

## Phase-1

**Student Name:** A. Akshaya Mandira

**Register Number:** 410723106001

**Institution:** Dhanalakshmi College Of Engineering

**Department:** Electrical And Electronics Engineering

**Date of Submission:** 28.04.2025

---

### 1.Problem Statement

#### **“Enhancing Road Safety with AI-driven Traffic Accident Analysis and Prediction”**

Road traffic accidents are a leading cause of injuries and fatalities globally.leveraging AI, we aim to analyze historical accident data, uncover contributing factors, and predict potential accident-prone conditions. This will aid policymakers, traffic authorities, and the general public in making informed decisions to enhance road safety.

### 2.Objectives of the Project

- Analyze historical traffic accident data to identify patterns and critical factors contributing to accidents (e.g., weather, time of day, road type).
- Develop a predictive model to estimate the probability of accidents under various conditions.

- Generate actionable insights for accident prevention strategies.
- Visualize key trends and predictions for easier interpretation by non-technical stakeholders.

### 3.Scope of the Project

- Accident frequency by location, time, and weather conditions.
- Severity prediction (minor, major, fatal).
- No real-time accident prediction (based on static historical data).
- Deployment (if done) limited to local or cloud-hosted prototype (no production-grade scaling).

### 4.Data Sources

- Kaggle (e.g., "UK Road Safety Data", "US Accident Dataset")
- UCI Machine Learning Repository
- Dataset must contain features such as date, time, location, weather, vehicle type, accident severity, etc.

#### Data Set Link:

<https://www.kaggle.com/datasets/denkuznetz/traffic-accident-prediction?resource=download>

### 5.High-Level Methodology

- **Data Collection** – Download publicly available datasets from Kaggle, UCI, or government portals.

- **Data Cleaning** – Handle missing values through imputation or removal.  
Remove duplicate records. Normalize or standardize inconsistent data formats (dates, times, categorical values)
- **Exploratory Data Analysis (EDA)** – Univariate and multivariate analysis.  
Visualization of accident frequency over time, weather, and location.  
Correlation analysis between features (e.g., weather and severity).
- **Feature Engineering** – Create new features like "hour of day", "day of week", "accident season" from timestamps. Encode categorical variables using label encoding or one-hot encoding.
- **Model Building** – Logistic Regression, Decision Trees / Random Forests, XGBoost or LightGBM
- **Model Evaluation** – Accuracy, Precision, Recall, F1-Score, ROC-AUC.  
Cross-validation to prevent overfitting. Confusion matrix and classification report for detailed performance.
- **Visualization & Interpretation** – Heatmaps for correlation matrices.  
Timeseries plots for accidents over months/years. Accident hotspot maps (if location data is available). Model feature importance plots.
- **Deployment** – Create a dashboard using Streamlit or Flask to present findings interactively. Allow user input (e.g., weather condition, time) to see accident risk prediction.

## 6.Tools and Technologies •

### Programming Language – Python

- **Notebook/IDE** – Google Colab or Jupyter Notebook

- **Libraries** –

**Data Processing:** pandas, numpy

**Visualization:** seaborn, matplotlib, plotly

**Modeling:** scikit-learn, XGBoost, LightGBM

**(Optional Advanced Modeling):** TensorFlow/Keras (if using neural networks)

**Mapping (Optional if geolocation data available):** folium, geopandas

- **Optional Tools for Deployment** – Streamlit (for fast dashboards)and Flask/FastAPI (for building simple web apps)

## 7.Team Members and Roles

NAME	ROLES	RESPONSIBILITY
Akshayamandira A	Team Leader	Data collection & data cleaning

Ankitha R	Member	Visualization & interpretation
Aswaya sajeev CK	Member	Exploratory data analysis (EDA) & feature engineering
Jayasri J	Member	Model building, Model evaluation