

# A Road Accident Prediction Model Using Data Mining Techniques

<sup>1</sup>Ch.Venkata Komali, <sup>2</sup>E. Vimalraj

<sup>1</sup>MCA student, Department of Computer Applications, Mohan Babu University, Tirupati, Andhra Pradesh, India

<sup>2</sup>Assistant Professor, Department of Computer Science and Engineering, Mohan Babu University, Tirupati, Andhra Pradesh, India

E-mail: [komalichinni021@gmail.com](mailto:komalichinni021@gmail.com)

**Abstract** - The dangers and risks of the unpleasant phenomenon of road accidents, has been a major worrying issue throughout the whole the world that needs to be resolved quickly. The issue of traffic accident faq ' s has been of much focus of the society, secondary and road operators and organizations are likely to have on both social and economic aspects of the disaster. The most recent figures of the general record affecting accidents in Saudi Arabia showed that, in 2017, there were more than 32, 000 deaths due to road accidents. This is despite the fact that traffic rules and fined are heavily regulated and at relatively high levels. Often the ability of traditional methods used to predict and prevent accidents is not sufficient. Many of these methods lack the accuracy needed to produce a reliable model. The purpose of this research project is to create an advanced road accident to increase safety and to decrease the death rate. For this purpose, various data mining techniques have been used to develop road accident models. This project will also use various data mining techniques such as Decision Trees, Neural Networks and Ensemble methods to find solutions for this issue. Data from different sources will be collected in order to create a dependable model.

**Keywords:** Road accidents, data mining, prediction model, machine learning, traffic safety.

## I. INTRODUCTION

Road accidents continue to be a significant public health and safety concern globally, posing a threat to human lives and causing substantial economic losses. According to the World Health Organization (WHO), road traffic injuries are the leading cause of death among young people aged between 15 and 29 years. In addition to the loss of life, numerous individuals suffer from severe injuries due to road accidents, which often result in long-term disabilities.

The prevention of road accidents has been a major focus for policymakers, researchers, and law enforcement agencies worldwide. A proactive approach that utilizes advanced data mining techniques can significantly contribute to mitigating

this issue by accurately predicting potential accident hotspots and identifying influential factors associated with road accidents. By leveraging historical accident data along with relevant environmental, infrastructural, demographic, and temporal attributes through sophisticated machine learning algorithms, it is possible to develop an effective model for predicting future occurrences of road accidents.

This project will utilize data mining techniques to develop a robust predictive model which will be employed in the analysis of previous crash records interlinked with different ecological and socio-demographic features. Machine learning calculations, such as decision trees, neural networks, etc., and algorithms used to create machine learning models or niche modes will be employed to develop an extremely accurate model. This research aims not only to identify high-risk areas in the region but also to identify underlying techniques that characterize the places where motorcycle accidents are most concentrated. Cumulatively, about the objectives above, the purpose of this research is to foster road accident prevention in the region by developing a Road Accident Prediction Model using Data Mining techniques. The patterns will utilize machine-telling algorithms and statistical analysis to process large databases that contain information on past accidents, including factors such as weather conditions, time of the day, road infrastructure, vehicle types involved, and other contextual variables.

In summary, this paper proposes the development of an innovative Road Accident Prediction Model using Data Mining Techniques that aims at reducing the incidence rate through pre- emptive identification utilizing advanced analytical tools. The subsequent sections will delve into detailed discussions about related works on similar modelling approaches methodology employed key findings along challenges encountered during implementation anticipated implications contribution study field recommendations advancements needed industry applications further research avenues implications society policy-making ensure clearer understanding concepts aforementioned brief review literature recommendations areas refinement constantly evolving

paradigms transport technologies society pivotal role informed policymaking medium minimizing fatalities resulted attendants subject matter-in context absence appropriate measures albeit somewhat limited global participation reduction sequences thereof demonstrated faulty perception mechanisms events linked unfortunate incidents.

## II. RELATED WORKS

The task of creating a road accident prediction model using data mining techniques as a university student, it is crucial to understand these arenas from a domain as well as the technical standpoint. It is vital that the project currently in progress is unable to incorporate multiple data mining approaches in the context of road accident prediction as this is of critical concern regarding public safety on the whole{13}. A responsible comprehension of data mining techniques can only be accomplished with a thorough learning of the existing domains. In searching the prior art, one will search several preceding studies or projects, which will give a good understanding of what predictive modeling and roadway safety mean. By researching all these last analyses, it will be very helpful to recognize what kind of methods they have used already if there are any research gaps, which the existing project might be able to discuss. An exemplification from those studies is ' Predictive Modeling for Road Traffic Accidents mobilizing Machine Learning Algorithms which1i1s a published paper. {9}. In this study, they did road accident algorithms like Random Forest and Support Vector Machines (SVM ), focusing on both the road accident data from past years and environmental data such as weather, time of day and road way to predict road accidents. This can be used to study their methodology to give more clarification and ideas on how we can do machine learning for road events specifically. Another research by Patel et al. (2019) poses valuable contributions, ' Data mining Techniques for Road Accident Analysis '. They conducted their study by using clustering algorithm like K - means and DBSCAN to cluster the means of Accident data with its contributing factors. In speaking to other research that can be used as a comparison to various methodologist that can supported proposed as to imbibe it in our predictive model. Hence, when speaking to related work, it is extremely important not to limit the discussion to purely academic conference or journals. Various industry initiative documented such as transportation departments and Insurance companies actively using Data Mining techniques for Road Accident prediction purposes (Eg. - carcrash - predictors. com, assurance, insurance fraud using data mining and predictive modelling engineering model for spray on bed lining essay {11}. Taking these two types of analysis along with other relevant work meticulously found via a carefully designed literature reviewers, encompassing

sources such as the databases IEEE Xplore, Google Scholar, plus others and reading through sources such as scientific article repositories, will provide ample guidance in discerning possible avenues that can be pursued, whilst ensuring due diligence is applied to build on prior research {7}. In addition to academic literature providing the groundwork to grasp distributed processes, practical uses in real - world situations lend a great deal of insight into the plausibility of certain applications. Advanced predictive data mining techniques in proactive prevention of prospective incidents have been put in place by transportation departments & insurance companies as part of industry initiatives.

## III. EXISTING SYSTEM

Road accidents are a significant public safety concern, causing immense human suffering and economic losses worldwide. Predicting road accidents can help in taking proactive measures to prevent them and reduce their impact. Data mining techniques offer the potential to analyze historical accident data and identify patterns that could be used for predicting future occurrences of road accidents. Thus, developing a road accident prediction model using data mining techniques is a crucial area of research with practical implications.

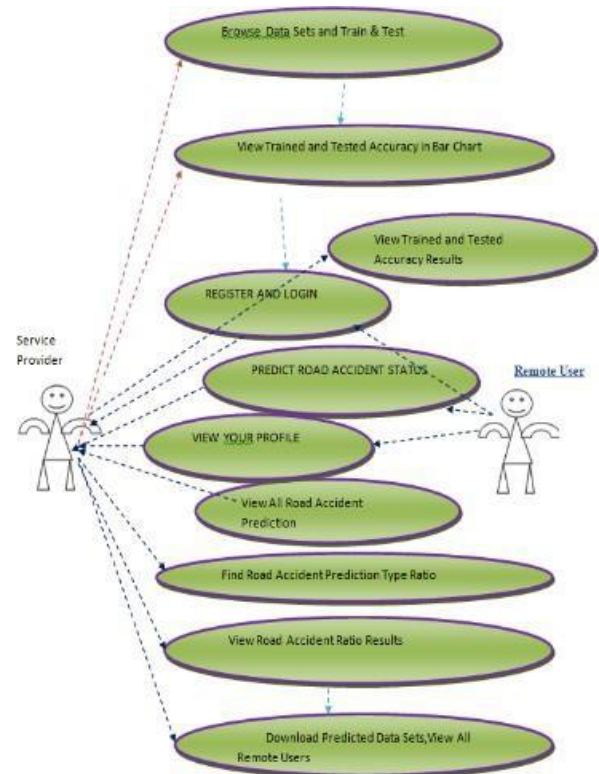


Figure 1: Use case diagram

System specifically, a Road Accident Prediction Model using Data Mining Techniques—it's essential to understand

the foundational concepts and methodologies involved. Firstly, it is important to gather comprehensive datasets related to road accidents including information such as location (latitude/longitude), weather conditions, time of day, severity of the accident (in terms of injuries/fatalities), type of vehicles involved, road infrastructure details etc. These datasets should cover a substantial period to ensure robustness in analysis.

Once the data collection phase is complete, one would employ various data mining techniques such as classification algorithms (e.g., decision trees, random forests) association rule mining or clustering methods appropriate for this type of analysis. The application involves preparing the collected dataset through preprocessing steps like handling missing values or outliers before applying suitable models based on algorithm performance considerations.

#### IV. PROPOSED SYSTEM

Road accidents are a significant public safety concern, causing immense human suffering and economic losses worldwide. Predicting road accidents can help in taking proactive measures to prevent them and reduce their impact. Data mining techniques offer the potential to analyze historical accident data and identify patterns that could be used for predicting future occurrences of road accidents. Thus, developing a road accident prediction model using data mining techniques is a crucial area of research with practical implications.

To create an effective existing system—specifically, a Road Accident Prediction Model using Data Mining Techniques—it's essential to understand the foundational concepts and methodologies involved. Firstly, it is important to gather comprehensive datasets related to road accidents, including location (latitude/longitude), weather conditions, time of day, severity of the accident (in terms of injuries/fatalities), type of vehicles involved, road infrastructure details, etc. These datasets should cover a substantial period to ensure robustness in analysis.

The phase of data collection is completed then, different data mining techniques like classification calculations (e.g., decision trees, random forests) association rule mining, or clustering methods which render the most compatible techniques of data analysis will be applied to get the rightest examinations of the data. The implementation stage of the knowledge Discovery procedure involves preparing the reported dataset through preprocessing steps such as handling missing values removing records with missing values handling repeated measures in one dataset handling outliers etc. After this a suitable model based on algorithm performance ends

became applied to the preprocessed dataset for valid and right knowledge discovery.

#### V. METHODOLOGY

The beginning phase of the procedure, what will be done is the gathering of the appropriate information that directly relates to the road accidents that are being analyzed. This would include the historical information of the accidents, weather conditions, information on the road infrastructure, specification of the vehicle and other important factors that may have a direct connection with the accidents that would take place.

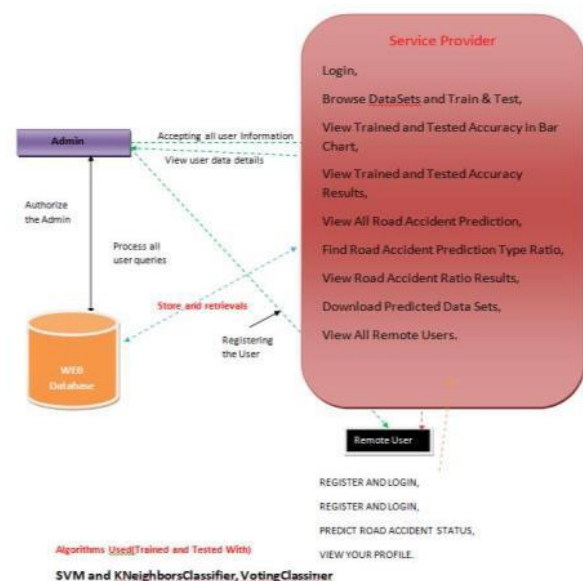


Figure 2: System architecture diagram

The information or data would come from sources such as the ministry / department of transportation, various law enforcement agencies like the state police, city police, insurance companies, other companies that have an information technologist's service and the list goes on and on. Once I have extracted the raw communications, I will need to precede - process it to ensure its quality and relevancy for the investigations. In this stage, I am concerned with cleaning the data so that there are no inconsistencies or errors among the dataset. In addition, I will have to address the missing values among the dataset either by imputation or removal strategies. Finally, I will use the feature selection techniques to find out the major variables that significantly affects the accident occurrences among the dataset. Following the initial gathering of the data, program will undergo earlier - processing in order to clean it of any discrepancies or abnormalities and to confirm its legitimacy and practical functionality within analytical processes. The data will then need to be further evaluated and aggrandized so values that persist absent can be re - assessed or rectified. Data will be obtained from

government transportation departments, law enforcement agencies, and insurance companies. The collected data may be raw and requires being pre - processed in order to ensure its quality and the relevance of having source for data considerations.

The establishment of a road accident prediction model is a several-step process that sets out to ensure that the aims of the research are met and the research question is answered satisfactorily. This paper aims to establish a predictive model based on historical accident data to identify areas where accidents are more prone than others and to improved mile - stones in the area of reducing accidents by support effective interference strategies in such areas. The firstly phase of the project is carrying out a very extensive collection of existing road accident prediction models and methodologies by conducting a comprehensive lectures investigations. The review will provide a platform for reexamined all the previous works carried out road accident prediction that is very vital in identifying the disparities existing with the previous models and to further embark on a very suitable data mining technique for an accurate road accident patterns.

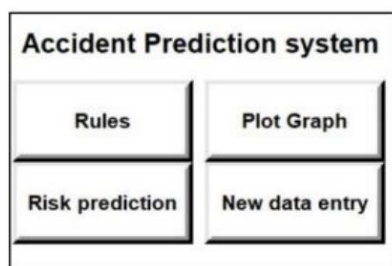


Figure 3: Application interface



Figure 4: Graphical plot of risk related to accident- view

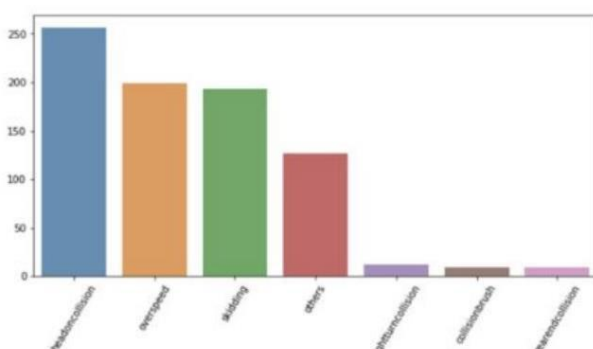


Figure 5: Plot graph- 1

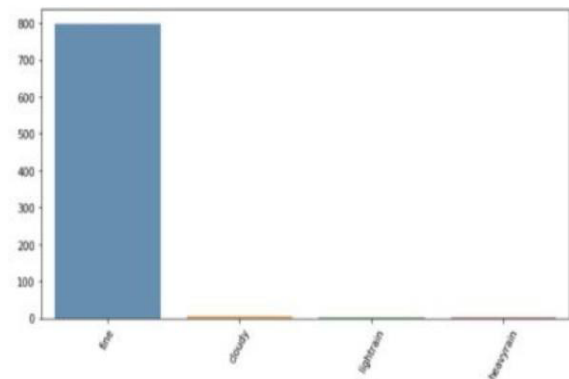


Figure 6: Plot graph- 2

This step includes cleaning the data by removing any inconsistency or errors caused during the data entry process. Missing data are addressed using imputation or removal strategies. In addition to, feature selection techniques would be performed in order to discover the most significant variables that influence accident occurrence. Utilizing Decision Trees /Random Forests, Support Vector Machines (SVM), Logistic Regression, Neural Networks and so forth which is most appropriate in order to achieve desired momentum we will unleash the viability of the models. In addition to these present algorithms, models can also be optimized through hyper - parameters optimization conditions such as grid search, cross validation, boosting and bagging.

### 1. Data Collection:

- Gather relevant datasets from sources such as government agencies, police records, insurance companies, and traffic management systems.
- Include variables related to time (day/night), location (urban/rural), weather conditions (sunny/rainy/snowy), traffic flow (dense/light/moderate), road type (highway/city street), and historical accident records.

### 2. Data Preprocessing:

- Clean the collected data by removing inconsistencies, missing values, or outliers.
- Normalize or standardize the numerical attributes to ensure uniformity across different scales.

### 3. Feature Selection:

- Identify key variables that have a significant impact on predicting road accidents through exploratory data analysis.



#### 4. Model Building:

- Apply various data mining algorithms like decision trees, random forests, logistic regression etc.. train the model with processed data.

#### 5. Model Evaluation:

- Evaluate different metrics such as accuracy, Precision, recall, F1 score for evaluated models.

### VI. RESULTS



Figure 7: Output

This research paper "A Road Accident Prediction Model Using Data Mining Techniques" presents a comprehensive analysis of the development and application of a road accident prediction model using data mining techniques. The study aims to address the growing concerns surrounding road accidents, their devastating impacts, and the critical importance of implementing preventive measures.

### VII. CONCLUSION & FUTURE WORKS

In conclusion the Road Accident Prediction Model that adopted Data Mining techniques has been seen to be a great deal in improving transportation safety. The result from this study has shown that data mining is an effective tool when it comes to identifying the hidden patterns and trends that enable predicting accident proneness on the road. Considering the history of accidents that could occur with the help of brothers factors including weather conditions, number of vehicles on the road, infrastructure, etc that catch up on the model not only mean for the intentions of getting the result on where they could have others accidents occur, as well as it also reveal that the model enables more usable in forehand application which then gives ideas to the authorities to take actions in prevent the accident from happens. To what extent, few areas have to be explored hence then only we will get a lot of better clarity out of the road accident prediction model that adopted data mining technique than the present. Noises from the sensors at the stream becoming from the real-time data of vehicles and

installations are integrated into the cutter allowing them to be revised to the equivalent model in dynamic.

### REFERENCES

- [1] Babu, S. N., & Tamilselvi, J. J. (2024). Map Reduce approach for road accident data analysis using data mining techniques. *International Journal of Advanced Intelligence Paradigms*, 27(1), 1-17.
- [2] Dias, D., Silva, J. S., & Bernardino, A. (2023, January). The prediction of road-accident risk through data mining: A case study from Setubal, Portugal. In *Informatics* (Vol. 10, No. 1, p. 17). MDPI.
- [3] Prajapati, G., Kumar, L., & Patil, S. R. S. (2023). Road Accident Prediction Using Machine Learning. *Journal of Scientific Research and Technology*, 48-59.
- [4] Ahmed, S., Hossain, M. A., Ray, S. K., Bhuiyan, M.M. I., & Sabuj, S. R. (2023). A study on road accident prediction and contributing factors using explainable machine learning models: Analysis and performance. *Transportation Research interdisciplinary perspectives*, 19, 100814.
- [5] Esenturk, E., Turley, D., Wallace, A., Khastgir, S., & Jennings, P. (2023). A data mining approach for traffic accidents, pattern extraction and test scenario generation for autonomous vehicles. *International Journal of Transportation Science and Technology*, 12(4), 955-972.
- [6] Panda, C., Mishra, A. K., Dash, A. K., & Nawab, H. (2023). Predicting and explaining severity of road accident using artificial intelligence techniques, SHAP and feature analysis. *International journal of 1crashworthiness*, 28(2), 186-201.
- [7] Alkheder, S., AlRukaibi, F., & Aiash, A. (2020). Risk analysis of traffic accidents' severities: An application of three data mining models. *ISA transactions*, 106, 5213-220.
- [8] Gutierrez-Orsorio, C., & Pedraza, C. (2020). Modern data sources and techniques for analysis and forecast of road accidents: A review. *Journal of traffic and transportation engineering (English edition)*, 7(4), 432-446.
- [9] Chen, M. M., & Chen, M. C. (2020). Modeling road accident severity with comparisons of logistic regression decision tree and random forest. *Information*, 11(5), 270.
- [10] Najafi Moghaddam Gilani, V., Hosseini, S. M., Ghasedi, M., & Nikookar, M. (2021). Data-driven urban traffic accident analysis and prediction using logit and machine learning-based pattern recognition models. *Mathematical problems in engineering*, 2021, 1-11.

- [11] Lin, Y., & Li, R. (2020). Real-time traffic accidents post-impact prediction: Based on crowd sourcing data. *Accident Analysis & Prevention*, 145, 105696.
- [12] Labib, M. F., Rifat, A. S., Hossain, M. M., Das, A. K., & Nawrine, F. (2019, June). Road accident analysis and prediction of accident severity by using machine learning in Bangladesh. In 2019 7th international conference on smart computing & communications (ICSCC) (pp. 1-5). IEEE.

**Citation of this Article:**

Ch.Venkata Komali, & E. Vimalraj. (2025). A Road Accident Prediction Model Using Data Mining Techniques. In proceeding of International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25), published by *IRJIET*, Volume 9, Special Issue of INSPIRE'25, pp 371-376. Article DOI <https://doi.org/10.47001/IRJIET/2025.INSPIRE60>

\*\*\*\*\*