ACROPOLIS INSTITUTE OF TECHNOLOGY AND RESEARCH

Computer Science & Engineering (Data Science)

<u>Synopsis</u>

on

Proactive Road Safety: A Hybrid Machine Learning Model for Predicting Traffic Accidents and Severity

1.Introduction

The rise in urbanization and vehicle numbers has led to a surge in road accidents, disrupting transportation and endangering lives. Various factors, such as traffic density, weather, and infrastructure, contribute to the frequency and severity of these accidents. This study focuses on analyzing and predicting road accidents and their severity using machine learning models. By examining historical data, traffic patterns, and external influences, the research aims to develop reliable forecasts and propose solutions to enhance road safety and reduce accident impact.

1.1 Overview

This project focuses on the development of a predictive system for analyzing road accidents, their contributing factors, and the severity of those accidents. By leveraging advanced machine learning techniques, specifically Random Forest and Time Series Analysis, the study aims to predict both the occurrence and severity of accidents on an hourly and weekly basis. The system will analyze historical data on road accidents, traffic patterns, weather conditions, and other relevant variables to provide accurate forecasts. These predictions will enable proactive measures to improve road safety, reduce accident rates, and anticipate the severity of accidents. This approach will contribute to enhancing emergency response strategies, optimizing infrastructure planning, and preventing future accidents through data-driven insights.

1.2 Purpose

The primary purpose of this project is to reduce the incidence and severity of road accidents. By accurately predicting when and where accidents are likely to occur, along with their potential severity, the project aims to provide actionable insights for traffic authorities to enhance road safety, optimize emergency response, and implement preventive measures. Additionally, the system will aid urban planners in designing safer infrastructure and developing strategies that minimize accident risks, improving overall safety and reducing the economic impact of accidents on transportation systems.

2. Literature Survey

2.1 Existing problem

Research in road accident prediction has advanced significantly with the use of various machine learning models. A comprehensive review [2] of these techniques reveals that many approaches rely on algorithms such as Decision Trees, Support Vector Machines (SVM), and linear models like ARIMA and SARIMA, each with its limitations in handling complex, non-linear relationships and generalizability.

For instance, [4] Hemalatha and Dhuwarananath (2024) combined Decision Trees, K-Nearest Neighbors (KNN), and Naive Bayes, but their model’s reliance on social media and government data introduces biases, limiting real-time predictions. Similarly, [6] Krishna et al. (2023) achieved an 88.89% accuracy using Decision Tree and Random Forest in Tamil Nadu; however, their dataset's geographic specificity restricts broader applicability.

Other researchers, such as [5] Gupta et al. (2022) and [1] Augustine & Shukla (2022), applied models like Random Forest, XGBoost, and SVM, achieving high accuracy but encountering challenges like computational complexity and overfitting to localized data. [8] Venkat et al. (2020) faced similar issues with SHAP analysis, while Mor et al. (2020) highlighted the limitations of static data in Haryana, India, which fails to account for dynamic factors like real-time traffic fluctuations. [3] Chelule et al. (2023) employed ARIMA for time series analysis, but its simplicity and data quality concerns may not fully capture the complexity of road accident patterns.

Problem Statement

Road accidents are a significant threat to urban safety, often resulting in fatalities, injuries, and economic losses. Various factors, including traffic congestion, weather conditions, and road infrastructure, contribute to the likelihood and severity of these accidents. This study aims to develop a predictive model that not only forecasts the occurrence of road accidents but also assesses their potential severity. By identifying patterns in accident data, the study seeks to propose proactive solutions that can minimize accident risks, improve emergency responses, and enhance overall road safety in urban areas.

2.2 Proposed Solution

In this project, we are using a combination of Random Forest and Time Series Analysis to develop a more accurate and robust predictive model for road accidents and their severity. Random Forest is chosen for its ability to handle non-linear relationships and identify key contributing factors, while Time Series Analysis captures temporal trends in accident data. This combination ensures better generalization across regions and time periods, offering a more adaptable solution compared to previous models.

The project consists of three primary components:

1. Accident Prediction and Severity Analysis: Using historical accident data obtained from Kaggle, Random Forest and time series models will predict both accident occurrence and severity on an hourly and weekly basis, allowing authorities to take proactive measures.

2. Correlation Between Contributing Factors and Accidents: Statistical techniques will be applied to assess the relationship between factors such as traffic density, weather conditions, and accidents, identifying high-risk periods and locations.

3. Dynamic Safety Recommendations: Based on the predicted accident data, recommendations for improving road safety, such as targeted enforcement, infrastructure improvements, and real-time traffic management adjustments, will be provided to reduce accident risks and enhance emergency response strategies.

3.Theoretical Analysis

3.1 Block Diagram

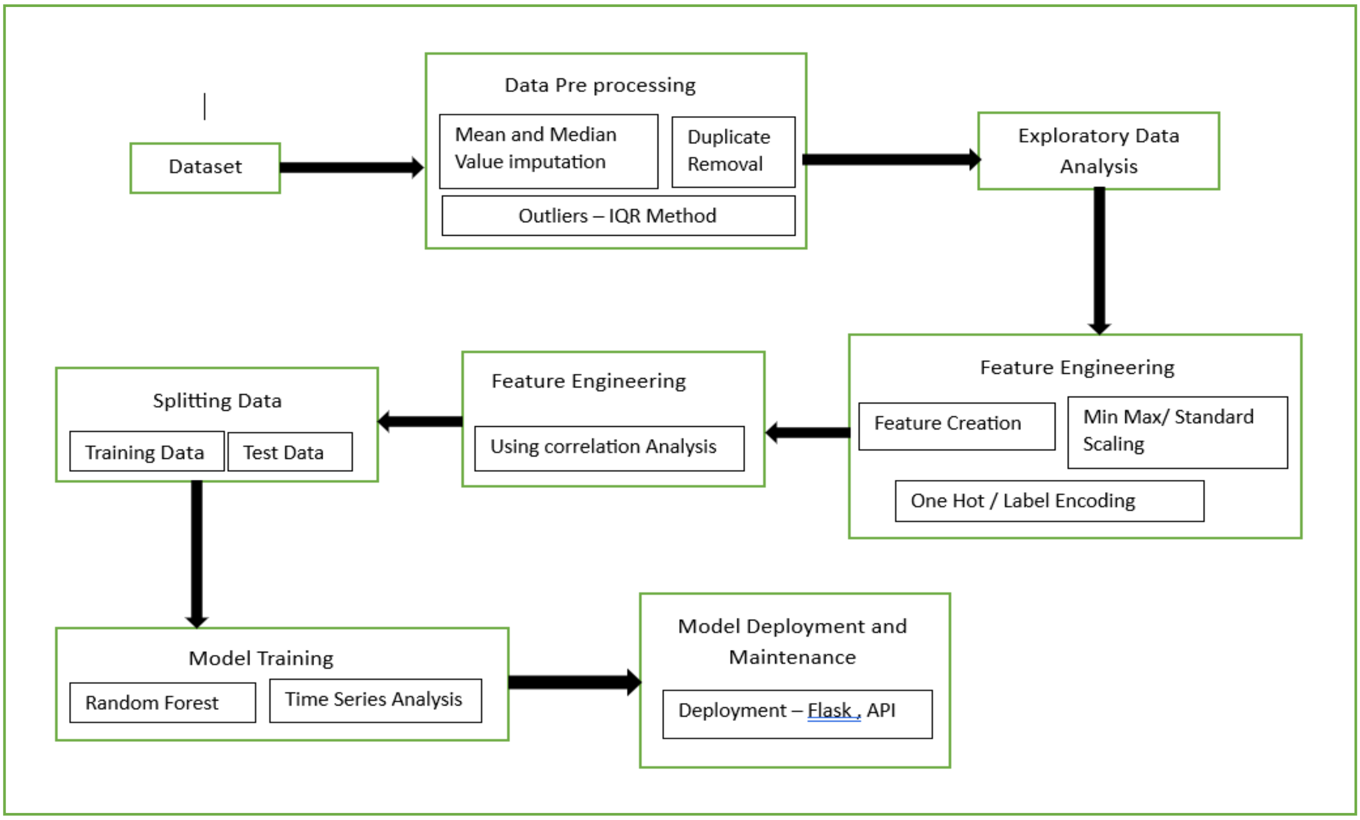

Fig. Block Diagram

This diagram illustrates a machine learning workflow, detailing the steps involved in the modeling process:

1. Dataset: The raw traffic data is the starting point for the process.

2. Data Preprocessing: Mean and Median Value Imputation: Handling missing values by filling them with the mean or median.

- **Duplicate Removal**: Identifying and removing any duplicate records from the dataset.
- **Outlier Detection (IQR Method):** Removing or adjusting values that deviate significantly from the rest of the data using the Interquartile Range (IQR) method.

3. Exploratory Data Analysis (EDA): A stage to visually and statistically explore the dataset to gain insights and better understand the data.

4. Feature Engineering:

- **Feature Creation:** Generating new features from the existing data.
- **Min-Max/Standard Scaling:** Normalizing the data to bring all features onto a similar scale.
- **One Hot/Label Encoding:** Converting categorical variables into numerical form for machine learning algorithms.

5. Feature Selection:

- **Correlation Analysis:** Selecting the most relevant features by checking their correlation with the target variable and among themselves to avoid multicollinearity.

6. Splitting Data: Dividing the dataset into Training Data (to train the model) and Test Data (to evaluate the model).

7. Model Training: Using machine learning algorithms like Random Forest and Time Series Analysis to train the model for predicting traffic congestion or accidents.

8. Model Deployment and Maintenance: Deploying the trained model using Flask and an API for real-time predictions and integrating it into a production environment. And ensuring continuous maintenance of the model for updates and improvements.

3.2 Hardware/Software Design

The design of the road accident prediction system incorporates both hardware and software components to effectively collect, process, analyze, and predict road accident data using advanced algorithms such as Random Forest and Time Series analysis.

Hardware Requirements:

1. **Server:** 32 GB RAM, 2 TB SSD, with high-speed internet connectivity.
2. **User Interface:** Monitors and mobile devices, GPS-enabled for delivering traffic and accident alerts.

Software Requirements:

1. **Operating System:** Windows**.**
2. **Languages**: Python, JavaScript, CSS and HTML for the user interface**.**
3. **Libraries:** Scikit-learn, TensorFlow, ARIMA, Pandas, NumPy.
4. **Version Control:** Git

4. Applications

1. **Proactive Accident Management:**
   Predictive models will enable authorities to manage accident risks proactively. By anticipating high-risk periods and locations, timely measures can be implemented, such as rerouting traffic or deploying emergency services in advance.
2. **Urban Infrastructure Planning:**
   Long-term accident and risk data will inform infrastructure development, such as optimizing road design, building safer intersections, and enhancing pedestrian and cyclist safety features.
3. **Public Transportation Efficiency:**
   Accident prediction data will help optimize public transportation schedules, ensuring that buses, trains, and trams are not delayed by road accidents and can reroute to avoid accident-prone areas.
4. **Smart City Integration:**
   The system can be integrated into smart city frameworks, delivering real-time accident risk data to drivers, autonomous vehicles, and traffic management systems, contributing to smoother and safer traffic flow.
5. **Environmental and Economic Benefits:**
   Reducing accident-related delays lowers vehicle emissions, improving air quality. Economically, reduced congestion and fewer accidents translate into lower costs for commuters and businesses, enhancing overall productivity.

Conclusion

The study of road accidents and the development of predictive models are essential for improving urban safety and mobility. By forecasting accidents and their contributing factors, this research will provide valuable insights for reducing accident risks and enhancing emergency responses. The model's application will extend beyond immediate safety, aiding in long-term infrastructure planning and supporting smarter, more efficient city traffic management. This proactive approach will ultimately contribute to safer, more sustainable urban environments.

References

[1] Augustine, T., & Shukla, S. (2022). Road Accident Prediction Using Machine Learning Approaches. *International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, IEEE, pp. 808–811.

[2] Behboudi, N., Moosavi, S., & Ramnath, R. (2024). Recent Advances in Traffic Accident Analysis and Prediction: A Comprehensive Review of Machine Learning Techniques. A review paper, 26 pages. *Subjects: Machine Learning (cs.LG). arXiv:2406.13968 [cs.LG]*. Available: https://doi.org/10.48550/arXiv.2406.13968.

[3] Chelule, J. C., Ngetich, M. K., Anapapa, A., & Imboga, H. (2019). Time Series Analysis of Road Accidents Using Autoregressive Integrated Moving Average (ARIMA) Model. *International Journal of Science and Research (IJSR)*, 8(4), 2319-7064. From: www.ijsr.net.

[4] Dr. M. Hemalatha, & Dhuwaraganath, S. (2024). Road Accident Prediction Using Machine Learning. *PG & Research Department of Computer Science, Sri Ramakrishna College of Arts and Science*, Coimbatore, Tamil Nadu, India, vol 11, Issue 2202, pp. 454-457.

[5] Gupta, D., Goel, V., Gupta, R., Shariq, M., & Singh, R. (2022). Road Accident Predictor Using Machine Learning. *International Research Journal of Modernization in Engineering, Technology and Science*, vol. 4, May 2022.

[6] Krishna, U. V., Sudhakaran, S., Sanju, S., Vignesh, E., & Kaladevi, R. (2023). Accident Prediction and Analysis Using Machine Learning Models. *International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, IEEE, pp. 37–40.

[7] Mor, N., Sood, H., & Goyal, T. (2020). Application of Machine Learning Technique for Prediction of Road Accidents in Haryana - A Novel Approach. *Journal of Intelligent & Fuzzy Systems*, 38(5), 6627–6636.

[8] Venkat, A., KP, G. V., Thomas, I. S., et al. (2020). Machine Learning Based Analysis for Road Accident Prediction. *IJETIE*, 6(2), 2020

Guided By:

Prof. Pawan Makhija

Group Members:

Ananya Panwar(0827CD221010)

https://github.com/Ananya2517

Bhoomika Sharma(0827CD221021)

https://github.com/Bhoomika71

Devanshi Joshi(0827CD221027)

https://github.com/Ddevanshii10

Hemanshi Adlak(0827CD221036)

https://github.com/Hemanshi-Adlak