Phase-1 Submission

Student Name: Janani S

Register Number: 712523106006

Institution: PPG INSTITUTE OF TECHNOLOGY

Department: Electronics and Communication Engineering

Date of Submission: 26/04/2025

1.Problem Statement

Every year, thousands of people are injured or lose their lives due to road accidents. Traffic congestion, poor road conditions, weather changes, and driver behavior all contribute to these accidents. In many places, traffic management systems only react after an accident happens, instead of preventing it.

With the help of Artificial Intelligence (AI) and data analysis, we now have the ability to study traffic patterns and predict when and where accidents are more likely to happen. This project focuses on using AI to analyse real-time and historical traffic data to find accident-prone areas and times. By doing this, we aim to help traffic authorities take early action, such as setting up alerts, changing signal timings, or increasing patrol in high-risk zones, ultimately making roads safer for everyone.

2.Objectives of the Project

Develop a predictive model for traffic congestion and accident-prone zones.

Identify key factors contributing to unsafe traffic conditions.

Analyze real-time or historical traffic data to find trends and anomalies.

Generate actionable insights to assist city planners and authorities in enhancing road safety.

3.Scope of the Project

Analyze traffic volume, speed, accident records, and weather conditions.

Build AI models to forecast traffic congestion and accident probability.

Limitations: Focus on a specific city/dataset; real-time deployment optional; limited to publicly available or simulated datasets.

4.Data Sources

Datasets from platforms like Kaggle, UCI Machine Learning Repository, or city transport APIs.

Open-source traffic data, accident records, and GPS movement logs.

Mostly static data, with optional integration of real-time APIs if time permits.

- *Kaggle – https://www.kaggle.com*
  - *Example datasets: Traffic accident analysis, real-time traffic flow, weather impact on roads.*
- *UCI Machine Learning Repository – https://archive.ics.uci.edu*
  - *Relevant datasets: Road safety data, sensor logs, and vehicle telemetry.*
- *City Transport APIs or Government Portals*
  - *For Indian context:*
    - *data.gov.in*
    - *Chennai Smart City Traffic API*
    - *Regional traffic police portals*
- *Real-Time APIs (optional for future scope):*
  - *Google Maps API – https://developers.google.com/maps/documentation*
  - *HERE API – https://developer.here.com*
  - *OpenWeatherMap API (for weather data) – https://openweathermap.org/api*

5.High-Level Methodology

- **Data Collection** – *We will collect traffic data from publicly available sources such as government traffic departments, Kaggle datasets, or real-time traffic APIs (e.g., Google Maps or HERE API). The data may include accident records, vehicle count, speed patterns, weather conditions, and road types.*

- **Data Cleaning** –*The dataset will be cleaned to handle missing values, remove duplicates, and standardize formats (e.g., date/time formats, location names). Inconsistent or irrelevant data points will be filtered out to maintain accuracy.*

- **Exploratory Data Analysis (EDA)** –*EDA will be performed to discover trends and patterns using visualizations like heatmaps (for accident-prone zones), time series plots (for peak traffic hours), and correlation analysis (between weather, traffic flow, and accidents).*

- **Feature Engineering** –*New features such as "time of day", "day of the week", or "weather conditions" will be created. These will help improve model accuracy by highlighting contextual factors influencing traffic and accidents.*

- **Model Building** –*Predictive models such as Decision Trees, Random Forests, or LSTM (for time-series forecasting) will be used to analyze traffic patterns and predict potential accident hotspots. Classification or regression models may also be applied.*

- **Model Evaluation** –*The models will be evaluated using metrics like accuracy, precision, recall, F1-score (for classification), and RMSE/MAE (for regression or prediction tasks). Cross-validation will be used to ensure reliability.*

- **Visualization & Interpretation** – *Insights and predictions will be presented using dashboards and interactive charts (e.g., via Tableau, Plotly, or Matplotlib) to help authorities easily interpret and act upon the results.*

- **Deployment** – *If applicable, the solution will be deployed as a web-based dashboard or app using tools like Streamlit or Flask, enabling real-time monitoring and prediction of traffic risks for safety departments or the public.*

6.Tools and Technologies

- **Programming Language** – *Python*

- **Notebook/IDE** – *Google colab/ Jupyter Notebook*

- **Libraries** –*Pandas, NumPy, matplotlib, seaborn, scikit-learn, Xgboost, Tenserflow.*

- **Optional Tools for Deployment** – *Streamlit, flask, Gradio*

7.Team Members and Roles

MEMBERS	ROLE	DESCRIPTION
NAKSHATRA. V	Team lead & Data Visualization Engineer	Takes care of Visualization, Interpretation, and Deployment – designing dashboards, interpreting results, and optionally building a web interface for showcasing the project.
JANANI. S	Data processing Engineer	Handles Data Cleaning & Preprocessing – managing missing values, normalizing formats, and preparing data for analysis.
KALPANA. S	Data analyst Engineer	Works on Exploratory Data Analysis (EDA) and Feature Engineering – uncovering insights, trends, and creating new features to enhance

		model accuracy.
AJITH. P	Machine learning Engineer	Focuses on Model Building and Evaluation – experimenting with algorithms, tuning hyperparameters, and assessing model performance.
AKSHAYA KEERTHI. V	Data acquisition engineer	Responsible for Data Collection – sourcing, downloading, and preparing datasets from various platforms and APIs.