A Road Accident Prediction Model Using Data Mining Techniques

Prof. M. N. Jadhav
Department of Computer Technology
K. K. WAGH POLYTECHNIC, Nashik
mnjadhav@kkwagh.edu.in

Pradeep Uttam Barve
student of Computer Technology
K. K. WAGH POLYTECHNIC, Nashik
pradeepbarve755@gmail.com

Yashwant Balasaheb Velamkar
student of Computer Technology
K. K. WAGH POLYTECHNIC, Nashik
yashwantvelamkar@gmail.com

Rushikesh Kishor Gaikwad
student of Computer Technology
K. K. WAGH POLYTECHNIC, Nashik
rushikeshgaikwad364@gmail.com

Aishwarya Santosh Avhad
Student of Computer Technology
K. K. WAGH POLYTECHNIC, Nashik
aishwaryaavhad97@gmail.com

---***---

Abstract - Implantation of study on road accidents and the development of a prediction model using data mining techniques is indeed very important. The use of the Apriori algorithm and Support Vector Machines (SVMs) in developing an accident prediction model is a well-established approach123.The Apriori algorithm is used to explore the causal relationship between multiple factors, such as the condition of a road and environmental factors, to understand the mechanism of accidents4. SVMs, on the other hand, are powerful tools for classification and regression problems and can be used to analyze the severity of crashes. As for the Bangaloreroad accident datasets for the years 2014 to 2017, there are indeed datasets available on the internet that can be used for such studies567. These datasets, along with the insights from your study, can indeed be beneficial for various stakeholders, including government public work departments, contractors, and automobile industries, in better designing roads and vehicles based on the estimates obtained. It's worth noting that road safety is a complex issue that involves various factors. While data analysis and prediction models can provide valuable insights, it's also crucial to consider other aspects such as enforcement of traffic rules, driver education, and infrastructure development to effectively reduce road accidents.

Key Words: Accident prediction, Data mining, Apriori algorithm, Rule mining, Classification

INTRODUCTION

This research is focused on understanding and predicting road accidents. This is a critical issue as the number of vehicles on the road is increasing exponentially, leading to a rise in accidents. By predicting these incidents, we can take preventive measures to reduce them. You're using data from Bangalore, collected between 2014 and 2017, to conduct your study. This data includes various details about each accident, such as the location, time, weather conditions, and more. To analyze this data, you're using two main techniques: the Apriori algorithm and Support Vector Machines (SVMs). The Apriori algorithm is a popular method used in data mining. It helps you find patterns in the data. In the context of your study, it helps you understand how different factors are related to accidents. For example, it can help you find out if accidents are more likely to happen on certain types of roads or in certain weather conditions. Support Vector Machines (SVMs), on the other hand, are used for classification and regression problems. In your study, SVMs help you analyze the severity of crashes. For instance, it can help you understand if accidents on certain types of roads or in certain weather conditions are more likely to be severe. The results of your study can be incredibly beneficial for various stakeholders.

Government public work departments can use your findings to improve road conditions and implement safety measures. Contractors can consider your results when designing and constructing roads. Automobile industries can use your insights to design safer vehicles. In conclusion, your study is a significant step towards making our roads safer. By understanding and predicting road accidents, we can take informed actions to prevent them and save live

I. LITERATURE SURVEY

The escalating frequency of road accidents in India has ignited a fervent quest among researchers to dissect the multifaceted factors underpinning these unfortunate events. Unlike traditional statistical methods that often necessitate rigid assumptions between dependent and independent variables, data mining techniques offer a more malleable approach, affording researchers the latitude to explore diverse avenues in constructing predictive models for road accidents, each honing in on distinct attributes. Srivastava et al. and Ghazizadeh et al., for instance, have embarked on meticulous investigations into accidents transpiring at intersection points, employing sophisticated techniques such as Multi-layered Perceptron (MLP) to discern and categorize the severity of these incidents with remarkable precision. Meanwhile, the probing inquiries by Chen et al. have spotlighted highways as veritable hotspots for accidents, underscoring the urgent need for targeted interventions in these critical zones.

Venturing further into the labyrinth of road safety, Williams et al. have cast a revealing spotlight on the pivotal role played by driver age and experience in the intricate tapestry of accident causation, shedding light on a crucial yet often overlooked aspect of road safety dynamics. In a parallel vein of inquiry, Suganya and Vijayarani have conducted a comparative analysis of various classification algorithms, unveiling the superior efficacy of K-Nearest Neighbors (KNN) in predictive accuracy. Similarly, Sarkar et al.'s exhaustive study has unearthed compelling evidence corroborating the prevalence of accidents on highways, echoing the findings of previous research endeavors. In a bid to harness the power of cutting-edge technology, Stewart et al. have pioneered the development of a neural network model tailored for accident prediction, heralding a new era of efficiency and efficacy in preemptive accident mitigation strategies. Delving deeper into the intricacies of accident dynamics, Zheng et al. have meticulously scrutinized the spectrum of injuries stemming from motor vehicle accidents, while also delving into the nuanced realm of driver emotions as potential causal factors, unveiling a rich tapestry of interconnected variables influencing accident outcomes. On a broader analytical front, Arun Prasath N and Muthusamy Punithavalli have meticulously surveyed a panoply of road accident detection techniques, meticulously dissecting their respective merits and demerits, thus laying the groundwork for future advancements in this crucial domain.

Meanwhile, the comprehensive perspective offered by George Yannis et al. has transcended national boundaries, offering a panoramic vista of international practices in the development of accident prediction models, informed by meticulous data collection and incisive analysis. In a complementary vein, Anand has pioneered a groundbreaking methodology for evaluating the impact of diverse variables on atmospheric deterioration, presenting a compelling case for a similar approach to understanding the intricate interplay of factors shaping road accidents.

The pivotal importance of historical data in identifying accident hotspots has been underscored by Tiwari et al., echoing the sentiments echoed by Singh et al., whose conceptual framework for detecting accident-prone areas serves as a cornerstone for targeted interventions aimed at mitigating risks and enhancing road safety. Rounding off this comprehensive panorama, Kaur et al.'s incisive analysis, informed by correlation analysis and advanced visualization techniques, has culminated in the development of a robust accident prediction model tailored specifically for state highways and district roads, thus furnishing policymakers and stakeholders with invaluable insights into the intricate dynamics of road safety.

Collectively, these groundbreaking research endeavors converge on a singular mission: to furnish stakeholders with actionable insights and empirical evidence, thus empowering them to embark on targeted interventions aimed at curbing the scourge of road accidents, safeguarding lives, and ensuring the safe traversal of our nation's thoroughfare.

II. PROPOSED METHODOLOGY

In this paper, we present an innovative application designed to forecast the likelihood of road accidents based on available accident data. The first step involves data pre-processing, where the road accident data undergoes thorough cleaning to eliminate null and erroneous values, followed by normalization to ensure consistency. Subsequently, relevant features are selected through a meticulous feature selection process, thereby forming a refined dataset for analysis.

The refined dataset is then subjected to various data mining techniques, starting with clustering to identify distinct patterns within the data. These clusters are further analyzed using advanced algorithms such as Support Vector Machines (SVM) and Apriori. SVM is employed to predict the potential risk of accidents, particularly in scenarios where the underlying data distribution is unknown. On the other hand, Apriori is utilized for rule mining, extracting frequent item sets that indicate common combinations of factors contributing to accidents across different road types and weather conditions.

By establishing rules based on these frequent item sets, we gain valuable insights into the interplay of various factors in accident causation. For instance, our analysis might reveal a heightened risk of accidents due to over speeding, even in fine weather conditions at a junction, potentially resulting in severe consequences as demonstrated by the training dataset. Additionally, SVM classification aids in categorizing each accident event into high or low-risk categories, facilitating a comprehensive understanding of accident severity. Leveraging various data mining techniques and exploratory visualization methods, we extract meaningful interpretations from the accident dataset. The architectural framework of our model, as depicted in Figure 1, provides a visual representation of the underlying methodology employed in this study. By estimating the contribution of different factors to road accidents, our analyses pave the way for devising effective solutions aimed at reducing accident rates and minimizing fatalities. In essence, our research endeavors to harness the power of data-driven insights to address the pressing challenge of road safety, ultimately striving towards the overarching goal of creating safer roadways and mitigating the human toll of accidents.

IV. Implementation

A. Dataset:

The dataset utilized in this study was sourced from the Open Government Data (OGD) Platform, India, focusing on accidents within the Bangalore region spanning the years 2014 to 2017. This comprehensive dataset encompasses crucial details including accident date, time, and location, as well as specifics regarding the nature of the accident such as head-on collisions, over-speeding, skidding, or other causes. Additionally, it includes road-related factors such as road type (straight or curved), number of lanes, presence of junctions, and the number of fatalities. These multifaceted factors collectively form the basis for modeling in this study, necessitating the adoption of a stochastic model rather than a simple deterministic model to yield meaningful insights. Thus, the integration of machine learning algorithms alongside data mining techniques becomes imperative to achieve desired outcomes.

B. System Architecture: The raw road accident data undergoes pre-processing to generate a structured dataset serving as input for the model. Subsequently, the model undergoes training using the provided training data, enabling it to predict potential accident risks in user-defined areas. The model further provides users with graphical representations based on derived statistics. The operational framework of this model comprises four distinct modules: Rule Mining, Risk Prediction, Graph Plot, and New Data Entry. Rule Mining leverages the Apriori Algorithm to generate frequent item sets from the input dataset. Risk Prediction employs the Support Vector Machine (SVM) Algorithm for classification, assessing accident risks in specific areas. The Graph Plot module generates informative bar charts illustrating weather conditions, previous accident patterns, and causal factors contributing to accidents. Finally, the New Data Entry module facilitates the reporting of new accident cases.

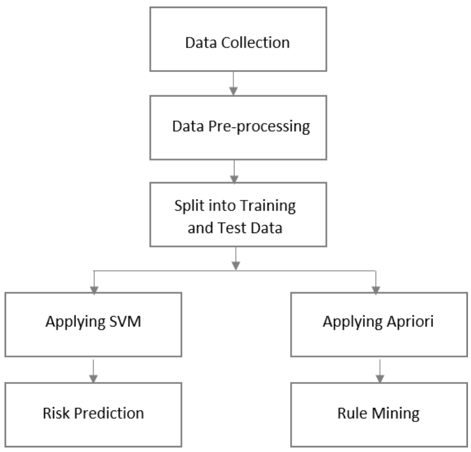

Fig.1. Architecture of the implemented model

C. Software and Languages Used:

The application is developed using the Python programming language within the Anaconda Spyder software environment. Leveraging the robust capabilities of Python and the user-friendly interface of Anaconda Spyder, the implementation process is streamlined and efficient.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

D. Simulation:

R tools are utilized for simulation purposes, enabling the application of various data mining techniques and exploratory visualization methods to the accident dataset. These tools empower researchers to extract interpreted results through graphical representations, charts, statistical analyses, and other insightful outputs. The interactive user interface developed using R tools facilitates comprehensive analysis of diverse factors contributing to accidents, thus aiding in informed decision-making and proactive measures aimed at mitigating accident risks.

1) User Interface

In the final User Interface of the application based on this model, we have four buttons, each corresponding to a particular module in the model. They have been named as.
Rules: This button will generate the frequent item sets based on the support and confidence values using Apriori Algorithm.
Plot Graph: This generates four graphs for the given area. The first graph is based on the attributes such as over speed, skidding, etc., which have been the causes of accidents. The second is based on the weather conditions such as fine, cloudy, rainy, etc., during the time of reported accidents. The third graph is generated based on the number of accident cases reported to nearby hospitals, and the fourth graph predicts the number of accidents due to heavy vehicles such as trucks or lorries over the years in a particular area.
Risk Prediction: For a particular area submitted, this button predicts the possibility of accidents in the area as either HIGH or LOW.
New Data Entry: New accidents can be reported through this option. Information about the accident such as date, time, location, type of accident, etc., can be entered here which will be used for the collection of datasets in the future. (Fig. 10).

A. Screenshots of Application

To illustrate the working of the application based on the prediction model, the following screenshots from the application have been included below:

Fig.2: Application Interface

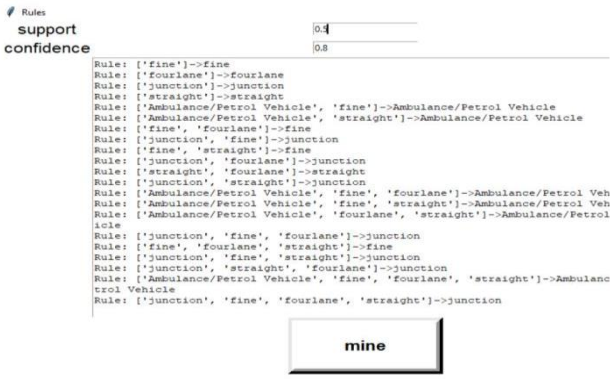

Fig.3: Rule Mining – View

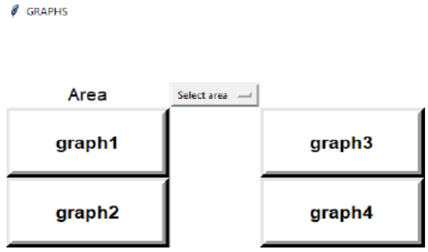

Fig.4: Graphical plot of risk related to accident – View

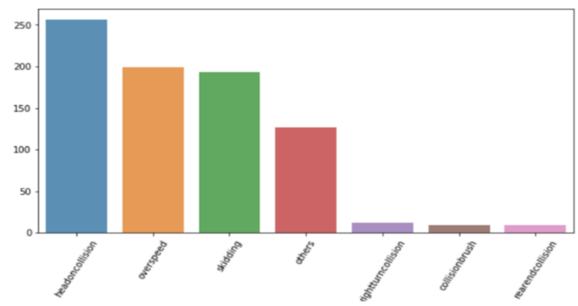

Fig.5: Plot graphs – Graph 1

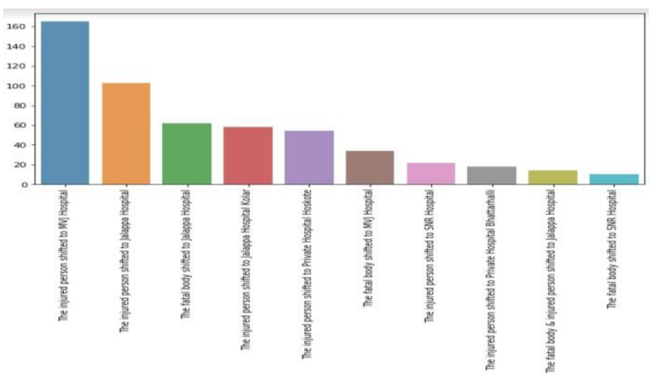

Fig.6: Plot graphs – Graph 2

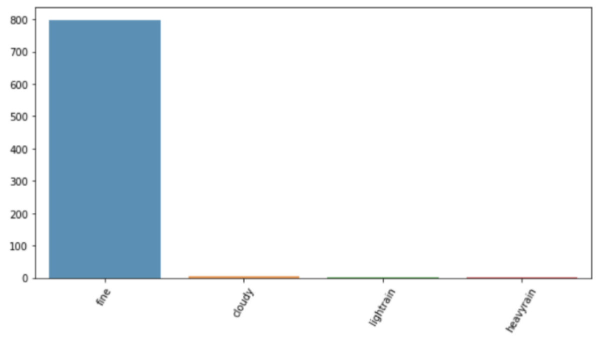

Fig.7: Plot graphs – Graph 3

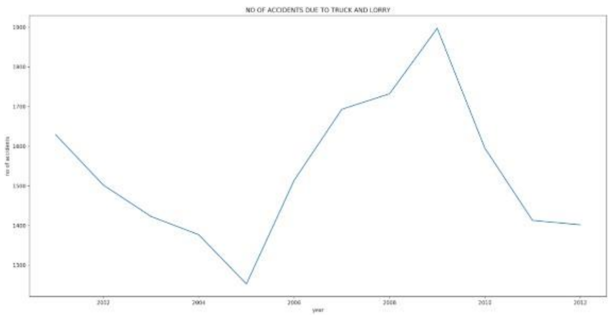

Fig. 8: Plot graphs – Graph 4

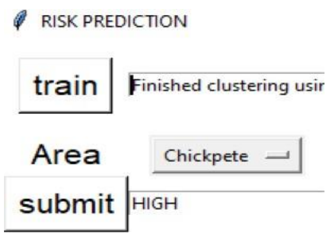

Fig.9: Risk Prediction for the chosen area – View

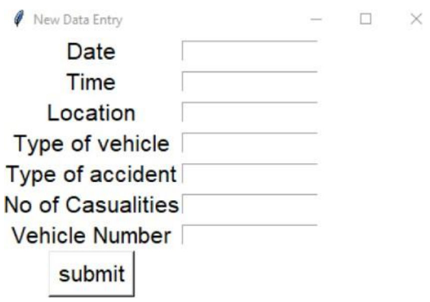

Fig.10: New data entry – View

Discussions

In this paper, we have developed and implemented a road accident prediction model, focusing on various causative factors such as the condition of the road, weather influences, and the nature of accident causes. While previous literature has often overlooked the emotional state and experiential influence of drivers, our model emphasizes these critical factors in predicting accident likelihood. Figures 5, 6, 7, and 8 provide insights into the parameters used in our study and the creation of the prediction model. Figure 5 offers a comparative analysis of different types of accidents reported, including head-on collisions, over-speeding, and skidding. Figure 6 illustrates the observed weather conditions during reported accidents, while Figure 7 indicates the response activities initiated after each accident. Figure 8 presents a graphical representation of the increasing number of accidents involving heavy-duty vehicles, an alarming trend that has been integrated into our dataset.

Our model serves as the foundation for an intuitive application designed to visualize and predict accident risk in user-specified areas. The user interface offers graphical representations of factors contributing to accidents in the past, enabling users to make informed decisions about potential accident risks in their chosen areas. By leveraging data-driven analysis and predictive modeling, our application empowers stakeholders to proactively address road safety concerns and minimize the occurrence of accidents.

Overall, our model provides valuable insights into fatal accident scenarios by analyzing the combinations of factors contributing to accidents. Furthermore, the option to input details of new accident cases ensures continuous improvement of the dataset for future use, reinforcing the utility and efficacy of our predictive model and application.

CONCLUSIONS

Accidents have a significant impact on people's lives, and it's our responsibility to reduce their occurrence. Safe driving practices, along with improvements in road design and vehicle safety, are crucial. This project successfully developed an application using data mining and machine learning to predict road accidents in Bangalore based on factors like vehicle types, driver age, weather, and road conditions. The model accurately predicts accident risk and can be optimized further to include additional constraints. Future developments could include integrating the model into mobile apps for route planning, alerting drivers to potential risks, and aiding emergency services. This technology could also inform better road safety measures and surveillance of accident-prone areas.

ACKNOWLEDGEMENT

I would like to express my gratitude to all those who have contributed to the completion of this project. Firstly, I am thankful to my supervisor for their guidance, support, and valuable insights throughout the research process.

I would also like to extend my appreciation to the road development authorities for providing access to relevant datasets and for their cooperation in this study. Additionally, I am grateful to the automobile industries for their input and collaboration.

I am thankful to my colleagues and friends for their encouragement and assistance during the development and implementation phases of the project.

Finally, I would like to acknowledge the support of my family for their unwavering encouragement and understanding throughout this endeavor. Their patience and encouragement have been instrumental in the successful completion of this project.

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): 1189-1203. 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 expectationmaximization 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 V001T06A009-V001T06A009. 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.