Enhancing Road Safety with AI-Driven Traffic Accident Analysis and Prediction

**STUDENT NAME:S.shanmugapriya**

**REGISTER NUMBER:422223243051**

**INSTITUTION:SURYA GROUP OF INSTITUTIONS**

**DEPARTMENT:B.TECH (AI&DS)**

**DATE OF SUBMISSION:04/05/2025**

**GITHUB REPOSITORY LINK:** <https://github.com/sahana230810/Sahana-A.git>

**1. Problem Statement**

Road traffic accidents remain a major cause of injury and death worldwide. Traditional approaches to road safety rely on reactive measures rather than proactive prediction and prevention. With increasing urbanization and vehicle usage, there is a pressing need to predict high-risk scenarios and accident-prone zones using AI and data analytics.This project addresses the problem of predicting road accidents using machine learning, enabling authorities to take timely preventive actions. The problem is primarily a classification and regression problem—classifying accident severity and predicting accident likelihood based on historical and real-time data.

**2. Project Objectives**

- Predict Accident Risk: Develop AI models to predict the probability and severity of traffic accidents.- Identify High-Risk Zones: Use geospatial data to locate accident hotspots.- Improve Road Safety: Provide insights for urban planners and traffic authorities to implement preventive strategies.- Real-Time Monitoring: Enable predictive analytics using live weather, traffic, and time data.- Evaluate Model Performance: Measure and optimize model performance using real-world accident data.

**3. Flowchart of the Project Workflow**

1. Data Collection – Gather data from sources like open government accident datasets, weather APIs, and traffic sensors.2. Data Preprocessing – Clean and normalize data, handle missing values, encode categorical features.3. Exploratory Data Analysis (EDA) – Visualize trends across time, weather, and location.4. Feature Engineering – Create features like time of day, weather severity, traffic volume, and road type.5. Model Development – Train classification/regression models for accident prediction.6. Model Evaluation – Use metrics like accuracy, F1-score, and MAE to assess performance.7. Deployment (Optional) – Simulate or integrate with traffic management systems.

**4. Data Description**

- Source: Open datasets from government portals (e.g., UK Road Safety, Kaggle, US DOT), weather APIs.- Type: Structured tabular data.- Records & Features: ~100,000+ records; features include time, location, weather, vehicle type, etc.- Target Variable: Accident severity (e.g., minor, major, fatal).- Static/Dynamic: Static historical data with potential for real-time extension.

**5. Data Preprocessing**

- Remove duplicate and inconsistent entries.- Impute or remove missing values (e.g., unknown weather).- Encode categorical variables like road type and weather condition.- Normalize features such as traffic volume and temperature.- Detect and handle outliers (e.g., unusually high speeds).

**6. Exploratory Data Analysis (EDA)**

- Univariate Analysis: Frequency of accidents by hour, day, road condition.- Bivariate Analysis: Heatmaps showing correlation between accident severity and features.- Spatial Analysis: Identify hotspots using geolocation clustering.- Temporal Patterns: Accident trends by month, weekday, and rush hour.- Insights: Rain and night conditions are highly correlated with severe accidents.

**7. Feature Engineering**

- Time-based features: hour of day, day of week.- Weather severity index.- Traffic density categories.- Distance to nearest hospital or junction.- PCA for dimensionality reduction (if needed).

**8. Model Building**

- Models Used: Random Forest, XGBoost (classification), Linear Regression (severity score).- Data split: 80% train, 20% test (stratified sampling).- Metrics: - Classification: Accuracy, Precision, Recall, F1-Score. - Regression (optional): MAE, RMSE.- Hyperparameter tuning using GridSearchCV.

**9. Visualization of Results & Model Insights**

- Confusion Matrix – Visualize classification performance.- Feature Importance – Understand key predictors of accidents.- Accident Hotspot Map – Interactive map showing dangerous zones.- Temporal Graphs – Line/bar plots of accident frequency over time.

**10. Tools and Technologies Used**

* **Languages:** Python
* **Libraries:** pandas, scikit-learn, matplotlib, seaborn, XGBoost, folium (for maps)
* **IDE:** Jupyter Notebook / Google Colab
* **Visualization:** Plotly, Tableau (optional)
* **Data Sources:** Kaggle datasets, government APIs, weather data APIs

**11. Team Members and Contributions**

Data Cleaning , EDA : **Sahana A**

Feature Engineering : **Shanmugapriya S**

Model Development: **Sindhumathi E**

Documentation & Reporting: **Sahana A**