e-ISSN:2582-7219

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 12, December 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

© 2024 IJMRSET | Volume 7, Issue 12, December 2024| DOI: 10.15680/IJMRSET.2024.0712055

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# Road Accident Prediction Using Machine Learning

**Mr. M.Ajay Kumar[1], Mr. K.Vasu[2], Mr. K.Karunakar[3], Mr. K.Uday Kiran[4]**

Assistant Professor, Department of Computer Science & Engineering, Guru Nanak Institute of Technology, India[1]

Student, Department of Computer Science & Engineering, Guru Nanak Institute of Technology, India[2,3,4]

**ABSTRACT:** Road accidents are a major public safety concern globally, leading to significant loss of life, injuries, and economic costs. This project aims to develop a predictive model for road accident occurrence using data mining techniques, providing a tool for proactive accident prevention and resource allocation. The proposed system utilizes the Naive Bayes classifier to predict the likelihood of road accidents, based on various contributing factors such as weather conditions, time of day, road type, and traffic density. In the existing system, Logistic Regression is employed to model the relationship between these factors and accident probability. While Logistic Regression provides a strong baseline due to its interpretability and effectiveness in binary classification problems, it may not fully capture the probabilistic dependencies among the predictors. The Naive Bayes classifier, by assuming conditional independence among predictors, offers a simpler yet powerful alternative for accident prediction. It is computationally efficient and well-suited for handling large datasets, making it an attractive choice for real-time prediction scenarios. This study compares the performance of both models using historical accident data, evaluating metrics such as accuracy, precision, recall, and F1-score. Preliminary results indicate that the Naive Bayes classifier achieves comparable, if not superior, performance to Logistic Regression, highlighting its potential for practical implementation in traffic management systems. The findings underscore the importance of advanced data mining techniques in enhancing road safety measures and inform the deployment of predictive analytics in transportation infrastructure planning.

## I. INTRODUCTION

Road safety is a critical concern in today's fast-paced world, with road accidents causing significant loss of life and economic damage globally. Predicting road accidents is essential to minimize these risks and enhance public safety. The Road Accident Prediction using Machine Learning project aims to address this issue by leveraging advanced algorithms and historical data to predict accident probabilities based on various influencing factors, such as weather conditions, road types, traffic violations, and human behaviour. The project integrates data preprocessing, feature selection, and machine learning techniques to develop an effective predictive model. By using the Naive Bayes Classifier, it achieves high accuracy in determining accident-prone scenarios. The predictive insights generated by the model can be used by authorities to implement preventive measures, optimize resource allocation, and enhance traffic management systems. This project represents a step forward in using technology to promote safer and smarter transportation systems.

## II. LITERATURE SURVEY

***Williams t et. Al(2016)***The National Transportation Safety Board in the United States and the Transportation Safety Board of Canada publish reports about major railroad accidents. The text from these accident reports were analyzed using the text mining techniques of probabilistic topic modeling and k-means clustering to identify the recurring themes in major railroad accidents. The output from These analyses indicate that railroad accidents can be successfully grouped into different topics. The output also suggests that recurring accident types are track defects, wheel defects, grade crossing accidents and switching accidents. A major difference between the Canadian and U.S. reports is the finding that accidents related to bridges are found to be more prominent in the Canadian reports.

***Suganya, E. and S. Vijayarani(2017)***Classification is a model finding process which is used for segmenting the data into different classes based on some constraints. This work analyzes the road accidents in India data set using classification algorithms namely linear regression, logistic regression, decision tree, SVM, Naive Bayes, KNN, Random Forest and gradient boosting algorithm. The performance measures used are accuracy, error rate and execution time. This analysis is done in R data mining tool. The performance of KNN is better than other algorithms.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

***Sarkar S (2017)***Steel industry is considered to be an economic sector with a higher number of accidents. Workers in this industry are exposed to a wide variety of hazards during working hours. Thus, the database maintained in industry varies in terms of the types of data indicating nature of accidents, causes of accidents, dates and timestamp etc. The objective of this study is to give predictive solutions of accident occurrences in a steel industry based on free-text data or narratives logged in the database from previous incidents report. Utilizing the database comprised of 9488 observations of occupational accidents during the year 2010 to 2013 in an integrated steel manufacturing plant in India, text mining technique has been used, and its outputs have been fed into three classifiers namely Support Vector Machines, Random Forest, and Maximum Entropy which are tested for the evaluation of the best fit model that could predict the injury occurrences in workplace. Furthermore, with the help of the same three classifiers, the Causes of injuries are also predicted. Maximum Entropy and Random Forest classifiers are found out to be the best of all other classifiers in terms of higher area under curve (AUC) value in binary and multi-class prediction models, respectively.

***Mohamed K Nour(2020)***Road safety researchers working on road accident data have witnessed success in road traffic accidents analysis through the application data analytic techniques, though, little progress was made into the prediction of road injury. This paper applies advanced data analytics methods to predict injury severity levels and evaluate their performance. The study uses predictive modeling techniques to identify risk and key factors that contribute to accident severity. The study uses publicly available data from UK department of transport that covers the period from2005 to 2019. The paper presents an approach which is general enough so that it can be applied to different data sets from other countries. The results identified tree-based techniques such as XGBoost outperform regression-based ones, such as ANN. In addition to the paper, it identifies interesting relationships and acknowledged issues related to quality of data

***Annie Racheal Rajkumar et. al(2020)***Injuries due to road accidents are one of the most prevalent causes of death apart from health-related issues. The World Health Organization states that road traffic injuries caused an estimated 1.35 million deaths worldwide in the year 2016. That is, a person is killed every 25seconds. This calls for the need to analyze road accidents and the factors affecting them and come up with a method to reduce the probability of their occurrence. The analysis of road accident severity was done by running an accident dataset through several machine learning classification algorithms to see which model performed the best in classifying the accidents into severity classes such as slight, severe and fatal. It was observed that logistic regression to perform multi label classification gave the highest accuracy score. It was also observed that factors such as the number of vehicles, lighting conditions and road features played a role in determining the severity of the accident.

## III. METHODOLOGY

Modules Name:

- Data collection
- Dataset
- Data Preparation
- Model Selection
- Analyze and Prediction
- Accuracy on test set
- Saving the Trained Model

**Modules Explanation**

**1) Data collection:** Collecting high-quality data is crucial for building a reliable machine learning model. Data is sourced through techniques like web scraping and manual inputs; the provided dataset contains 576 records with 15 columns

**2) Dataset:** The dataset includes features such as states, junction types, vehicle and human details, traffic violations, weather conditions, and accident occurrences (yes/no). It offers comprehensive data for training and testing.

**3) Data Preparation:** The data is cleaned, duplicates are removed, and missing values are addressed. It is randomized, visualized for patterns, and split into training and evaluation sets.

**4) Model Selection:** The Naive Bayes classifier is chosen for its simplicity and efficiency, achieving an accuracy of 94.7% on the test set.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

**5) Analyze and Prediction:** Fourteen features are used for prediction, with the model delivering a test set accuracy of 94.8%.
**6) Saving the Trained Model:** The trained model is saved as a .pkl file using the Python pickle library, ensuring it is ready for deployment.

**Existing System Disadvantages:**

- Logistic regression assumes that there is a linear relationship between the independent variables and the log odds of the dependent variable.
- This method can be sensitive to multicollinearity, which occurs when two or more independent variables are highly correlated.

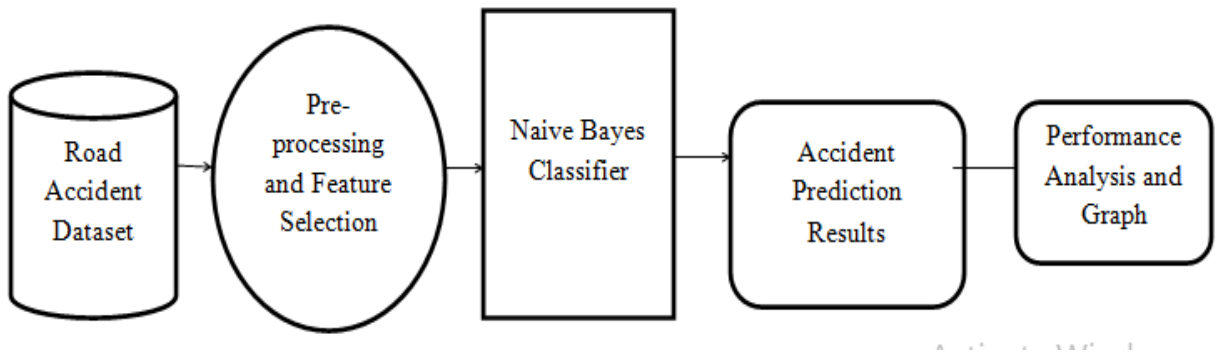
## IV. PROPOSED SYSTEM

Naive Bayes is a family of simple, yet highly effective probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Despite the simplification in its assumptions, Naive Bayes classifiers have demonstrated impressive performance in various complex real-world applications, particularly in text classification, spam detection, and sentiment analysis. Naive Bayes classifiers, the goal is to calculate the posterior probability for each class given the input features and assign the class label with the highest posterior probability.

**PROPOSED SYSTEM ADVANTAGES:**

- Naive Bayes is easy to implement and computationally efficient. It requires a relatively small amount of training data to estimate the parameters necessary for classification.
- Naive Bayes classifiers scale well with the number of predictors and the size of the dataset.

## V. SYSTEM ARCHITECTURE




**Explanation:**
The system architecture diagram represents the workflow of the Road Accident Prediction Model. It begins with the Road Accident Dataset, which contains information such as weather conditions, road types, vehicle details, traffic violations, and accident occurrences. The data undergoes Pre-processing and Feature Selection, where it is cleaned, irrelevant data is removed, and the most important features are selected to enhance the model's accuracy and efficiency. The refined data is then passed into the Naive Bayes Classifier, which is used to train the machine learning model and make predictions on accident occurrences based on the input features. The results of these predictions are presented in the Accident Prediction Results stage, indicating whether an accident is likely to occur under specific conditions. Finally, the model's performance is evaluated in the Performance Analysis andGraph stage, where the accuracy is measured, and visualizations are generated to present the results effectively.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

**Implementation**

**Input:**

- Source: Historical road accident datasets.
- Input Types: Factors such as weather conditions, road type, traffic density, and time of day.

**Processing:**

- Data Preprocessing:
  - Cleanses the dataset by handling missing values and outliers.
  - Normalizes the data for feature scaling and ensures data consistency.
- Feature Selection:
  - Identifies critical factors influencing road accidents using statistical methods.
- Algorithm Integration:
  - Utilizes the Naive Bayes Classifier for predicting the probability of accidents.
  - Incorporates existing Logistic Regression for baseline comparison.

**Prediction Model:**

- Training:
  - Trains the Naive Bayes Classifier on the pre-processed data.
  - Tunes hyperparameters for optimal performance.
- Testing:
  - Evaluates the model using metrics such as accuracy, precision, recall, and F1-score.
  - Compares the performance with the Logistic Regression model.

**Augmentation:**

- Generates risk predictions based on user-defined inputs or real-time factors.
- Visualizes accident probabilities and highlights critical patterns or risk zones.

**Output:**

- Accident Prediction Results:
  - Displays the likelihood of an accident occurring under specific conditions.
- Performance Metrics:
  - Outputs performance graphs (accuracy, precision, recall, F1-score) for evaluation.

**Termination:**

- Ends the process upon completing predictions or at the user's discretion.

## VI. EXPERIMENTAL RESULTS

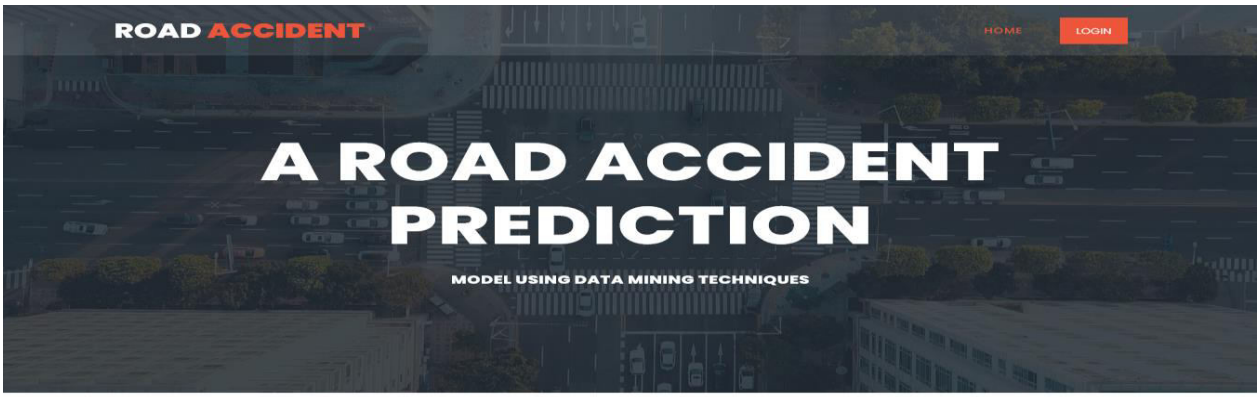


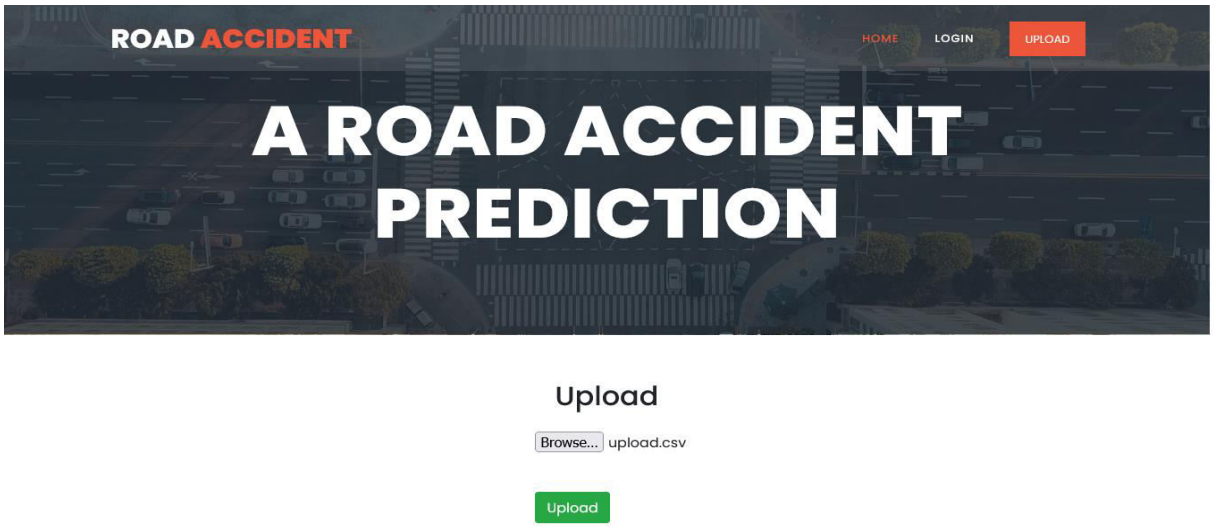

Fig:Users can log in to access the Road Accident Prediction

© 2024 IJMRSET | Volume 7, Issue 12, December 2024| DOI: 10.15680/IJMRSET.2024.0712055

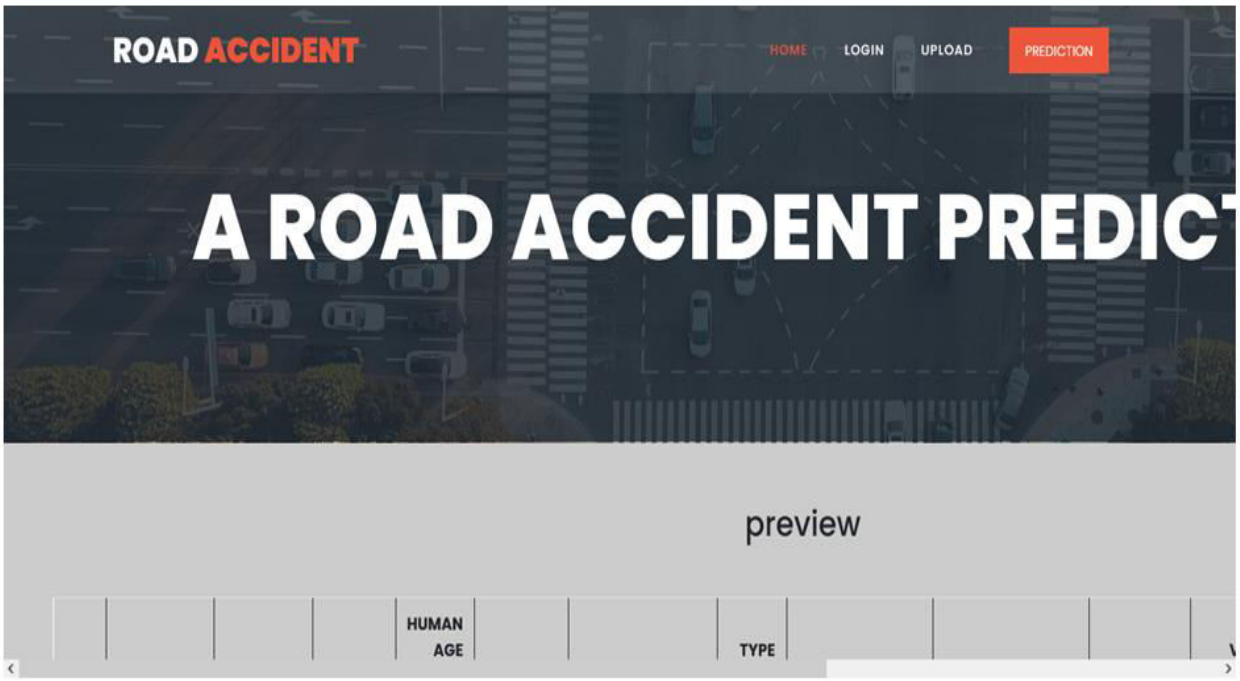
ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)




Users upload datasets or relevant information for analysis.



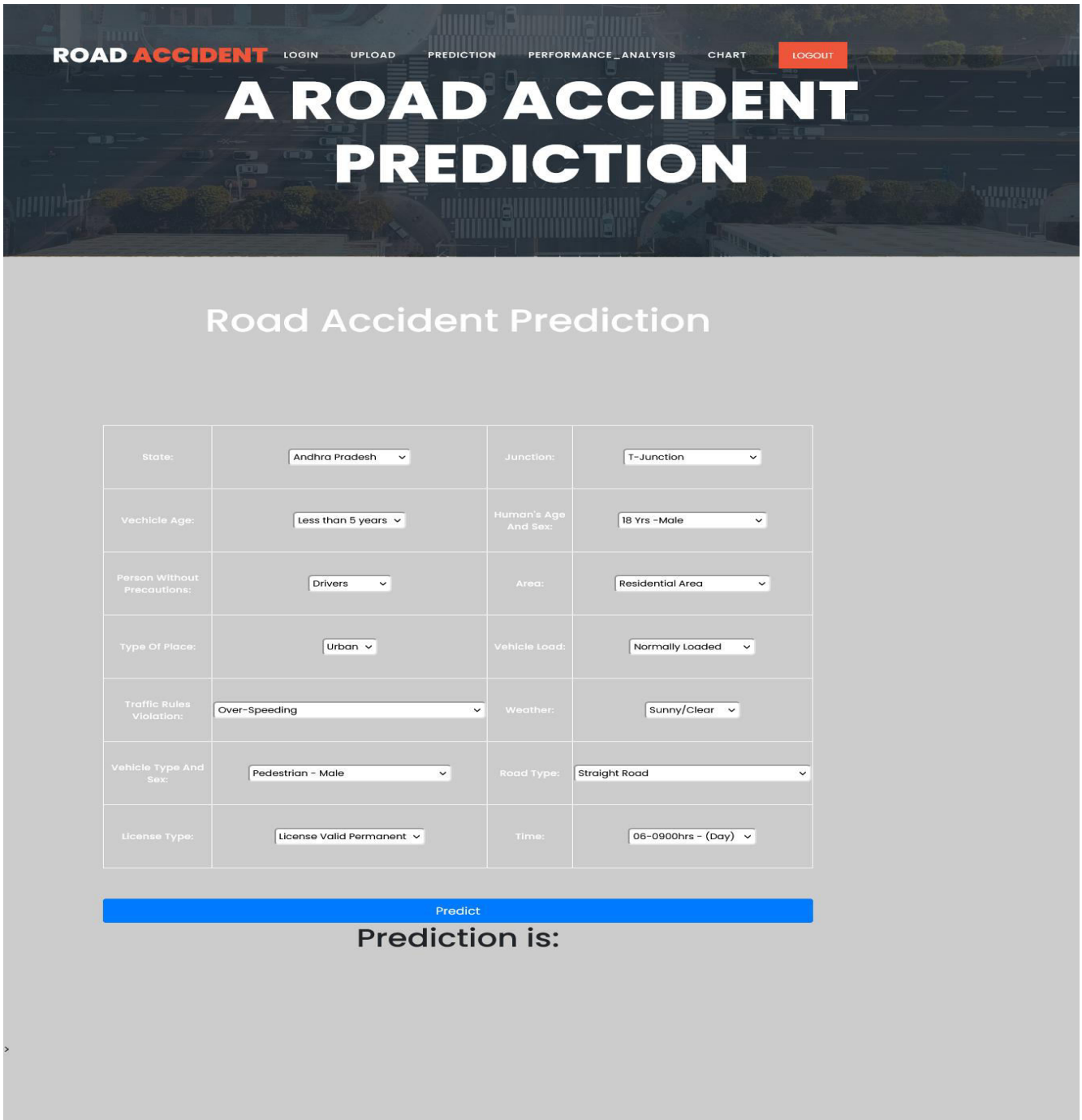

A table preview shows the dataset details for verification.

© 2024 IJMRSET | Volume 7, Issue 12, December 2024| DOI: 10.15680/IJMRSET.2024.0712055

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



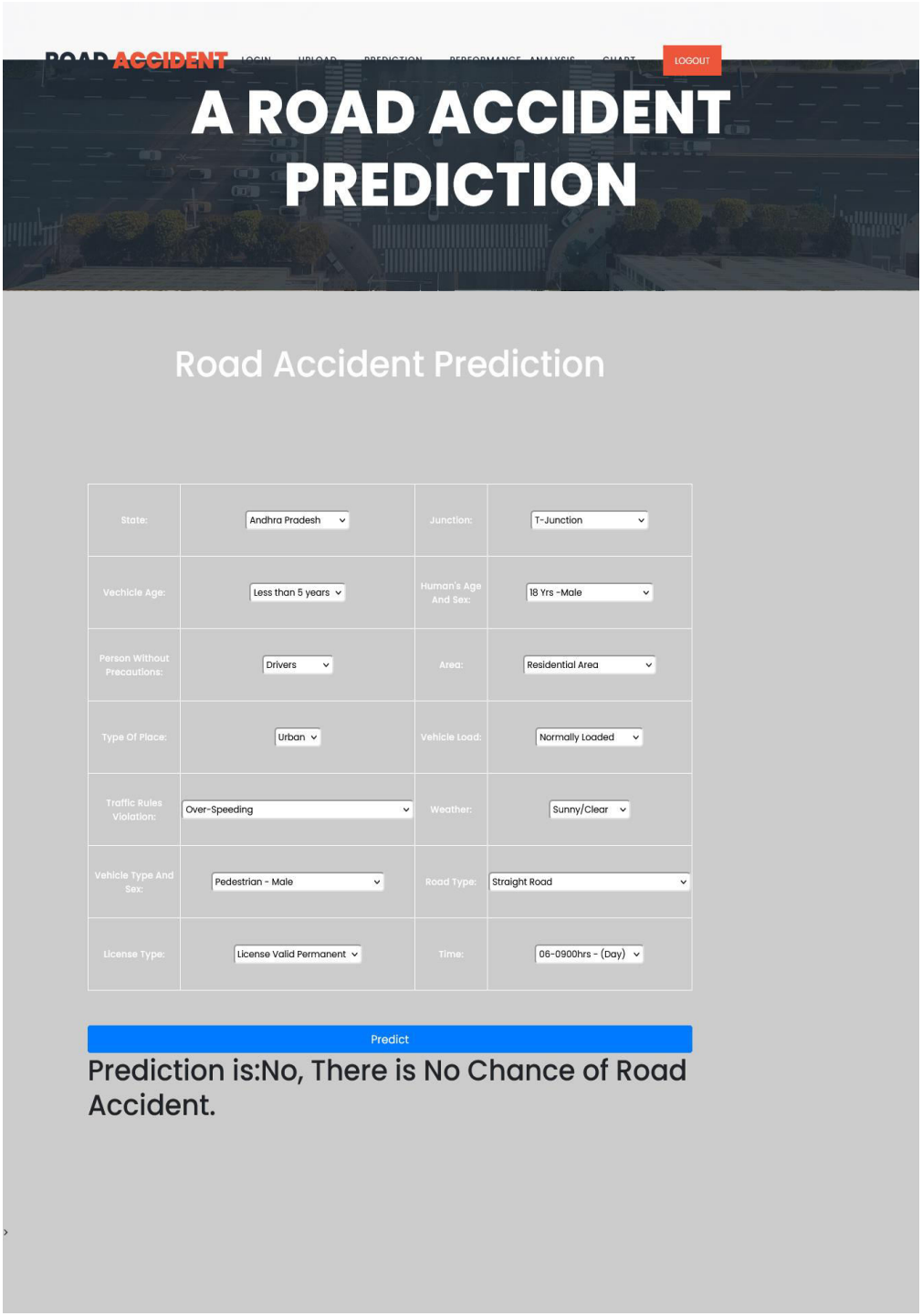

Users input parameters to predict the likelihood of a road accident.

© 2024 IJMRSET | Volume 7, Issue 12, December 2024| DOI: 10.15680/IJMRSET.2024.0712055

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

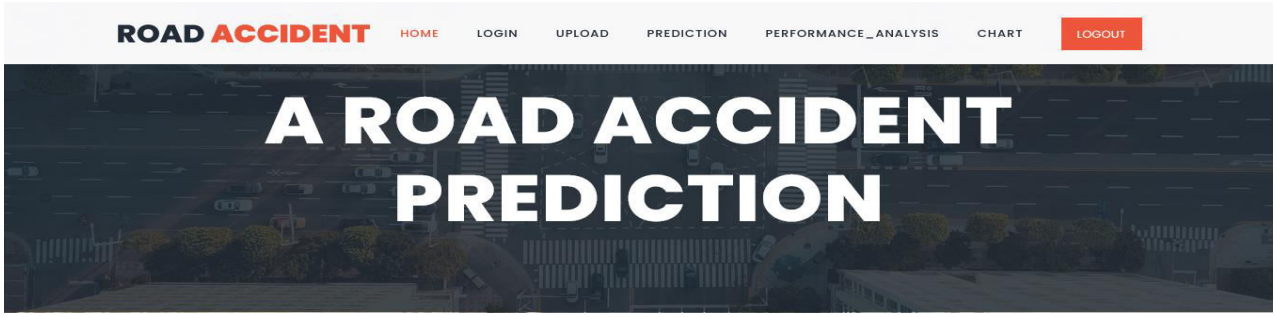



Results indicate whether a road accident is likely to occur.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

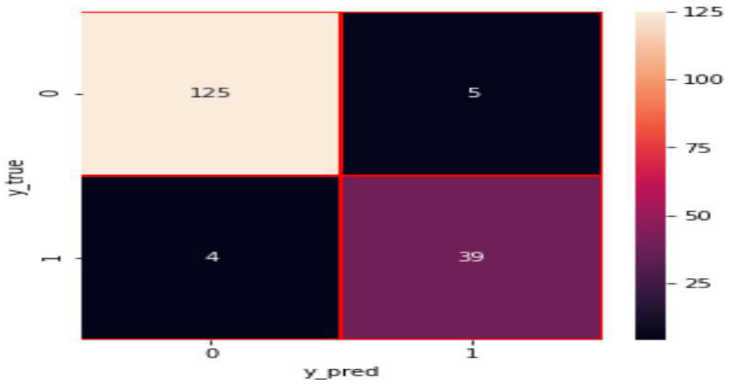



PERFORMANCE ANALYSIS

Precision and recall

| | Recall | Precision |
|---|---|---|
| No Road Accident(0) | 0.98 | 0.99 |
| Road Accident(1) | 0.95 | 0.97 |

Confusion Matrix



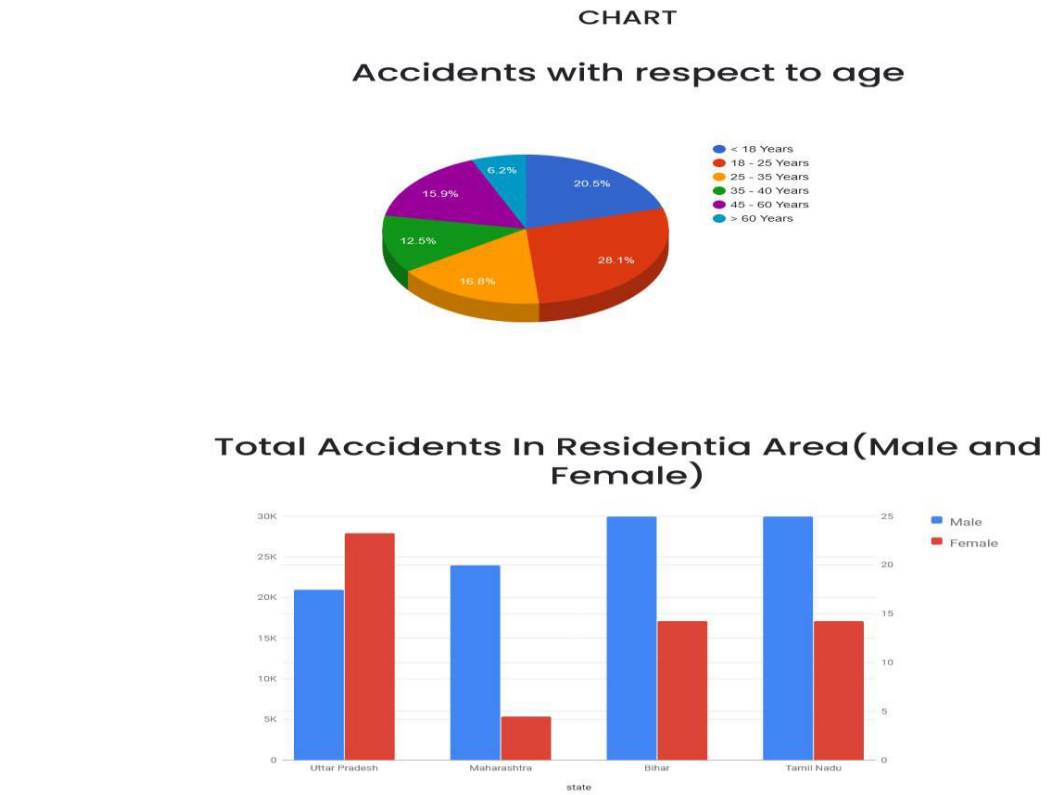

Precision, recall, and confusion matrix are displayed for model performance.

© 2024 IJMRSET | Volume 7, Issue 12, December 2024| DOI: 10.15680/IJMRSET.2024.0712055

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)


Charts and graphs summarize accident trends or demographics.

## VII. CONCLUSION

An accident can change the lives of many people. It is up to each of us to bring down this increasing number. This can be made possible by adopting safe driving measures to an extent. Since all instances of accidents cannot be attributed to the same cause, proper precautionary measures will also need to be exercised by the road development authorities in designing the structure of roads as well as by the automobile industries in creating better fatality reducing vehicle models. One thing within our capability is to predict the possibility of an accident based on previous data and observations that can aid such authorities and industries. This project was successful in creating such an application that can help in efficient prediction of road accidents based on factors such as types of vehicles, age of the driver, age of the vehicle, weather condition and road structure, so on. This model was implemented by making use of several data

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

mining and machine learning algorithms applied over a dataset for India and has been successfully used to predict the risk probability of accidents over different areas with high accuracy.

## VIII. FUTURE ENHANCEMENT

The model can be further optimized in future to include several constraints that have been left out in the current study. These optimized models can be efficiently utilized by the government to reduce road accidents and to implement policies for road safety. Another scope of this work would be to develop a mobile app that will help the drivers in choosing a route for a ride. A call out to the driver through the maps service can also be implemented that would also announce the risk probability in a chosen route along with the directions.

## 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 nhtsavehicle 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 expectation maximization method. J. Inf. Sci. Eng31(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. Neuro computing 275: 2444-2458.http://dx.doi.org/10.1504/IJIT.2018.090859.
[10] ArunPrasath, N and MuthusamyPunithavalli. "A review on road accident detection using data mining techniques." International Journalof Advanced Research in Computer Science 9 (2018): 881-885.
[11] George Yannis, AnastasiosDragomanovits, Alexandra Laiou, Thomas Richter, Stephan Ruhl, Francesca La Torre, Lorenzo Domenichini, Daniel Graham, NioviKarathodorou, 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] PrayagTiwari, Sachin Kumar, Denis Kalitin (2017). "Road-UserSpecific 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. HarpreetKaur. "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.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

## IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com