Phase 1

Student Name: S. Pooja

Register Number:620123106078

Institution: AVS Engineering College

Department: ECE

Date of Submission: 30.04.2025

---

1. Problem Statement

Road traffic accidents are a major cause of death and injury worldwide, with increasing urbanization and vehicle usage escalating the problem. Traditional methods of accident analysis lack real-time insights and predictive capabilities. This project aims to address this issue by leveraging AI to analyze traffic accident data, identify patterns, and predict accident-prone zones to proactively mitigate risks and enhance road safety.

2. Objectives of the Project

- Analyze historical traffic accident data to identify key contributing factors.
- Develop predictive models to forecast potential accident hotspots.
- Visualize data to highlight high-risk areas and contributing variables.
- Provide actionable insights to support traffic management and urban planning.

3. Scope of the Project

- Features to analyze/build:
    - Accident frequency by time, location, weather, and road conditions.
    - Predictive heatmaps for accident hotspots.
    - Identification of patterns using machine learning models
- Limitations/Constraints:
    - Reliance on publicly available datasets.
    - No real-time data collection or deployment to live traffic systems.
    - Models limited to interpretability rather than real-time automation.

4. Data Sources

- Dataset Source: Kaggle (e.g., "US Accidents (2016-2021)" dataset)
- Accessibility: Public dataset
- Nature: Static dataset downloaded once for analysis
- Additional datasets may include weather or traffic flow data if required.

5. High-Level Methodology

- Data Collection:
    - Datasets will be downloaded from Kaggle or UCI repository.
- Data Cleaning:
    - Handle missing values, remove duplicates, standardize formats (e.g., date-time).
- Exploratory Data Analysis (EDA):
    - Use histograms, box plots, and heatmaps to analyze relationships and distributions.

- Feature Engineering:
    - Create time-based features (e.g., hour of day, day of week), encode categorical variables
- Model Building:
    - Use classification algorithms like Decision Trees, Random Forests, Logistic Regression, and potentially Neural Networks.
- Model Evaluation:
    - Accuracy, precision, recall, F1-score, and confusion matrix.
    - Cross-validation for model robustness.
- Visualization & Interpretation:
    - Use maps, bar charts, and dashboards to present predictions and insights.
- Deployment:
    - Optional; if deployed, will be a simple dashboard using Streamlit or Flask

6. Tools and Technologies

- Programming Language: Python
- Notebook/IDE: Google Colab, Jupyter Notebook
- Libraries: pandas, numpy, matplotlib, seaborn, plotly, scikit-learn, XGBoost, folium, geopandas
- Optional Tools for Deployment: Streamlit or Flask

7. Team Members and Roles

S.Pooja - Responsible for data collection and preprocessing

B.Poornasri - Handled exploratory data analysis and visualizations

R.Sowmiya - Developed and evaluated machine learning models

All members collaborated on documentation, interpretation, and presentation