Phase-2

Student Name: Dhivy Bharathi S

Register Number: 412023104304

Institution: Sri Krishna Engineering College

Department: B.E.Computer Science and Engineering

Date of Submission:04-05-2025

Github Repository Link: https://github.com/dhivya-1906?tab=repositories

---

1. Problem Statement

- You're aiming to reduce traffic-related accidents and fatalities by analyzing and predicting accident patterns using AI.

- Type of Problem: Supervised learning – primarily classification (e.g., predicting accident severity or likelihood).

- Importance: Helps governments and transport authorities optimize traffic infrastructure, enforce safety measures, and proactively mitigate high-risk situations.

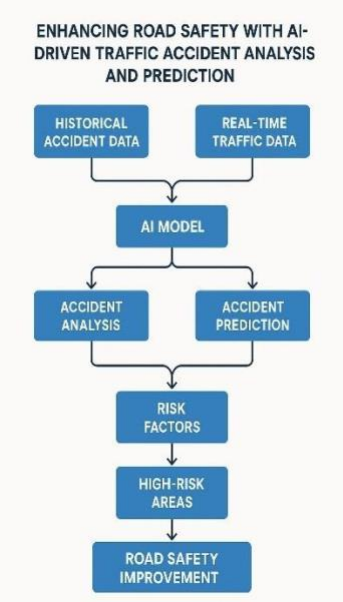
2. Project Objectives

- Identify accident hotspots and causes using data analysis.

- Build a predictive model to forecast accident likelihood or severity.

- Achieve high accuracy and interpretability to support real-world deployment.

3. Flowchart of the Project Workflow



4. Data Description

Source: https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents

Dataset Name: US Accidents (2016 - Present)

- **Type of data**: structured (CSV format)
- **Number of records**: 2.5 million
- **Features**: 40+ (e.g., time, location, weather, severity)

- Static dataset

5. Data Preprocessing

- **Steps to include:**
- Drop duplicates and irrelevant columns (e.g., ID)
- Impute missing values (weather, location)
- Normalize numerical columns if needed

6. Exploratory Data Analysis (EDA)

- **Univariate Analysis:**
    - Countplot of severity, histogram of time-of-day
- **Bivariate/Multivariate Analysis:**
    - Severity vs. weather, time, and location.
- **Insights Summary**:
    - Accidents peak during rush hours, in rain or fog, or at specific intersections

7. Feature Engineering

Create features like

- is_weekend, hour_of_day, weather_risk_level
- Use PCA or feature selection to reduce dimensionality
- Justify each transformation

8. Model Building

- **Compare models like:**
- Random Forest: For interpretability and performance
- **XGBoost***:* For handling imbalanced classes and boosting accuracy
- **Split data (e.g., 80/20), evaluate with:**
- **Classification metrics**: Accuracy, F1-score, ROC-AUC

9. Visualization of Results & Model Insights

Include:

- Confusion matrix.
- ROC Curves

- Feature Importance

Model Comparison Charts

Explain what each visual tells about model performance and reliability.

10. Tools and Technologies Used

- **Programming Language**: Python
- **IDE/Notebook***:* Google Colab, Jupyter Notebook.
- **Libraries:** pandas, numpy, seaborn, matplotlib, scikit-learn, XGBoost, etc.
- **Visualization Tools***:* Plotly

11. Team Members and Contributions

Member Name	Role	Key Contributions
Maladevi G	Project Coordinator & Analyst	Project planning, accident data collection, preprocessing, EDA, User interface design,
Dharshini M	AI Model Developer & Documentation	Model building, training, prediction algorithm design, visualization of results, documentation,

Dhivya Bharathi	software & Integration Engineer	Backend integration, API creation, deployment of AI models, final presentation