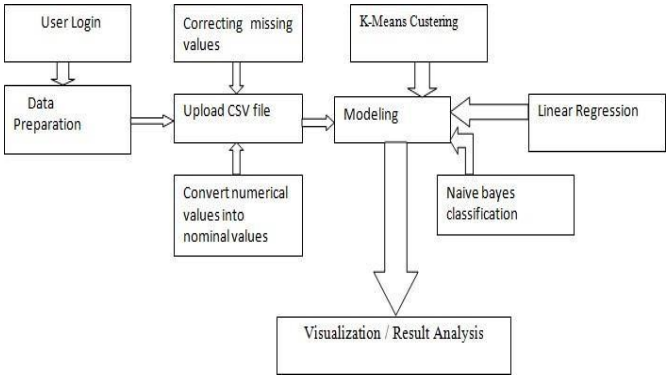

Fig No.1 System Architecture

5.2 System Requirements

5.2.1 Hardware requirements

Hard disk:

128 GB

RAM: 512 MB

Processor: Pentium and above Input device:

Keyboard and Mouse

Output device: Monitor

5.2.2 Software requirements

Operating System: Windows

7/Linux Front End: HTML, JS,Bootstrap

Back End: MySQL, Oracal 10 g, Python, PHP UML Design: Star Uml

6. ANALYSIS

6.1 Mathematical Model

A mathematical model is a description of a system using mathematical concepts and language. The process of developing a mathematical model is termed as mathematical modeling

6.2 Set theory

Let the system be described by S, S= {I, P, R, O}

Where,

S: is a System.I:

is Input

R: is set of Rules
O: Final Output. I= {I1; I2 ;}

Where,

I1 = Dataset/Accident recordsI2 =

Username and

Password

P is set of procedure or function or processesor methods. P = {P1, P2, P3};
Where,
P1 = Check login for Admin.

P2 = Check linear regression to predict the independent Variable.

P3 Check Naïve Bayes for

Classification R is set of

RulesR=R1, R2;

R1 = Enter Valid Information for login. O = {O1, O2}

Where,

O1 = Result analysis Report.

O2 = Predict accidental zone

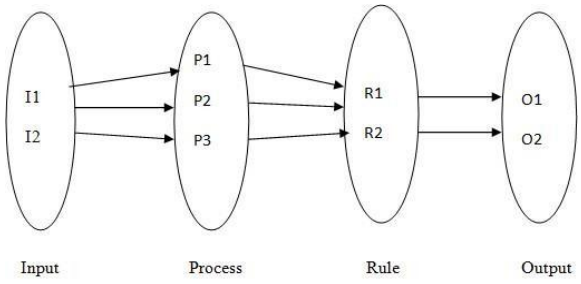

Fig 2:Venn Diagram

Fig. 2 shows Venn diagram

Where,

I1, I2 are inputs,

P1, P2, P3 are process R1. R2 is rules And O1, O2 are output.

1. RESULT ANALYSIS

The percentage of fatal accidents is depend on four variables: SPEED_LMT (speed limit), LIGT_CONDITION (light condition), WEATHER_CONDITION (weather condition) and SURFACE_COND (road surface condition).

Collision Type: The percentage of fatal accidents happened on different collision types the percentageof people and fatal involved are much higher than thepercentage of accident number, which reveals that head-on collision has higher fatal rate in a fatal accident.

- **Speed Limit:** The percentage of fatal accidents happened at different speed limit.
- **Light Condition:** The percentage of fatal accidentshappened on different light condition. Most fatal accidents happen in day light condition because much more roadway.
- **Weather Condition:** The percentage of fatal accident happened on different weather. Most fatalaccidents happened at clear/cloud weather.
- **Surface Condition:** It's the percentage of fatal accident happened on different roadway surface condition. Most fatal accidents happened on dry surface. This is understandable because the most usual case of road condition is that the road surfaceis dry.

(1) Table No.1 Cleaned Data for association rule mining & classification

Light	weather	surface	collision type	drunk driver	rate
daylight	clear/cloud	dry	not collision with motor vehicle in transport	no	low
dark but lighted	clear/cloud	dry	angle-front-to-side, right angle (includes broadside)	no	low
dusk	clear/cloud	dry	sideswipe – same direction	no	low
daylight	clear/cloud	dry	angle-front-to-side, opposite direction	no	low
dark	clear/cloud	dry	angle-front-to-side, right angle (includes broadside)	no	low
daylight	clear/cloud	dry	not collision with motor vehicle in transport	no	low
daylight	clear/cloud	dry	front-to-front (include head-on)	no	low
daylight	rain	wet	angle-front-to-side, opposite direction	no	low
dark	clear/cloud	dry	front-to-front (include head-on)	no	low
dark	clear/cloud	dry	not collision with motor vehicle in transport	yes	low
dark	clear/cloud	dry	front-to-front (include head-on)	no	low
dark but lighted	clear/cloud	dry	not collision with motor vehicle in transport	no	low
.....	.....	...	.....	...	...

Table 2: Results of the Native bayes classification

	TP rate	FP rate	Precision	Recall	F-Measure	ROC Area	Class
	0.996	0.996	0.681	0.996	0.809	0.561	High
	0.004	0.004	0.342	0.004	0.009	0.561	Low
Weighted Avg.	0.679	0.679	0.573	0.679	0.553	0.561	

	TP rate	FP rate	Precision	Recall	F-Measure	ROC Area	Class
	0.996	0.996	0.681	0.996	0.809	0.561	High
	0.004	0.004	0.342	0.004	0.009	0.561	Low
Weighted Avg.	0.679	0.679	0.573	0.679	0.553	0.561	

B. 7. TESTING

We perform the Performance testing of our system. Performance testing will determine whether software meets speed, scalability and stability requirements under expected workloads. For this we uses Apache J meter testing tool.

Load testing - checks the application's ability to perform under anticipated user loads. The objective is to identify performance bottlenecks before the software applicationgoes live.

Stress testing - Involves testing an application under extreme workloads to see how it handles high traffic or data processing. The objective is to identify the breaking point ofan application.

Scalability testing - The objective of scalability testing is to determine the software application's effectiveness in "scaling up" to support an increase in user load. It helps plan capacity addition to your software system.

C. 6. CONCLUSION

As seen in statistics, Linear Regression, and the classification, the environmental factors like road surface, weather, and light condition do not strongly affect the fatal rate, while the human factors like being drunk or not, and the collision type, have stronger on the fatal rate. From the clustering result we could see that some states/regions have higher fatal rate, while some others lower, through the task performed, we realized that data seems never to be enough to make a strong decision. If more data, like non-fatal accident data, weather data, mileage data, and so on, are available, more test could be performed thus more suggestion could be made from the data.

D. ACKNOWLEDGMENTS

I would like to take this opportunity to express my profound gratitude and deep regard to my Project Guide Prof. M. V. Kumbharde, for his exemplary guidance, valuable feedback and constant encouragement throughout the duration of the project. Also the valuable help of Prof. I. R. Shaikh (HOD Comp. Dept.) and Prof. V. N. Dhakane (PG coordinator) who provided facilities

to explore the subject with more Enthusiasm. I express my immense pleasure and thankfulness to all the teachers and staff of the Department of Computer Engineering, S.N.D. College of Engineering and Research Center, Yeola, Nasik for their co-operation and support.

II. REFERENCES

[1] Amira A El Tayeb, Vikas Pareek, and Abdelaziz Araar. Applying association rules mining algorithms for traffic accidents in dubai. International Journal of Soft Computing and Engineering, September 2015**.**

[2] William M Evanco. The potential impact of rural mayday systems on vehicular crash fatalities. Accident Analysis & Prevention, 31(5):455–462, September 1999.

[3] K Jayasudha and C. Chandrasekar. An overview of data mining in road traffic and accident analysis.
Journal of Computer Applications, 2(4):32– 37,
2009.

[4] S. Krishnaveni and M. Hemalatha. A perspective analysis of traffic accident using data mining techniques. *International Journal of Computer Applications*, 23(7):40–48, June 2011

[5] Sachin Kumar and Durga Toshniwal. Analysing road accident data using association rule mining. In Proceedings of International Conference on Computing, Communication and Security, pages 1–6, 2015.

[6] Eric M Ossiander and Peter Cummings. Freeway speed limits and traffic fatalities in Washington state. *Accident Analysis & Prevention*, 34(1):13–18,

[7] KMA Solaiman, Md Mustafizur Rahman, and Nashid Shahriar. Avra Bangladesh collection, analysis & visualization of road accident data in Bangladesh. In *Proceedings of International Conference on Informatics, Electronics & Vision*,pages 1–6. IEEE,

III. BIOGRAPHIES

SATHISH KUMAR G is an Undergraduate student, department of computer science and engineering inPaavai College of engineering.

NITHEESH R is an Undergraduate student, department of computer science and engineering inPaavai College of engineering.

PRAKASH M is an Undergraduate student, department of computer science and engineering inPaavai College of engineering.

Mrs. DIVYA. K. M.E., Assistant Professor, Department of computer science and Engineering, Paavai College of engineering.