Road Safety Accident Prediction

Mukul Prasad Ray
mukulray83@gmail.com
Lovely Professional University

Arbas Rahaman
rahamanarbas@gmail.com
Lovely Professional University

Harsh Ranjan
harshranjan9966@gmail.com
Lovely Professional University

Pardeep Kumar
Assistant Professor
pardeep.25237@lpu.co.in
Lovely Professional University

ABSTRACT

The number of incidents happening every day is rising alarmingly as a result of the constantly growing number of automobiles on the road. Due to the high amount of traffic fatalities and incidents, today, the transportation department needs to be able to predict the amount of accidents over a specific period of time in order to make informed judgements. In this situation, it will be beneficial to assess the frequency of accidents so that we may utilize this information to further aid in the development of strategies to lessen them. Even if the bulk of accidents are characterized by ambiguity, there is a degree of pattern that is noticed when watching the accidents occurring in different areas in India. Making accurate forecasts about the likelihood of accidents occurring in a given location and creating accident prediction models can both benefit from this regularity. In this essay, we've examined the connections between traffic accidents, road conditions, and the impact of environmental factors on the likelihood of an accident. Using the **AdaBoost Algorithm and Classifier**, we have developed an accident prediction model utilizing data mining approaches. The datasets for Bangalore road accidents from recent years that were made available online were used for this investigation. Several parties, including the government's public works departments, contractors, and other car sectors, can benefit from using the study's findings to improve the design of roads and vehicles on the basis of the data obtained.

Keywords— Accident prediction, Data mining, AdaBoost Algorithm, Rule mining, Classification.

I. INTRODUCTION

India's worrisome rate of accident growth is now a major source of worry. In light of some recent statistics India owns just 1% of all vehicles worldwide, but it is responsible for around 6% of all traffic incidents worldwide. Two-wheeler's irresponsibility is a major cause of recorded accidents, and over speeding is another contributing element. Accidents resulting from drunk driving or other types of traffic offences are also frequent. The negligence of people towards the speed of the vehicle, the condition of the vehicle, and their own negligence of not wearing helmets has led to many accidents despite the existence of set regulations and highway codes. Although the growing number of vehicles is the primary contributor to traffic accidents, the importance of the state of the roads and other environmental elements cannot be understated.

In India, the amount of fatalities from traffic accidents is undoubtedly concerning. With more than 137,000 persons suffering injuries from traffic accidents, the situation is exceedingly grim. This number is more than four times the number of people killed by terrorism each year. Accidents involving large trucks and even those involving buses, which are utilized for public transit, are among the deadliest kinds of accidents that can happen and cost the lives of innocent people. Rain, fog, and other weather conditions can greatly increase the likelihood of accidents. So, it would be easier to take action to reduce accidents if you have a proper estimation of incidents and are aware of accident hotspots and contributing variables. This calls for careful research on accident and development of safety predication models. It is frequently necessary to construct a well-designed road framework management system for considering concerns of road security. The model is intended to create an improved accident prediction model that could examine potential problems brought on by infrastructure failures and estimate

the contribution of current models to minimizing the likelihood of accidents. The appraisal of the relative importance of each variable's contribution to the accident and the question of how the model should be best constructed to take the effects of all such factors into account are two of the primary issues involved in developing such a model. In the past, a range of fields, including but not limited to credit risk management, fraud detection, healthcare informatics, recommendation systems, and other areas, have found data mining techniques and models to be helpful for the goal of data interpretation. These investigations have also been enhanced by methods integrating machine learning and artificial intelligence. In order to write this study, we looked at the connections between the incidence of traffic accidents and the roles that underlying road conditions and environmental factors played in causing the same. In order to capture all the factors impacting accidents in such a study, we can employ these vast amounts of data would normally be meaningless without the proper interpretation being applied to them, therefore data mining techniques are being used to analyze them and extract pertinent information from them.

II. LITERATURE SURVEY

The increasing rate of accidents in India has prompted numerous researchers to investigate the factors influencing road accidents. To create prediction models for road accidents, various categories of data mining techniques have been utilized because they do not require assumptions between dependent and independent variables, which are required in traditional statistical techniques. Researchers have focused on different sets of attributes when developing such models. For instance, Srivastava et al. [2] and Ghazizadeh et al. [3] concentrated on accidents occurring at intersection points and used Multi-layered perceptron (MLP) and feed forward MLP techniques to analyze the impact of several factors such as day or night, traffic conditions, etc. on accidents. Chen et al. [4] found that highways were the common area where most accidents occurred.

Williams et al. [5] discovered that a driver's age and experience played a major role in accident occurrence. Suganya, E. and S. Vijayarani [6] compared the performance of different classification algorithms and found KNN to be the best. Sarkar et al. [7] conducted a comparative study on the type of roads prominent in accidents and found that accidents on highways were more common. Stewart et al. [8] used original data to build a neural network model that predicted accidents more quickly than other models used on Indian roads.

Zheng et al. [9] examined the range of injuries resulting from motor vehicle accidents and analyzed the emotions of drivers involved in the accidents that may have played a role in causing them. Meanwhile, Arun Prasath N and Muthusamy Punithavalli [10] conducted a thorough survey on road accident detection techniques, discussing their advantages and disadvantages over the years.

In their paper, George Yannis et al. [11] discussed the current practices used internationally in the development of accident prediction models, gathering detailed information from questionnaires to identify the most useful models. Similarly, Anand, J.V [12] created a method to determine the impact of different variables on atmospheric deterioration using clustering techniques, which could be applied to evaluate the impact of various factors on road accidents. Tiwari et al. [13] classified road accident data using various techniques, while N. Singh et al. [15] identified accident hotspots to create a model for detecting accident-prone areas. Kaur, G et al. [14] used correlation analysis and visualization techniques to study road accidents and create an accident prediction model for state highways and normal district roads.

III. METHODOLOGY

In this research, we developed an application that utilizes road accident data to predict the likelihood of accidents occurring. The data underwent a pre-processing step that involved cleaning, normalization, and feature selection. The processed data was then subjected to various data mining techniques, including clustering, Support Vector Machines (SVM), and AdaBoost . SVM was used to predict the potential risk of accidents, while AdaBoost was employed for rule mining to generate frequent item sets based on given support and confidence values. Rules were established by considering various combinations of factors that have caused accidents of different types and severities in diverse road and weather conditions. SVM classification was utilized to categorize each accident into a high or low-risk group. The architecture of the model implemented in this research is depicted in Figure 1. Through this analysis, we were able to identify various factors that contribute to road accidents and provide

potential solutions to reduce accident rates and fatalities. The use of different data mining and visualization techniques aided in the interpretation of results.

IV. IMPLMENTATION

A. Dataset

The study utilized data from the Open Government Data (OGD) Platform in India, specifically datasets pertaining to accidents that occurred in Bangalore between 2014 and 2017. The dataset contains various details, such as the date, time, and location of the accidents, the cause of the accident (such as overspeeding or skidding), the type of road (straight or curved), the number of lanes, and whether it occurred at a junction. Additionally, it includes information on the number of fatalities. To analyze this dataset, a stochastic model is required as opposed to a simple deterministic model. Therefore, machine learning algorithms are necessary to supplement the data mining techniques used in the study.

B. System Architecture

The gathered raw data on traffic accidents is first processed to create the dataset that will be used as model input. Using the training set of data, the model is further trained to make predictions and the potential for accidents in a space where a user will enter data. Based on the gathered statistics, the user is also presented with a graphical depiction. The four components of this model's operation are Rule Mining, Risk Prediction, Graph Plot, and New Data Entry. The AdaBoost algorithm is used for rule mining, and we create a frequent item set depending on the input dataset. The SVM (Support Vector Machine) algorithm, which is usually used for classification, is employed for risk prediction. This module forecasts the likelihood of accidents in a specific location. Using information from the weather, past accidents, and causes for the accidents. The new accident cases are reported using the New Data Entry module.

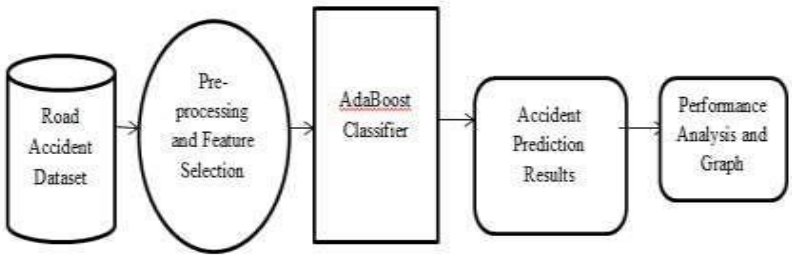

Fig 1. System Architecture of the Implemented Model

C. Software and Language Used

The implementation of the application was done using Anaconda Spyder software and the Python programming language.

D. Simulation

The R tools are utilized to conduct a simulation, applying a range of data mining and exploratory visualization techniques to interpret the accident dataset. Through this, an interactive user interface is developed, enabling the analysis of multiple factors that contribute to accidents through the generation of various graphs, charts, and statistical and graphical representations.

E. Findings of Dataset Analysis

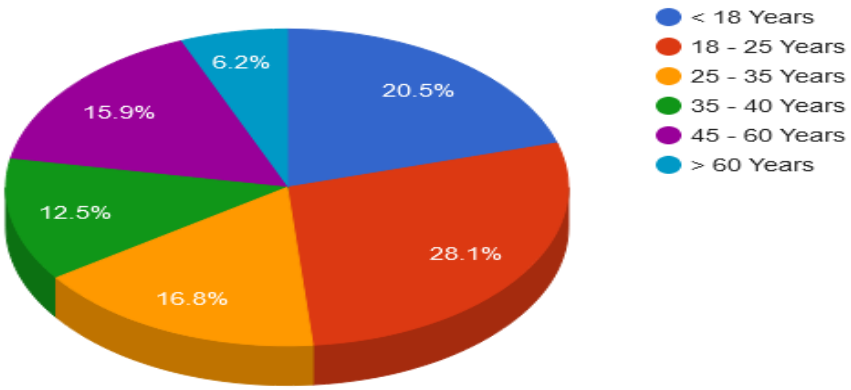

Fig 2. Chart of the Dataset Analysis (Accidents with respect to age)

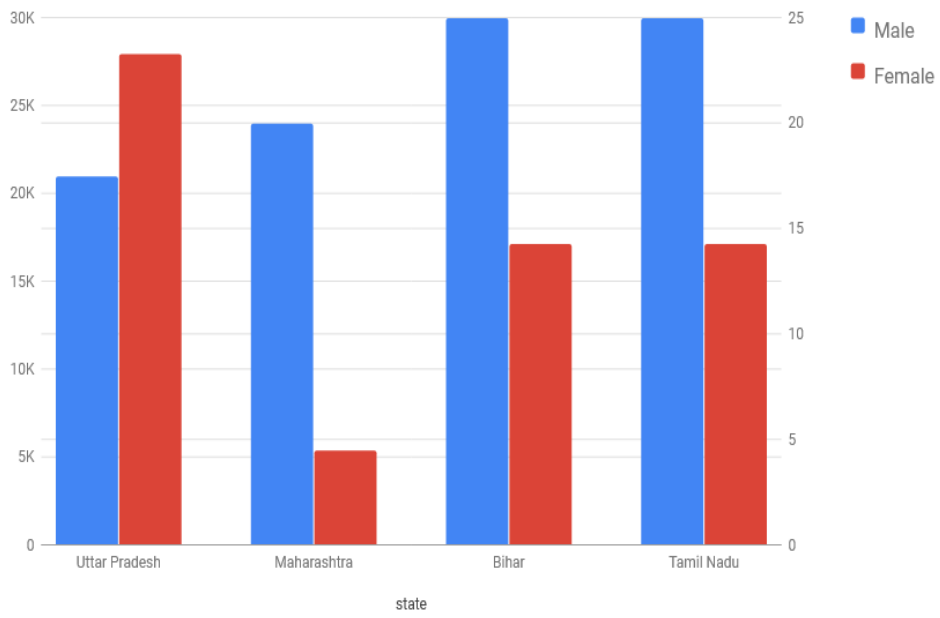

Fig 3. Bar Diagram of Total Accidents in Residential Area (Male and Female)

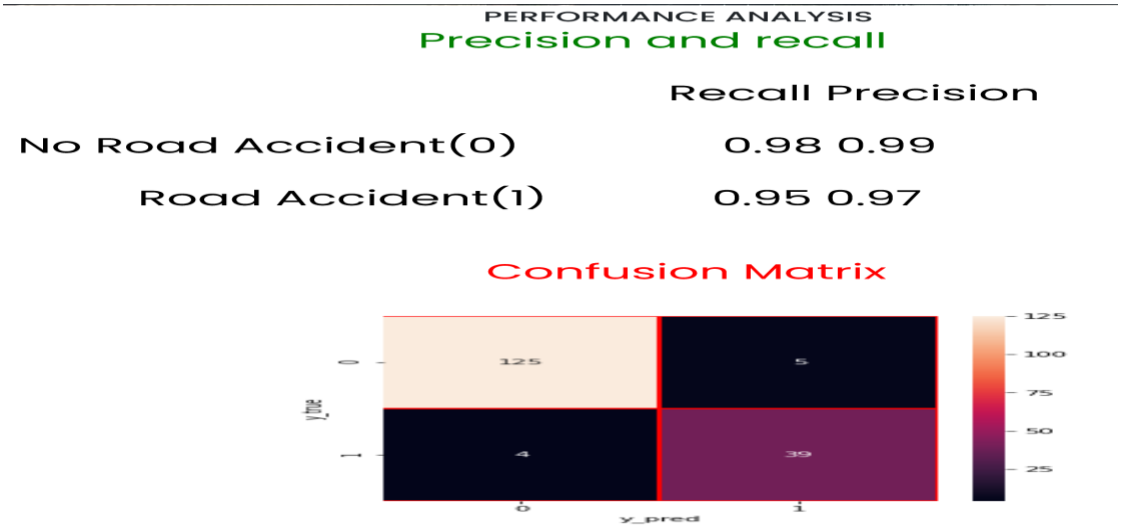

Fig 4. Performance analysis Matrix of the Dataset

V. DISCUSSION

The paper describes the development and implementation of a road accident prediction model that considers various causative factors, such as the condition of the road, weather influences, and the nature of the accident cause. Unlike past literature, this study does not take into account the emotional state of mind and experiential influence of the driver. Figures 5, 6, 7, and 8 provide an overview of the parameters used in creating the prediction model, such as the number of accidents reported for each type of accident, weather conditions observed during accidents, and response activities initiated after each reported accident. The increasing number of accident cases reported due to heavy-duty vehicles is also represented in Figure 8, which is included in the dataset used for this study. The model is integrated into an application that predicts the probability of accident risk over a specified area inputted by the user. The user interface of the application presents a graphical visualization of the factors responsible for causing accidents in the past, and a categorical prediction of high or low risk is made for the selected area. The model helps in understanding the combinations of factors that have resulted in fatal accident scenarios, and an option to enter details of new accident cases has been provided to improve the dataset for future use.

VI. CONCLUSION

The number of lives affected by accidents can be significantly reduced if we take responsibility and adopt safe driving practices. While accidents can occur due to various reasons, road development authorities and automobile industries can also take necessary measures to prevent them. Road design can be improved, and better safety features can be incorporated into vehicles. Predictive models that utilize previous data and observations can also aid in accident prevention. A successful example of such a model is an application that can efficiently predict the likelihood of an accident based on various factors such as vehicle type, driver age, vehicle age, weather conditions, road structure, etc. The application was developed using data mining and machine learning algorithms applied to a dataset for Bangalore and has been highly accurate in predicting the probability of accidents in different areas.

There is potential for the current model to be optimized in the future by incorporating several constraints that were not considered in the current study. These optimized models can be effectively utilized by the government to decrease the occurrence of road accidents and to implement road safety policies. Additionally, a mobile application could be developed to assist drivers in selecting safer routes for their journeys. The maps service could also incorporate a feature that alerts drivers to the risk probability of their chosen route and provides directions. This technology could be adopted by companies such as Uber and Ola in the future, leading to improved surveillance of accident-prone areas and more timely emergency services in the event of an accident. Moreover, the risk information obtained from this model could also be used to install better road safety instructions along highways.

VII. REFERENCES

1. https://www.statista.com/topics/5982/road-accidents-in-india/

2. Srivastava AN, Zane-Ulman B. (2005). Discovering recurring anomalies in text reports regarding complex space systems. In Aerospace Conference, IEEE. IEEE 3853-3862.

3. Ghazizadeh M, McDonald AD, Lee JD. (2014). Text mining to decipher free-response consumer complaints: Insights from the nhtsa vehicle owner's complaint database. Human Factors 56(6): 11891203. http://dx.doi.org/10.1504/IJFCM.2017.089439.

4. Chen ZY, Chen CC. (2015). Identifying the stances of topic persons using a model-based expectation maximization method. J. Inf. Sci. Eng 31(2): 573-595. http://dx.doi.org/10.1504/IJASM.2015.068609.

5. Williams T, Betak J, Findley B. (2016). Text mining analysis of railroad accident investigation reports. In 2016 Joint Rail Conference. American Society of Mechanical Engineers V001T06A009V001T06A009. http://dx.doi.org/10.14299/ijser.2013.01.

6. Suganya, E. and S. Vijayarani. "Analysis of road accidents in India using data mining classification algorithms." 2017 International Conference on Inventive Computing and Informatics (ICICI) (2017): 1122-1126.

7. Sarkar S, Pateshwari V, Maiti J. (2017). Predictive model for incident occurrences in steel plant in India. In ICCCNT 2017, IEEE, pp. 1-5. http://dx.doi.org/10.14299/ijser.2013.01.

8. Stewart M, Liu W, Cardell-Oliver R, Griffin M. (2017). An interactive web-based toolset for knowledge discovery from short text log data. In International Conference on Advanced Data Mining and Applications. Springer, pp. 853-858. http://dx.doi.org/10.1007/978-3- 319-69179-4_61.

9. Zheng CT, Liu C, Wong HS. (2018). Corpus based topic diffusion for short text clustering. Neurocomputing 275: 2444-2458. http://dx.doi.org/10.1504/IJIT.2018.090859.

10. ArunPrasath, N and Muthusamy Punithavalli. "A review on road accident detection using data mining techniques." International Journal of Advanced Research in Computer Science 9 (2018): 881-885.

11. George Yannis, Anastasios Dragomanovits, Alexandra Laiou, Thomas Richter, Stephan Ruhl, Francesca La Torre, Lorenzo Domenichini, Daniel Graham, Niovi Karathodorou, Haojie Li (2016). "Use of accident prediction models in road safety management – an international inquiry". Transportation Research Procedia 14, pp. 4257 – 4266.

12. Anand, J. V. "A Methodology of Atmospheric Deterioration Forecasting and Evaluation through Data Mining and Business Intelligence." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, no. 02 (2020): 79-87.

13. Prayag Tiwari, Sachin Kumar, Denis Kalitin (2017). "Road-User Specific Analysis of Traffic Accident Using Data Mining Techniques". International Conference on Computational Intelligence, Communications, and Business Analytics. 10.1007/978-981-10-6430-2_31.

14. Kaur, G. and 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) (2017): 1-7.

15. Irina Makarova, Ksenia Shubenkova, Eduard Mukhametdinov, and Anton Pashkevich, "Modeling as a Method to Improve Road Safety During Mass Events", Transportation Research Procedia 20 (2017) 43.

16. https://data.gov.in/resource/stateut-wise-total-number-road-accidents-india-2016-2019