



UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
DEHRADUN, UTTRAKHAND, INDIA

PROJECT SYNOPSIS

A DATA MINING FRAMEWORK TO ANALYZE ROAD ACCIDENT DATA

BACHELOR OF TECHNOLOGY

CSE (Artificial Intelligence and Machine Learning)

SUBMITTED BY

S. NO.	NAME	ROLL NO.	BATCH	PHONE NO.	EMAIL ADDRESS
1.	Piyush Malviya	R177219128	B4	6261622231	500075183@stu.upes.ac.in
2.	Priyal Gupta	R177219136	B6	9415209873	500076110@stu.upes.ac.in
3.	Rahul Dhanola	R177219139	B4	6397227162	500075154@stu.upes.ac.in
4.	Saksham Garg	R177219154	B5	9996619712	500076000@stu.upes.ac.in

GUIDED BY

Mr. Sumit Shukla

TABLE OF CONTENTS

S NO.	TOPICS	PAGE NO.
1.	Introduction	3
2.	Literature Review	4
3.	Methodology	5
4.	Facilities Required	6
5.	References	8

Figures Used

Fig. No	Fig Name	Page No.
Figure 1	Proposed framework for analysis	7

INTRODUCTION :

Nowadays, road accidents are one of the main causes of mortality worldwide. Then, measures are required to reduce or mitigate the accident impacts. The identification of the most effective measures requires an effective analysis of accidents able to identify and classify the causes that can trigger an accident.

Road accidents are uncertain and unpredictable incidents and their analysis requires the knowledge of the factors affecting them. Fatalities and injuries resulting from road traffic accidents are a major and growing public health problem in India. Every week nearly 2,650 people get killed and 9,000 get injured due to traffic accidents.

Traffic accidents have now earned India a dubious distinction; with nearly 140,000 deaths annually, the country has overtaken China to top the world in road fatalities. India is the only country in the world which faces more than 15 fatalities and 53 injuries every hour as a consequence of road crashes.

The roads of India has not abated their contribution in the traffic accident fatalities. The accident rate of India has been on an increase ever since the start of the century. Data mining analyses can help identify the major causes and help the transport authorities in improving safety requirements

We will study the states and the union territories of India against the contributing causes in order to facilitate road safety in the country. We are focused on taking the aid of clustering to group similar objects off this dataset in order to group regions on the basis of vulnerability.

The major problem in the analysis of accident data is its Heterogeneous nature. Thus, heterogeneity must be considered during analysis of the data. Road accident analysis aims to investigate the main factors that characterize an accident to understand patterns or behaviors and, consequently, to identify the appropriate countermeasures to adopt to avoid the accident.

LITERATURE REVIEW:

[1] A Analysis of road accidents in India using data mining classification algorithms.

By : E. Suganya and S. Vijayarani

In this, researchers have used classification algorithms namely linear regression, logistic regression, decision tree, SVM, Naïve Bayes, KNN, Random Forest and gradient boosting algorithm and measures accuracy, error rate and execution time. The performance of KNN is better than other algorithms.

[2] Data mining approach to analyze the road accidents in India

By : Ayushi Jain, Garima Ahuja, Anuranjana and Deepti Mehrotra

In this, researchers have used a data mining approach to create a model that smooths out the heterogeneity of the data by grouping similar objects together, cluster analysis helps to determine the accident prone states and territories of India. These clusters are labeled to be classified with the help of a decision tree to conclude the dominant factor, backing the accidents.

[3] Road Accident Analysis with Data Mining Approach: evidence from Rome

By : Antonio Comi , Antonio Polimeni , Chiara Balsamo

In this, researchers have aimed to find out which data mining techniques are most suitable to analyze road accidents, to identify the most significant causes and the most recurrent patterns of road accidents. K-means and Kohonen network algorithms are used for descriptive analysis and decision trees and neural networks are used for predictive analysis.

METHODOLOGY :

Proposed Methodology

In this system we will use different data mining algorithms to analyze the data. Different algorithms are applied to group the accident locations into clusters and mining techniques are used to characterize the locations. Most state of the art traffic management and information systems focus on data analysis.

Apart from analysis focus has been done towards the sense of classification. So our system also uses classification techniques to predict the severity of the accident which will bring out the factors behind road accidents that occurred and a predictive model is constructed using fuzzy logic to predict the location wise accident frequency. This system uses road accidents data to mine frequent patterns and important factors causing different types of accidents.

Proposed framework

- **Data preprocessing:** Data preprocessing is one of the important tasks in data mining. Data preprocessing mainly deals with removing noise, handling missing values, removing irrelevant attributes in order to make the data ready for the analysis. In this step, our aim is to preprocess the accident data in order to make it appropriate for the analysis.
- **Clustering algorithm:** There are several clustering algorithms in the literature. The objective of the clustering algorithm is to divide the data into different clusters or groups such that the objects within a group are similar to each other whereas objects in other clusters are different from each other. K means & Decision trees have been used in road accident analysis.
- **Association rules:** Association rule mining is a very popular data mining technique that extracts interesting and hidden relations between various attributes in a large data set. Association rule mining produces a set of rules that define the underlying patterns in the data set. The associativity of two characteristics of accident is determined by the frequency of their occurrence together in the data set. A rule $A \rightarrow B$ indicates that if A occurs then B will also occur. Further association rules are generated from the frequent itemsets and strong rules based on interestingness measures are taken for the analysis.

Facilities Required :

System Configuration

This project can run on commodity hardware. First part is the training phase which takes 10-15 mins of time and the second part is the testing part which only takes a few seconds to perform classification and calculate accuracy.

Hardware Requirements

Table 1: Table for Hardware Requirements

Type	Name
System Architecture	32 – bit or 64 – bit
Memory	8 GB 2400 MHz DDR4
Storage	500 GB SATA3 2.4 HDD
Central Processing Unit (CPU)	Intel or AMD - 2 GHz or Faster
Graphical Processor Unit (GPU)	2 GB Nvidia Graphic Processor

Software Requirements

Table 2: Table for Software Requirements

Type	Name
Programming Language with Version	Python 3.5 & above
Operating System	Windows 7 & above. Linux based OS Mac OS

PROJECT SCOPE :

- Identify the basic nature of accidents happening in the selected highway stretches.
- Identify the root cause of accidents based on the collected data.
- Identify the features of the road causing the accident.
- Identify and Compare various road segments for optimization.
- Identify road intersection types and the frequency of accidents.

Proposed framework for analysis :

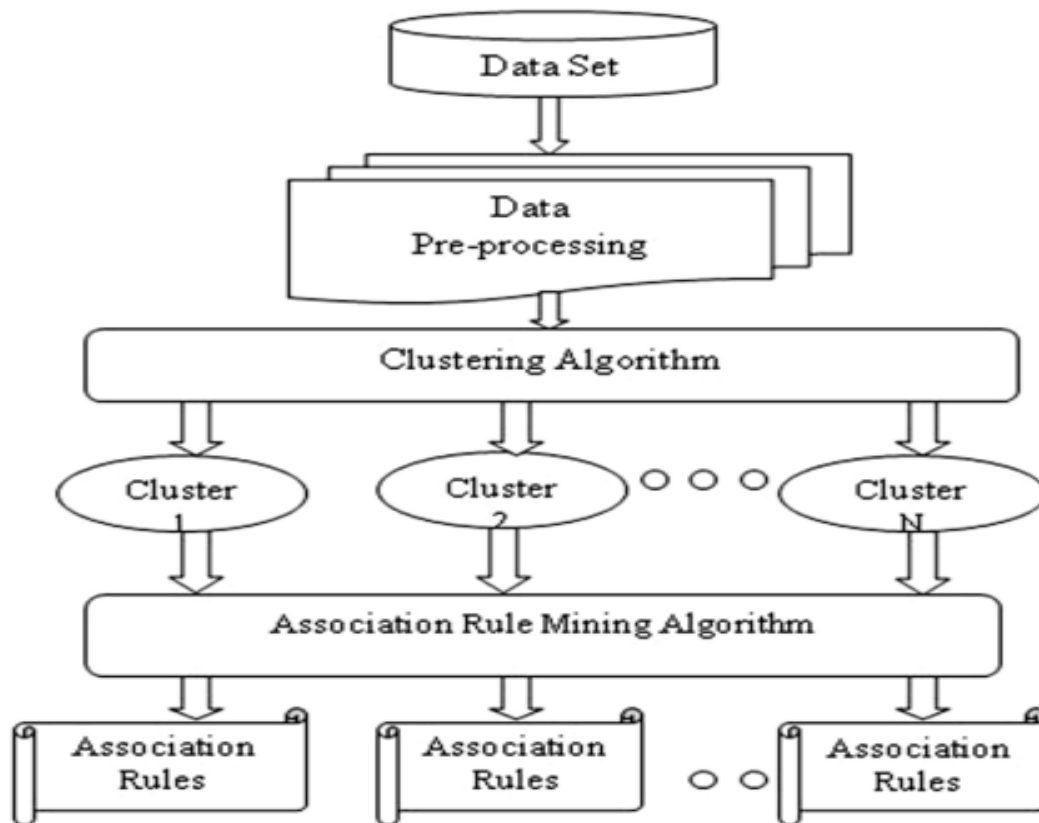


Figure 1: Proposed framework for analysis

REFERENCES :

- [1] Dataset(online) Available:
<https://data.gov.in/catalog/road-accidents-india-2018>
- [2] A. Priyanka and K. Sathiyakumari, "A comparative study of classification algorithm using accident data", International Journal of Computer Science & Engineering Technology (IJCSET)
- [3] Article(online) Available:
<https://www.hindustantimes.com/mumbai-news/india-had-most-deaths-in-road-accidents-in-2019-report/story-pikRXxsS4hptNVvf6J2g9O.html>
- [4] Article(online) Available:
<https://www.hindustantimes.com/mumbai-news/india-had-most-deaths-in-road-accidents-in-2019-report/story-pikRXxsS4hptNVvf6J2g9O.html>
- [5] Road Traffic Accidents in India: Issues and Challenges, Sanjay Kumar Singh
<https://www.sciencedirect.com/science/article/pii/S2352146517307913>
- [6] The Global status report on road safety 2018, launched by WHO in December 2018 <https://www.who.int/publications/i/item/9789241565684>
- [7] Thair Nu Phyu, "Survey of Classification Techniques in Data Mining", Proceedings of the International Multi Conference of Engineers and Computer Scientists 2009
- [8] T. Dipo, I Akomolafe and Akinbola Olutayo, "Using Data Mining Technique to Predict Cause of Accident and Accident Prone Locations on Highways", American Journal of Database Theory and Application 2012
- [9] India: 100% More Traffic Deaths Than China
<https://www.indiaspend.com/india-100-more-traffic-deaths-than-china-19692>