A Survey on Analyses of Factors Related to Road Accidents Using Data Mining Techniques

Suraj D, Sandeep Kumar S

M.Tech Student, Associate Professor
Department of Computer Science and Engineering

_____

Abstract: **This paper surveys on Analyses made with respect to factors influencing road accidents. Road accidents causes major damages to humans. It can cause lifetime injuries. There is a spike in the number of accidents over the recent years. So it is an important concern for departments that deal with road safety and for the citizens. Road traffic data must be critically analyzed to determine the factors that are closely related to road accidents. Factors like collision type, road condition, light effect, weather and drunk driver must be considered. Places near residential areas, zebra crossing, school regions are the major zone of accidents. This paper makes comparative study on the analyses done so far in this regard. It proposes an effective means for analyzing the road accident dataset.**

Index Terms— **Data Mining, Association Rule Mining, Classification, K-means clustering, Correlation Analysis, Exploratory Visualization Techniques**

_____

I. INTRODUCTION

There are 1.2 billion vehicles on world's roads now and expected to be 2 billion by 2035. This alarming rise in the number of vehicles causes the rise in number of accidents. Nearly 1.3 million people die every year due to road accidents. Traffic departments and citizens want to avoid accidents. Road traffic dataset is very huge and data mining methodologies are used to find the relation in this dataset. Effective use of accident data depends on the correctness of data, how well the records are retained and how well the data is analyzed. Heterogeneity is innate with accident data. Clustering must be used to segment the accident data. Studies shows that many of the accidents are due to improper interpretation of accidental factors, imperfect road design. With proper analyzation of accident, effective precaution can be taken to avoid accidents and major causalities.

Association rules helps to find relationships such as patterns, correlations and casual structures between seemingly unrelated data in a large data set. Classification is a data mining technique that assigns categories to a collection of data in order to aide in more accurate predictions and analysis. Also called sometimes called a Decision Tree, classification is one of several methods intended to make the analysis of very large data sets effective. Naive's 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. Clustering is a data mining technique that makes a meaningful or useful cluster of objects which have similar characteristics using the automatic technique. The clustering technique defines the classes and puts objects in each class.

II. RELATED WORK

One of the main focus of government is road safety. Finding out the reasons for road accidents helps to reduce and mitigate the damage caused by them. Results obtained by data mining can be used by the transportation department to discover the patterns and trends and foresee the future. Safety measurements can be taken to reduce the number of accidents. Segregating the type of collisions and their severity helps the road construction authority to design better roads and improve the existing ones. India has the more number of accidents in the world according to MORTH-2013. In the intersection of roads that connect major industries and mines, the main culprit are trucks. Other reasons for accidents are no speed breakers, dropping edges of roads. Attempts are done to build safer vehicles, but still accidents are still unavoidable.

Below Table shows the comparative study of

S no.	Research Paper	Focus	Limitations
1	GAGANDEEP KAUR, HARPREETH KAUR: Communication and Networking Technologies (ICCCNT), 2017 8th International Conference, IEEE Conferences, 2017	Depicts crucial aspects such as frequency distribution of enormous data categories and summarize dataset in pictorial form using RStudio and Exploratory Visualization Techniques	Does not analyze the cause of severity of accidents by considering other parameters such as non-restriction of speed, old girth trees on shoulder of roads
2	LILING LI, SHARAD SHRESHTA, GANGZHU HU: 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), IEEE	Certain safety driving suggestions were made based on statistics, association rules, classification model, and clusters obtained.	Factors like non-fatal accident data and weather data are not considered.

	Conferences, 2017		
3	RISHI SAI REDDY SUDHI REDDY; UTTAM MUNDE: 2016 8th International Conference on Computational Intelligence and Communication Networks (CICN), IEEE Conferences, 2016	Map Reduce algorithm is used to reduce the time associated with correlation analysis of road accident data.	It uses small number of nodes and a small dataset. Does not optimize Map reduce performance.
4	NIKHAT IKRAM, SHILPA MAHAJAN: Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference, 2016	Proposes a technique based on VANET and WSN that helps to avoid accident in blind or sharp turns.	Practical implementation may be difficult.
5	SUWARNA GOTHANE, M.V SARODE: 016 6th International Advanced Computing Conference, IEEE, 2016	Identifies the variables influencing accidents using Info Gain Attribute Evaluator function using WEKA tool.	It does not consider the parameters like curve road, straight road, intersection. It does not perform clustering and prediction.
6	SHEEBA RAZZAQ, FAISAL RIAZ, TAHEER MEMOOD: Computing, Electronic and Electrical Engineering (ICE Cube), 2016 International Conference on, IEEE, 2016	Implements Multi-Factors Based Road Accident Prevention System (MFBRAPS) that works on the rules defined in the Fuzzy system and take actions according to the severity of the condition to avoid the road collisions.	Tests are done in the single lane. It does not consider two lanes, T-intersection and on lane changing scenarios
7	SACHINI KUMAR, DURGA TOSHNIWAL: 2015 International Conference on Computing, Communication and Security (ICCCS), IEEE 2015	Uses association rule mining to analyze accident patterns for different types of accidents on the road network of Dehradun district.	Does not use k-means clustering or machine learning techniques to predict the accident prone location

III. PROPOSED WORK

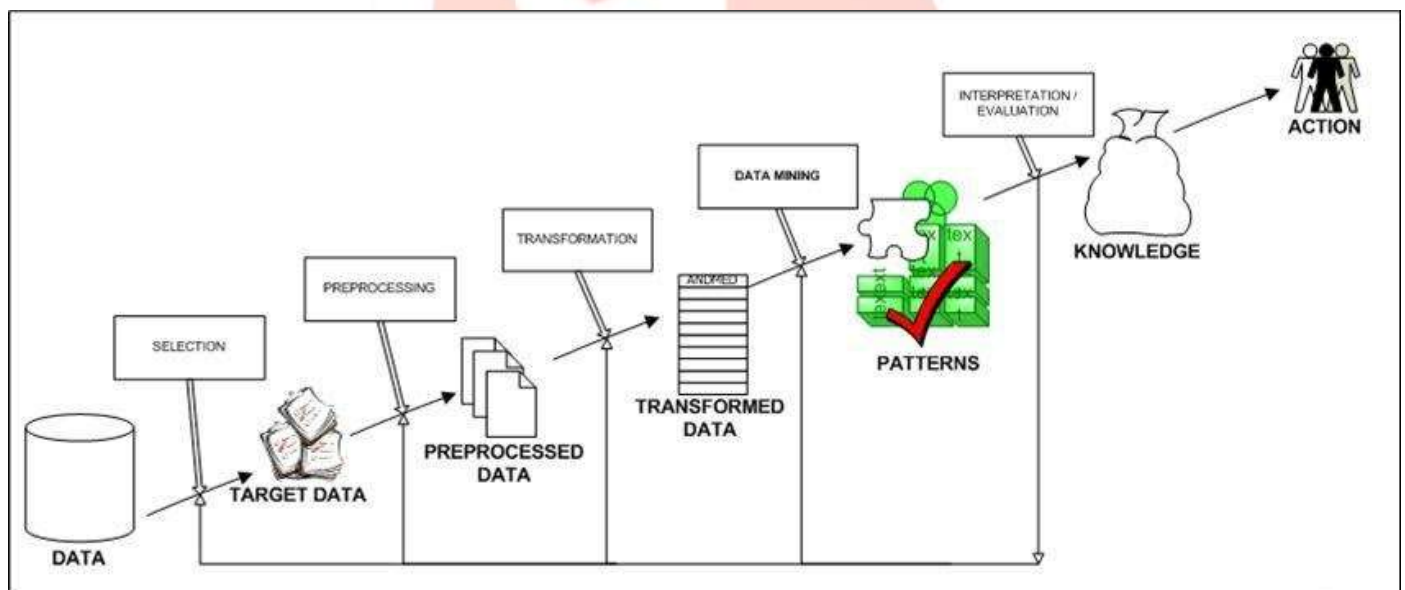The steps involved in data mining are depicted in the figure.

Fig 1: Steps involved in data mining

Collect the realistic data from the transportation department. Data cleaning is done using the imputation method. This phase is also called data preparation. Complexity of the dataset is reduced by Data selection. Discretization technique is used of data reduction. Apply data mining techniques such as Association rule mining, clustering to obtain the correlation between the factors influencing road accidents. Use map reduce programming for correlation to reduce the time taken for correlation analysis.

Then apply following machine learning techniques to predict the accident prone zones for the newly planned roads.

- Support vector machines
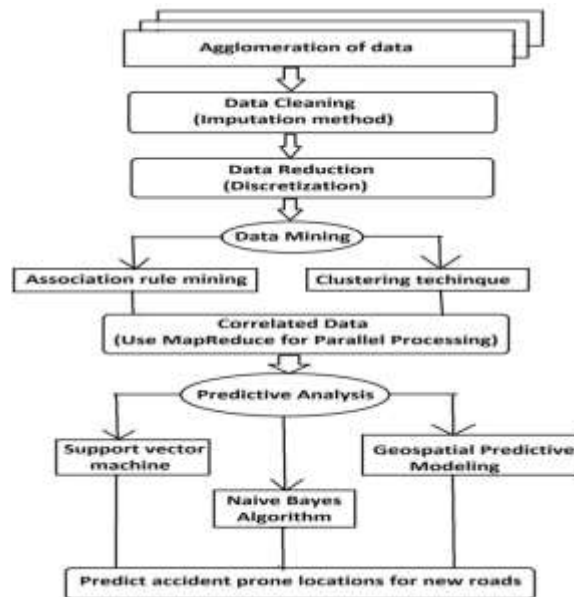- Naive Bayes Classification and
- Geospatial predictive modeling

Fig 2: Steps representing the proposed work

IV. CONCLUSION

Existing work mainly concentrates on only analyzing and finding correlation between the parameters influencing road accidents using the dataset for existing roads. It does not consider the newly planned roads and roads under construction. This paper aims at finding accident prone locations for roads that are planned for future and roads that are under construction.

REFERENCES

[1] Gagandeep Kaur, Er. Harpreet Kaur: "Prediction of the cause of accident and accident prone location on roads using data mining techniques", 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE Conferences, 2017.

[2] 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). IEEE Conferences, 2017.

[3] Rishi Sai Reddy Sudireddy; Uttam Mande: "Prediction of Road Accidents Using Correlation Based on Map Reducing", 2016 8th International Conference on Computational Intelligence and Communication Networks (CICN), IEEE Conferences, 2016.

[4] Nikhat Ikram, Shilpa Mahajan: "Road accidents: Overview of its causes, avoidance scheme and a new proposed technique for avoidance", 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), IEEE Conferences, 2016.

[5] Suwarna Gothane; M. V. Sarode: "Analyzing Factors, Construction of Dataset, Estimating Importance of Factor, and Generation of Association Rules for Indian Road Accident", 2016 IEEE 6th International Conference on Advanced Computing (IACC), IEEE Conferences, 2016.

[6] Sheeba Razzaq, Faisal Riaz; Tahir Mehmood; Naeem Iqbal Ratyal, "Multi-Factors Based Road Accident Prevention System", 2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), IEEE Conferences, 2016.

[7] Sachin Kumar, Durga Toshniwal: "Analysing road accident data using association rule mining", 2015 International Conference on Computing, Communication and Security (ICCCS), IEEE Conferences, 2015.