ENHANCING SAFETY WITH AI-DRIVEN TRAFFIC ACCIDENT ANALYSIS AND PREDICTION

Student Name: V.Sanjeeu

Register Number: 513523205040

Institution: ANNAI MIRA COLLEGE OF ENGINEERING AND TECHNOLOGY

Department: B.TECH INFORMATION TECHNOLOGY

Date of Submission: 03.05.2025

Github Repository Link:

https://github.com/Sanjeeuanime

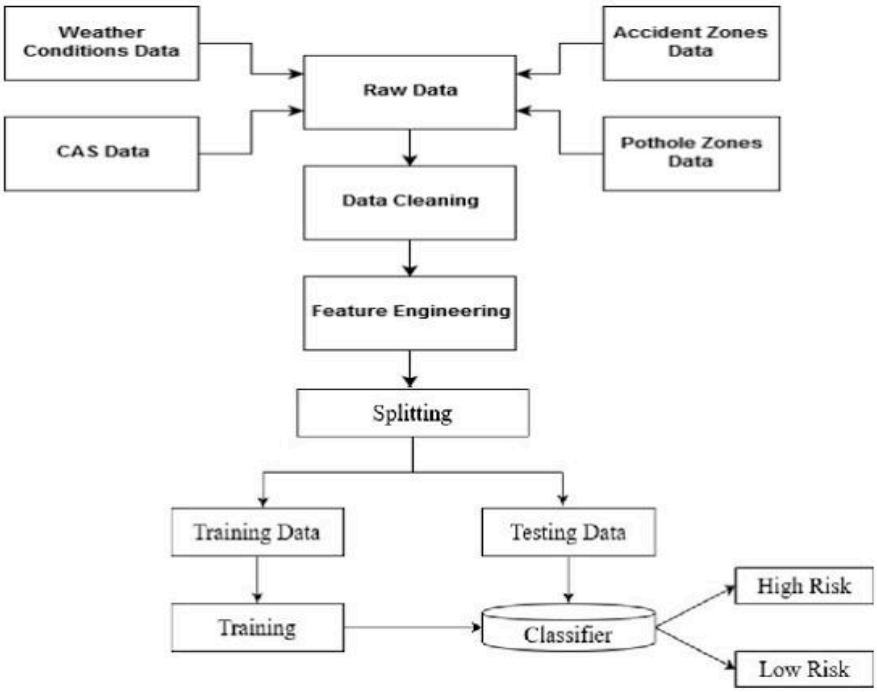
1. Problem Statement

Traffic accidents are a leading cause of fatalities and injuries worldwide. Despite increased awareness and infrastructure improvements, the frequency and severity of road accidents remain high. Traditional data analysis methods are reactive, often used only after accidents occur. There is a need for a proactive approach that can predict and help prevent accidents by analysing patterns in traffic data. AI and machine learning offer powerful tools to analyse large volumes of data and make accurate predictions that can enhance road safety.

2.Project Objectives

Analyze historical traffic accident data to identify significant patterns and trends. Detect accident-prone zones based on geographic, temporal, and environmental data. Build machine learning models to predict the likelihood of accidents under various conditions. Provide actionable insights through visual dashboards to assist policymakers and traffic authorities. Recommend safety measures and infrastructure improvements based on the model's findings.

3.Flowchart of the Project Workflow



4.Data Description

The dataset utilized in the **"Enhancing Safety with AI-Driven Traffic Accident Analysis and Prediction"** project comprises historical traffic accident records collected from public transportation authorities, law enforcement agencies, or open data portals such as Kaggle, Open Data platforms, or government repositories.

1. Data Sources

- Open government datasets (e.g., U.S. DOT, UK Road Safety Data)
- Real-time traffic APIs (for prediction models)
- Satellite and GPS-based traffic data (optional for advanced systems)

2. Data Format

- Structured format: CSV, JSON, or SQL database
- Temporal coverage: e.g., 2015–2024
- Geographical scope: City-wide or nationwide depending on source

5.Data Pre-processing

The dataset used includes records of road traffic accidents, containing features such as: Date and time of the accident, Location (latitude, longitude, city, road type),Weather conditions (rain, fog, visibility),Light conditions (daylight, dusk, night),Vehicle types involved, Severity of the accident (minor, serious, fatal).Driver-related factors (alcohol influence, age, experience)

6.Exploratory Data Analysis (EDA)

Handled missing values using mean/median imputation or deletion. Converted categorical variables into numerical values using encoding (e.g., One-Hot Encoding).Normalized numerical features to bring them to a similar scale. Detected and removed outliers using statistical methods.

7.Feature Engineering

Created new features such as:"Rush Hour Indicator" based on time," Weather Severity Index" based on conditions. "Accident Density Score" using geospatial clustering. Performed dimensionality reduction using PCA to remove redundant features.

8.Model Building

Used classification models to predict accident severity: Logistic Regression, Random Forest, XGBoost,Neural Networks, Performance evaluated using metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC.

9.Visualization of Results & Model Insights

Heat maps showing accident hotspots. Bar graphs for accident distribution across vehicle types and age groups. Confusion matrix and ROC curves for model

performance. Feature importance graphs indicating key factors like poor visibility and road surface condition.

10.Tools and Technologies Used

Programming Language: Python. Libraries: Pandas, NumPy, Matplotlib, Seaborne, Scikit-learn, XGBoost, TensorFlow.Visualization: Tableau

11.Team Members and Contributions

1.SAI KRISHNA K–Data Collection & Preprocessing Lead Responsible for gathering the dataset from reliable sources.
Handles cleaning and preprocessing of the text data, including tokenization, stop word removal, and vectorization.

2.SACHIN B– Exploratory Data Analysis (EDA) & Documentation
Conducts in-depth analysis and visualization to understand the data. Creates plots and graphs for patterns in fake vs. real news. Assists with compiling the report and presentation.

3.SANDHIYA N – Feature Engineering & NLP Specialist
Applies NLP techniques such as TF-IDF, N-grams, and lemmatization. Designs and optimizes feature extraction for improving model performance.

4.SANJAI R– Model Building & Evaluation Lead
Trains various machine learning models and fine-tunes them. Evaluates models using metrics like accuracy, precision, and recall. Handles confusion matrices and ROC curves for comparison.

5.SANJEEU V – Visualization & Final Presentation Designer
Creates charts, visual summaries, and flowcharts. Designs and prepares the final presentation slides for submission. Supports deployment planning if required.