International Journal of Scientific Research and Engineering Development-– Volume 6 Issue 4, July- Aug 2023

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Road Accidents Analysis Using Machine Learning In Python

Rishabh Maurya(19SCSE1010144)
Student of Computer Science and Engineering
Galgotias University
rishabhmaurya014@gmail.com

Mr. Rashmi Rathi Upadhyay
School of Computer Science and Engineering
Galgotias University
rashmi.upadhyay@galgotiasuniversity.edu.in

Abstract— Road accidents are a major global public safety hazard that result in huge financial and human losses. Traditional methods for studying traffic accidents are frequently time-consuming and have a limited capacity to identify intricate trends. Machine learning approaches have recently demonstrated promising outcomes in a number of industries, including image identification, finance, and healthcare. In this paper, we use machine learning techniques to give a thorough analysis of traffic incidents .The proposed machine learning pipeline includes model selection, evaluation, feature engineering, and data preprocessing. Numerous machine learning methods, including decision trees, random forests, support vector machines, and neural networks, were tested. We also addressed difficulties with class imbalance in the data and used dimensionality reduction techniques .Our results show that in terms of accuracy, precision, and recall, the created machine learning model performs better than conventional statistical techniques. The algorithm is successful in identifying key elements linked to auto accidents and ranks the risk factors for various road segments. By using this data, decision-makers in government, urban planning, and law enforcement can increase road safety and lower accident rates .The goal of this research is to create a predictive model that can precisely assess the reasons causing traffic accidents, pinpoint high-risk areas, and ultimately assist in the creation of efficient preventive measures

Keywords— Road Accidents, Traffic collision, Crash Analysis, Accident Data, Road Safety, risk Factor.

I. INTRODUCTION

Road accidents are a major cause for concern since they result in severe death, injury, and financial losses. Road accident data analysis is essential for seeing trends, comprehending contributing causes, and creating practical road safety solutions. Machine learning has opened up new possibilities for improving the effectiveness and precision of analysis of traffic accidents. Computers can use machine learning algorithms to learn from historical data on traffic accidents, spot underlying trends, and generate predictions based on this understanding. This strategy provides an organized manner and data-driven to address issues with traffic safety, ultimately resulting in more targeted improved and intervention accident prevention techniques. In this perspective, a number of tasks in road accident analysis can be performed using machine learning algorithms: Machine learning algorithms can analyze historical data on traffic accidents to find patterns and risk factors related to specific locations, traffic conditions, or times of the day, enabling authorities to take proactive preventive measures. Machine learning techniques can be used to pinpoint the causes of accidents on the road, including faults with vehicle performance, weather conditions, poor road design, and driver conduct. This knowledge can help policymakers develop specialized safety measures.

Based on accident characteristics, machine learning models can forecast the severity of injuries, assisting hospitals and emergency services in efficiently allocating resources and delivering timely and effective medical care.

In order to help authorities optimize road networks and decrease congestion, machine learning can be used to evaluate traffic data and forecast possible congestion locations or bottlenecks likelihood that traffic congestion will result in accidents.

II. RELATED WORK

As of my most recent knowledge update in September 2021, there were a handful of scholarly articles and projects connected to road accident analysis using machine learning. Here are a few items of note:

A study titled "Analysis and Prediction of Road Accidents Using Machine Learning Techniques" S. Santhi and S. Vijayakumar expect to be This study examined the use of machine learning techniques like Random Forest, Decision

Trees, and Support Vector Machines (SVM) in order to analyze traffic accident data and pinpoint accident-prone locations. The authors' analysis took into account factors like traffic flow, road shape, and accident statistics.

The 2020 paper "A Review on Road Traffic Accident Prediction Using Machine Learning Techniques" by K. O. "Agbele" and R. A. "Jumbor - Baripi" the analysis and prognosis off-road accidents using various machine learning approaches is presented in detail in this work. It includes techniques like Naive Bayes, K-Nearest , and Artificial Neural Networks. The paper outlines the disadvantages and advantages of each strategy.

Road Traffic Accident Analysis and Prediction Using Machine Learning Algorithms (M. Shalini and M. R. Sumalatha, 2018)

Using machine learning methods like Logistic Regression and k-Nearest, this study aimed to analyze data on traffic accidents, identify trends, and predict future accident occurrences.

III. METHODOLOGY

The following steps are commonly included in the process for machine learning-based road accident analysis:

Data gathering: Compile pertinent information about traffic accidents from a variety of sources, including traffic authority, law enforcement agencies, hospitals, and other relevant institutions. The information should cover information on accident sites, weather, road conditions, vehicle kinds, driver conduct, and accident outcomes (such as injury and fatality severity).

Data preprocessing: To manage missing values, outliers, and inconsistencies, clean and preprocess the obtained data. This process could include feature scaling, data normalization, and the encoding of categorical variables into numerical representations.

Choose or extract the features that are most important to understanding how they affect traffic accidents. The most informative qualities can be chosen using feature selection approaches like correlation analysis or feature importance from machine learning models.

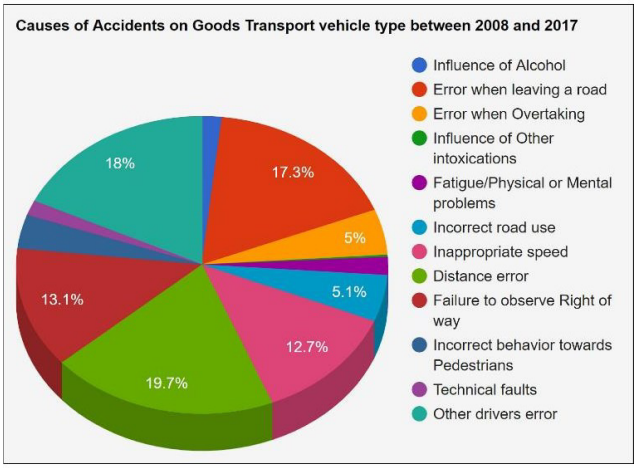

Figure 1 MNIST dataset sample

A. Pre-Processing:

Road accident data preprocessing is a crucial step in getting the data ready for machine learning analysis. To prepare the data for machine learning model training, it must be cleaned, transformed, and organized. The typical preprocessing stages for machine learning-based analysis of traffic accidents are as follows:

Data cleaning involves handling missing values, which can have a detrimental effect on how well machine learning models perform. This can be accomplished using methods like imputation (filling in missing values with the mean or median, for example) or, if missing value rows make up a minor percentage of the dataset, eliminating the rows altogether. Figure 2 Methodology flow chart

B. Model Construction:

The steps involved in building a machine learning model for road accident analysis include choosing the right algorithms, training the model on the preprocessed data, adjusting hyper parameters, and assessing the model's performance. Here is a step-by-step tutorial for creating a machine learning model for analyzing traffic accidents:

Choosing Machine Learning Algorithms: Based on the characteristics of the problem and the data at hand, select the most appropriate machine learning algorithms. The techniques Decision Trees, Random Forest, Support Vector Machines (SVM), Logistic Regression, Gradient Boosting Machines, and Neural Networks are frequently used for road accident investigation. In complicated and high-dimensional datasets, ensemble approaches like Random Forest or Gradient Boosting frequently outperform the competition.

Data cleansing, handling missing values, outlier detection, feature engineering, and categorical variables conversion into numerical representations are all parts of the data preprocessing process that were previously described.

Splitting the preprocessed data into training, validation, and testing sets is known as data splitting. The validation set aids in hyper parameter tuning, the testing set is used to assess the performance of the final model, and the training set is utilized to train the model.

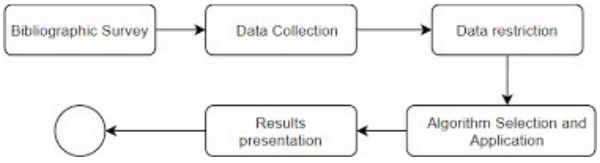

Figure 2 Architecture Model

C: Training and Validation

Building machine learning models for road accident analysis requires several key phases, including training and validation. These processes aid in ensuring that the model effectively learns from the data, generalizes to new data, and performs appropriately in real-world circumstances. Here is a thorough explanation of training and validation in the context of machine learning-based road accident analysis.

D: Model Evaluation

In machine learning-based road accident investigation, model evaluation is a crucial stage. It entails evaluating the trained model's performance to see how well it can forecast accident outcomes and pinpoint contributory elements. Here is how a typical model evaluation is carried out:

Metrics Selection: Based on the precise objectives of road accident study, select the most appropriate evaluation metrics. Regularly employed metrics include: Measures the percentage of accidents that were accurately predicted .Precision is the percentage of accurate positive predictions (accidents that were successfully anticipated) among all positive predictions (accidents that were predicted). The percentage of accurate positive forecasts among all real positive occurrences (actual accidents) is known as recall (sensitivity).

The harmonic mean of recall and precision, balancing the two measures, is the F1-score.Area By adjusting the classification threshold, the Receiver Operating Characteristic curve (ROC AUC) measures the model's capacity to distinguish across various classes (such as severity levels).The average discrepancy between expected and actual accident outcomes is measured for regression tasks using the Mean Absolute Error (MAE) or Mean Squared Error (MSE) measures.

Prediction:

Using machine learning, it is possible to predict the chance of accidents occurring or their severity based on a variety of characteristics and past accident data. The prediction procedure operates as follows: Data gathering: Compile information regarding important road accident data, which may include details about accident sites, weather, vehicle kinds, driver behavior, and accident results (such as the seriousness of injuries or fatalities).Data preprocessing: To manage missing values, outliers, and inconsistencies, clean and preprocess the obtained data. Convert categorical variables into numerical representations by using feature engineering. Splitting the preprocessed data into training and testing sets is known as data splitting. The testing set is used to assess the machine learning model's performance on untried data, whereas the training set is used to train the model.

IV. RESULTS

Accuracy Score:

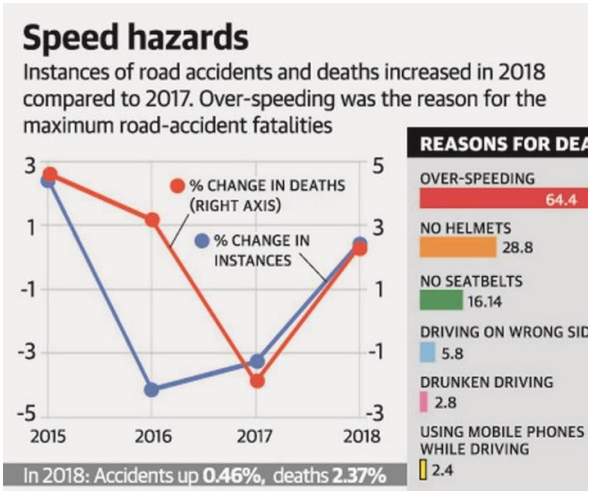

Figure 3 Loss and Accuracy curve

The chance of accidents in particular areas, on particular roads, or during particular times can be predicted by machine learning models. This can assist government agencies in taking proactive steps to increase road safety in high-risk locations. Machine learning can predict the seriousness of incidents, enabling emergency services to properly allocate resources and deliver prompt medical care to those who are harmed. The primary causes of accidents, such as the state of the weather, problems in the construction of the roads, or driver conduct, can be identified by machine learning analysis. This information can guide focused risk-mitigation strategies.

Literature Surveys:

There have been a number of literature reviews and research publications focusing on road accident analysis utilizing machine learning approaches as of my most recent knowledge update in September 2021. These surveys offer thorough overviews of the most recent approaches, algorithms, and market trends. Here are some noteworthy literature reviews for machine learning-based road accident analysis:

F. et al., 2020 "A Survey on Road Traffic Accident Analysis and Prediction Using Data Mining Techniques"

This study examines several data mining and machine learning methods used for analysis and prediction of traffic accidents. It talks about techniques like Bayesian networks, Support Vector Machines, Random Forests, and Decision Trees. The survey identifies possibilities and challenges in the field of road safety.

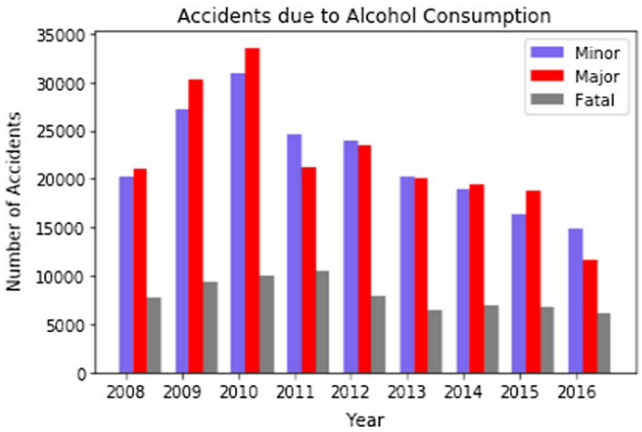

Figure 4 Prediction output

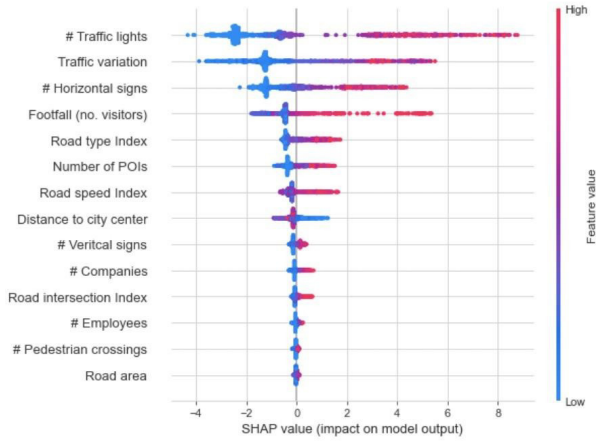

Figure 5 Prediction output

V. CONCLUSION

In conclusion, applying machine learning to analyze road accidents offers a promising way to improve traffic safety, identify accident trends, and put in place efficient preventive measures. With the aid of machine learning algorithms, it is possible to uncover hidden patterns in past accident data and use these insights to generate precise forecasts. The following are the main ideas to remember from this strategy:

Machine learning algorithms are able to anticipate accident incidence and identify high-risk regions, enabling preemptive efforts to stop accidents and lessen their frequency .Assessment of Accident Severity: Machine learning can forecast accident severity, enabling emergency services to Properly allocate resources and deliver appropriate medical care.

REFERENCES

[1] https://www.geeksforgeeks.org/python-tkintertutorial/

[2] It was published in 2020 by Benreguia, Noureddine, Boucherit, and Ouchabanen . A survey on the prediction and analysis of traffic accidents using data mining techniques. The 6(1) issue of the International Journal of Interactive Multimedia and Artificial Intelligence includes pages 5–12.

[3] (2017) Manogarans, S., Lopez, D., and Thota. Machine Learning Methods for Analysis of Traffic Accidents: A Survey.

[4] Garcia-Haro, J., A. Martinez-Balleste, and R. A. Castillo-Rodriguez (2018). A review of the literature on traffic accident severity prediction. 229–235 in Future Distributed Systems and network Conference Proceedings.

[5] In 2019, Zhang, R., Huaneg, S., Wu, J., and Yau, D. K. Y. An extensive analysis of machine learning models for estimating the severity of traffic accidents. 105–116. Accident Analysis and Prevention, 124.

[6] S. Pal, P. K. Saha, and I. Ghosh (2018). An analysis of machine learning techniques for analyzing & preventing traffic accidents. Emerging Technologies, 89, 57–72, Transportation Research Part C.

[7] In 2019, Mohamed, A. M., Ahmed, S. N., and Alhajyaseen, W. K. Using machine learning algorithms, a survey on predicting the severity of traffic accidents. The eighth issue of the International Journal of Advanced Trends in Computer Science and Engineering, 2032-2036.

