

STREET ACCIDENT PREDICTION USING INFORMATION MINING TECHNIQUES

¹Mr. Khapare Pankaj P., ²Prof. Kumbharkar M. V.

¹ P G Student Department of Computer Engineering SND College of Engg. Yeola

²Assistant Professor Dept. of Computer Engineering, SND College of Engg. Yeola, India

pk87924@gmail.com

Abstract - Street visitors accident are taken into consideration as primary public fitness problem. Which will supply safe driving hints, cautious analysis of avenue traffic facts is critical to discover out variables which can be carefully associated with fatal injuries. In this paper we apply probability analysis and records mining algorithms on FARS Fatal Accident dataset as an attempt to deal with this problem. The relationship between fatal rate and different attributes together with collision way, weather, floor condition, mild condition, and under the influence of alcohol motive force were investigated. Category version turned into constructed by naive bayes classifier, and clusters were fashioned by way of easy k-way clustering algorithm. Sure safety riding suggestions were made based on chance, type version, and clusters acquired.

Index Term -Roadway fatal accidents, classification, clustering, FARS

1. INTRODUCTION

There are is lot of motors riding on the avenue each day, and visitors injuries should happen at any time. Some of them accident entails fatality, method human beings die in that coincidence. As human being, we all want to avoid coincidence and stay secure. To locate out how to pressure safer, records mining method could be implemented on the traffic twist of fate dataset to discover some treasured statistics, consequently give driving proposal. Information mining uses many different techniques and algorithms to find out the dating in large amount of records. It is taken into consideration one of the maximum essential tool in information era in the preceding many years [2]. Association rule mining algorithm is a famous method to perceive the massive relations between the statistics stored in massive database and additionally plays a completely crucial function in common object set mining [1]. A classical association rule mining approach is the apriori set of rules who important project is to locate common item sets, that's the approach we use to research the road traffic statistics. Class in records mining ambitions at constructing a version (classifier) from a schooling facts set that may be used to categorise information of unknown elegance labels. The naive bayes technique is one of the very fundamental probability-based totally methods for classification this is based at the bayes' hypothesis with the presumption of independence among each pair of variables. We used fars dataset for our system.

DRAWBACK OF EXISTING

1. Inside the previous machine records preprocessing clustering and class is accomplished in another way in exceptional software
2. In preceding gadget twist of fate conditions will no longer be taken into consideration for a fatality .(collision kind , velocity restriction , light situation , climate situation , roadway floor situation).

II. PROPOSED SYSTEM

1. Classification in statistics mining technique targets at constructing a model (classifier) from a schooling statistics set that may be used to classify facts of unknown elegance labels
2. The naive bayes method is one of the very primary probability-based strategies for classification this is based totally at the bayes' hypothesis with the presumption of independence among each pair of variables.
3. Specific forms of parameter don't forget for fatality. linear regression is beneficial for locating relationship between continuous variables.

III. SYSTEM OVERVIEW

3.1 Stage

3.1.1 Data Preparation

Records guidance changed into carried out earlier than every version production. All records with missing cost within the chosen attributes have been removed. All numerical values had been converted to nominal cost in keeping with the information dictionary. Fatal price had been calculated and binned to two categories: high and low.

3.1.2 Modeling

First off we calculated numerous facts from the dataset to reveal the simple characteristics of the deadly accidents. Then we implemented 1, clustering, and naive bayse category to locate relationships many of the attributes and the styles.

3.1.3 Result

The outcomes of our evaluation include association regulations some of the variables, clustering of city, states in the usa on their populations and range of fatal injuries, and type of the areas as

being high or low threat of fatal coincidence. We used the information analytic tool weka to carry out these analysis.

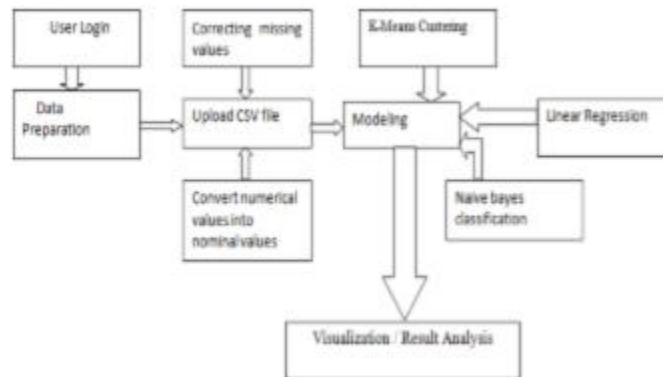


Fig No.1 System Architecture

3.2 System Requirements

3.2.1 Hardware requirements

Hard disk: 128 GB RAM: 512 MB

Processor: Pentium and above Input device: Keyboard and Mouse Output device: Monitor

3.2.2 Software requirements

Operating System: Windows 7/Linux Front End: HTML, JS, Bootstrap

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

IV. ANALYSIS

4.1 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. As the project is having finite input and finite output, it comes under P-Problem.

4.2 Set theory Let the system be described by S, $S=\{I, P, R, O\}$

Where, S : is a System.

I : is Input

ISSN

R : is set of Rules

O : Final Output. $I = \{ I1; I2; \}$

Where,

I1 = Dataset/Accident records

I2 = Username and Password

P is set of procedure or function or processes or 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 Rules

$R = R1, R2$; R1 = Enter Valid Information for login. $O = \{ O1, O2 \}$

Where,

O1 = Result analysis Report.

O2 = Predict accidental zone

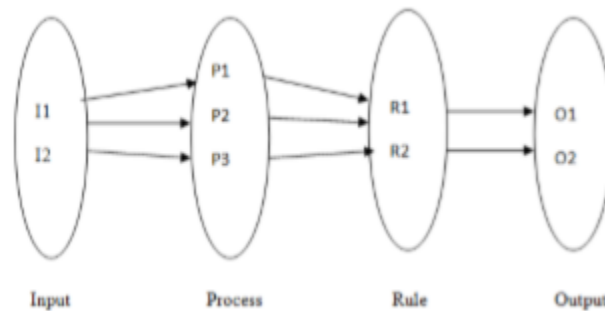


Fig. 2 shows Venn diagram

Where,

I1, I2 are inputs,

P1, P2, P3 are process R1. R2 is rules

And O1, O2 are output.

V. RESULT ANALYSIS

The proportion of deadly injuries is rely upon 4 variables: speed_lmt (speed restriction), lig_t_condition (mild condition), weather_condition (climate circumstance) and surface_cond (avenue floor circumstance). 1. Collision type: the proportion of deadly injuries took place on distinct collision kinds. In front-tofront (head-on collision), the proportion of people and fatals involved are a whole lot higher than the share of twist of fate wide variety, which famous that head-on collision has higher fatal price in a deadly twist of fate.

2. Pace limit: the percent of fatal accidents occurred at specific speed restriction.

3. Light situation: the proportion of deadly injuries passed off on special mild situation. Most deadly accidents happen in day mild condition because an awful lot greater roadway visitors occurs in day time apart from at night time.

4. Climate situation: the percentage of fatal twist of fate befell on distinct climate. Maximum deadly injuries occurred at clean/cloud climate. That is understandable due to the fact clear/cloud is the most typical case of climate condition.

5. Floor situation: its the percentage of deadly twist of fate passed off on special roadway surface condition. Most deadly injuries passed off on dry floor. That is comprehensible due to the fact the maximum usual case of street circumstance is that the road floor is dry

Table No.1 Cleaned Data for association rule mining & Classification

Light	weather	surface	collision type	dunk 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	

VI. TESTING

We perform the performance testing of our machine. Performance checking out will decide whether or not software program meets

speed, scalability and stability requirements below anticipated workloads. For this we makes use of apache jmeter testing tool. Load testing - checks the application's capability to perform underneath predicted consumer loads. The objective is to discover performance bottlenecks earlier than the software application is going live. Pressure trying out - entails checking out an utility beneath severe workloads to look how it handles excessive traffic or facts processing. The goal is to perceive the breaking point of an utility. Scalability checking out - the goal of scalability testing is to decide the software software's effectiveness in "scaling up" to aid an growth in consumer load. It facilitates plan capability addition for your software system.

VII. CONCLUSION

As seen in records, linear regression, and the class, the environmental elements like avenue surface, climate, and mild condition do no longer strongly have an effect on the deadly fee, while the human factors like being under the influence of alcohol or now not, and the collision type, have more potent have an effect on on the fatal charge. From the clustering end result we ought to see that some states/regions have higher fatal price, whilst some others decrease. We may also pay greater interest whilst driving within those risky areas. Via the challenge performed, we realized that data appears by no means to be enough to make a sturdy selection. If greater records, like non-fatal twist of fate statistics, climate data, mileage statistics, and so on, are available, more take a look at will be accomplished for that reason extra idea may be made from the facts

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,