Road Accident Data Analysis Using Machine Learning Techniques

Sannaila Praveen Kumar[1], M. Prashanti[2]

PG Scholar, Department of CS&SE, College of Engineering (A), Andhra University, Visakhapatnam, India[1]

Research Scholar, Department of CS&SE, College of Engineering (A), Andhra University, Visakhapatnam, India [2]

ABSTRACT: Today, traffic safety is one of the main priorities of governments. Considering the importance of topic, identifying the factors of road accidents has become the main aim to reduce the damage caused by traffic accidents. Data mining is the process of analysing data from different perspectives and summarizing it into useful information. Data mining allows users to analyze data from many different dimensions or angles, categorize it and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. One of the key objectives in accident data analysis is to identify the main factors associated with a road and traffic accident. association rule mining is used to identify the various circumstances that are associated with the occurrence of an accident. In this paper we are applying machine learning algorithms to find out how the vehicle speed is contributing to road accidents.

KEYWORDS: Association rule, Data mining, Machine Learning, Multi-Layer Perceptron.

I. INTRODUCTION

Consider the issue of providing a safety travelling measures on the road network within the urban and suburban one of the fundamental principles governing the engineering, traffic and transportation planning. Nearly 3,500 people die on the world's roads every day lots of people are injured or disabled every year. Children, pedestrians, cyclists and the elderly are among the most effected persons while using roads. WHO is combined with the partners of government and non-governmental to raise the issues including in accidents and try to suggest preventions on causing accidents suppose encourage good practices like using helmet is compulsory for two wheelers and wear seat belt in four wheelers, not drinking and driving not speeding and being visible in traffic. The cost of Road accidents can be very heavy burden to the government. Road accidents are a serious threat to public transportation. Transport system administrators also need to take the needs for road safety measures and campaign to the people and also take necessary actions to improve the road conditions reduction of road accidents is only possible by educating the drivers with effective traffic engineering needs. Data mining is an emerging technique in data analysis that in the recent years due to increased ability of collecting and storing data has been considered today rapid increase in the volume of databases is a form that human ability to understand this data is not possible without powerful tools. In this situation, decision-making rather than relying on the information relies on managers and users, because decision makers do not have powerful tools for extracting valuable information. Statistical models for analysis of road accidents and geometric relationship between the crashes and environmental factors are widely used. The main goal of researches have been analyzing data to extract patterns using data mining algorithms. Data mining is implied, previously unknown and potentially useful data extraction process in databases.

   Determining previous information, data cleaning and initial processing, constructing data warehouse, selecting target data set, finding used features and determining new features, visualizing data, selecting data mining operations, pattern extraction, evaluating and interpreting results and deleting inadequate patterns and finally and interpreting results of data and concluding from valuable information different methods have been used for accident data. However, a particular application so far that has been less in traffic safety analysis is association rules analysis and machine learning algorithms. Determining previous information, data cleaning and initial processing, constructing data warehouse, selecting target data set, finding used features and determining new features, visualizing data, selecting data

mining operations, pattern extraction, evaluating and interpreting results and deleting inadequate patterns and finally and interpreting results of data and concluding from valuable information.

II. LITERATURE REVIEW

Many researchers and scientists doing their research on road accident data analysis for finding the major and hidden factors that are causing accidents. Researchers proposed many prediction models for predicting accidents some of them are;

Road accident is one of the undesirable events that are uncertain and unpredictable. Road accident is one of the major causes of unnatural deaths, disability and property damage. MORTH, 2014 [1], proposal stated that 0.4 million accidents are reported each year in India that makes India as one of the countries with larger accident rate. The report says that there is slightly reduce rate recorded in accidents from 2012 to 2013 but these changes are not permanent. Raj V. P. [2] discussed that in India, the method used for collecting, compiling and recording the accident data needs lot of improvements. He states that the reports which are prepared at accident sites are very basic and non-analytical for research purpose. Authors from other countries have used quality data for accidents and analyzed those using statistical techniques. Poisson models [3-4] and negative binomial(NB) models [5-7] have been used extensively to identify the relationship between traffic accidents and the causative factors. In India, the technology used for data collection related to transportation systems is not up to mark as compared to other countries i.e. UK, US and European countries. Various studies in India [3-4] shows that accident data for analysis were collected on manual basis and hence this data is for short duration and use of data mining techniques cannot extra much information from this data as data mining requires large data to extract meaningful results.

III. PROPOSED METHODOLOGY

Data Mining is defined as the procedure of extracting information from huge sets of data. In other words, we can say that data mining is mining knowledge from data. Data pre-processing is a proven method of resolving such issues. Data pre-processing prepares raw data for further processing. This step consists of attribute selection, data cleaning and data transformation. Data gathered from different sources was consolidated, mapped and scrutinized. Some of the data that is not pertinent to the data mining exercise was ignored. Data cleaning is a process used to determine inaccurate, incomplete or unreasonable data and then improve the quality through correcting of detected errors and omissions. Generally, data cleaning reduces errors and improves the data quality. Some entries were clearly invalid, caused by either human error or the evolution of the problem reporting system. Those errors that were correctable were corrected. If all errors detected for a report were not corrected, that report was discarded from the study. Data transformation converts a set of data values from the data format of a source data system into the data format of a destination data system. In this process, few attributed were transformed into required formats for example, the attribute "Hour" was converted into 24-hour format. Also, the attribute values were hard-coded for better representation in the training dataset. A formal presentation of the rule and parameters of confidence, support and lift which quantify a rule is given below. The general form of the rule is as follows: "IF event X occurs THEN event Y occurs as well, in M% of times, and this pattern occurs in N% of all events in the dataset". Association rules defines correlation among a set of items, the relationship is defined in terms of the frequency of co-occurrence or appearance of the items together in each transaction process. In Association rule mining various rules are created in the form X=>Y, X is termed as antecedent whereas 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 associate rat reduce the feature selection method toning this paper we are proposing feature selection method using random forest to select the attributes that are mostly effecting the road accidents , apriority algorithm is applied on selected attributes to find the most important rules by discarding the irrelevant rules. This paper also proposing multi-layer perceptron algorithm to find how vehicles speed contributing the road accidents.

IV. RESULTS

Association rule mining is applied on dataset to find the feasible rules that are mostly contributing the road accidents. Feature selection is applied on the pre-processed dataset to find most effecting rules in road accidents by selecting the appropriate attributes. Selected attributes are given fig 1;
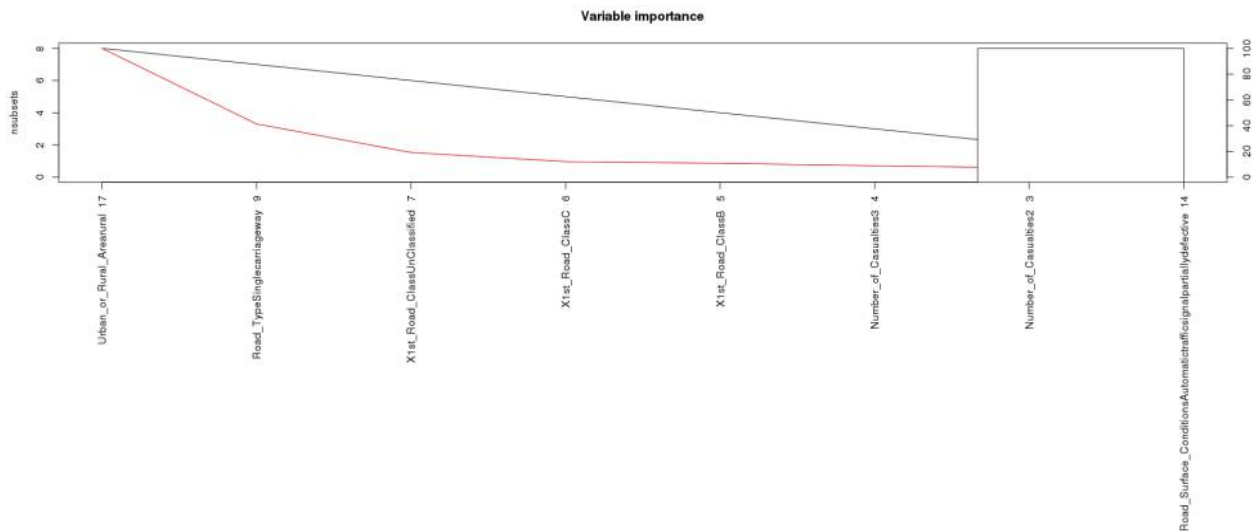

Fig.9: Selecting Attributesusing Feature Selection.

Rules generated before and after feature selection is given in Table 1.

Table 1. No of rules before and after feature selection

Algorithm	Before	After
Apriori	9864	113

The table clearly indicating that feature selection improves the road accidents pattern by selecting only the important rules. This paper implemented MLP to find how the speed of vehicles contributing the Road accidents and the Metric values of the algorithm are given in Table2.

Table 2. Metrics of MLP

Algorithm	Accuracy	Sensitivity	Specificity
MLP	0.8575	0.8545	0.8589

V. CONCLUSION

This paper proposed apriori algorithm and feature selection method to find the most effecting rules that are contributing the road accidents. This paper implemented machine learning algorithm to find the contributing factor of vehicle speed in road accidents. The proposed model can efficiently find the patterns involved in road accidents.

REFERENCES

1. MORTH,. "Road Accidents in India 2013", New Delhi: Ministry of Road Transport and Highways Transport Research Wing,Government of India. August, 2014.
2. R. V. Ponnaluri.," Road traffic crashes and risk groups in India: analysis, interpretations, and prevention strategies, IATSS Research, 35(2), 2012.
3. B. Jones, L. Janssen, and F. Mannering, "Analysis of the Frequency and Duration of Freeway Accidents in Seattle", Accident Analysis and Prevention, Elsevier, vol. 23, 1991.
4. S. P. Miaou., and H. Lum, "Modeling Vehicle Accidents and Highway Geometric Design Relationships", Accident Analysis and Prevention, Elsevier, vol. 25, 1993.
5. S. P. Miaou., "The Relationship between Truck Accidents and Geometric Design of Road Sections–Poisson versus Negative Binomial Regressions", Accident Analysis and Prevention, Elsevier, vol. 26, 1994.
6. M. Poch., and F. Mannering., "Negative Binomial Analysis of Intersection-Accident Frequencies", Journal of Transportation Engineering, vol. 122, 1996.
7. M. A. Abdel-Aty., and A. E. Radwan., "Modeling Traffic Accident Occurrence and Involvement", Accident Analysis and Prevention,Elsevier, vol. 32, 2000.
8. M. Parida., S. S. Jain., and C. N. Kumar., "Road traffic crash prediction on national highways" Indian Highways, Indian Road Congress, 40(6), 2012.
9. M. Parida., S. S. Jain., and C. N. Kumar., "Road traffic crash prediction on national highways" Indian Highways, Indian Road Congress, 40(6), 2012.