Phase-1 Documentation

Student Name: Kavya.R

Register Number: 712523121010

Institution: PPG Institute of Technology

Department: B.E BME (Bio Medical Engineering)

Date of Submission: 25-04-2025

1. Problem Statement

Road traffic accidents are a leading cause of injury and death globally, resulting in significant economic and social costs. Traditional methods of accident prevention often rely on reactive measures rather than predictive insights. This project aims to address the growing need for proactive road safety measures by utilizing AI to analyze patterns in traffic accident data. By identifying high-risk areas, times, and conditions, AI can enable authorities and drivers to take preventative actions, ultimately reducing accidents and saving live.

2. Objectives of the Project

- To build an AI-based model that can predict the likelihood of traffic accidents based on various factors such as time of day, weather conditions, road type, and traffic volume.
- To perform an in-depth analysis of historical traffic accident data to uncover hidden patterns and risk factors.

- To generate visualizations and dashboards that highlight accident hotspots and trends.
- To provide actionable insights that can help urban planners, traffic authorities, and drivers make data-driven decisions to improve road safety.

3. Scope of the Project

Features to be analyzed or developed:

- Historical accident data (date, time, location, severity, cause)
- Weather conditions and their correlation with accidents
- Traffic volume and road infrastructure data
- Accident prediction model using machine learning algorithms
- Interactive visualizations of accident-prone zones

Constraints:

- Use of only publicly available or authorized datasets
- Implementation limited to simulation or prototype level; not a real-time deployment
- Focused primarily on a specific city or region (depending on data availability)
- May use only Python and standard libraries such as Pandas, Scikit-learn, TensorFlow, and visualization tools like Tableau or Plotly.

4. Data Sources

Primary Dataset: Historical traffic accident data from Kaggle, such as the "UK Road Safety" dataset or similar region-specific databases.

Weather Data: Public APIs like OpenWeatherMap or historical weather datasets from NOAA.

Traffic Data: City or government transport department databases (if accessible), or datasets like TomTom Traffic Index.

Type: Public and static datasets primarily, with optional use of dynamic APIs for weather or traffic update9s in advanced phases.

5. High-Level Methodology

Data Collection:

Primary datasets will be downloaded as CSV files.

Weather and traffic data can be obtained through APIs and be integrated if required.

Data Cleaning:

Problems Found: Missing values (e.g., missing weather, road surface)

Duplicates (e.g., duplicate accident records)

Resolution Techniques:

Imputation methods

Dropping duplicates

Standardization of formats with Pandas and datetime parsing

Exploratory Data Analysis (EDA):

Univariate and bivariate analysis (e.g., histograms, boxplots, heatmaps)

Correlation matrices to determine relationships among features

Time-series plots to display accident trends over days, weeks, or seasons

Geographic visualizations (e.g., Folium or Plotly-based accident density maps)

Feature Engineering:

Production of new features like: Time bins (e.g., busy hours, weekend)

Weather severity score

Road risk score (based on the frequency of accidents by location)

Model Building:

Algorithms to be employed:

Logistic Regression (for base binary classification)

Random Forest and XGBoost (for managing non-linearity and feature importances)

Model Evaluation:

Accuracy, Precision, Recall, and F1-Score (for classification tasks)

ROC-AUC for overall model discrimination.

Visualization & Interpretation:

Matplotlib, Seaborn for preliminary analysis

Plotly and Folium for interactive visualizations

Dash or Tableau for dashboard creation

Deployment:

Deployment as a simple interactive dashboard using Streamlit or Dash

Alternately, deliver as a Jupyter Notebook with visualizations and model outputs

6. Tools and Technologies

Category	Tool
Programming Language	Python
Libraries and Frameworks	Pandas, NumPy
Visualization	Matplotlib, Seaborn, Plotly, Folium
Machine Learning	Scikit-learn, XGBoost, TensorFlow/Keras
Web/Dashboard	Streamlit, Dash
Other Tools	Jupyter Notebook or Google

7. Team Members and Roles

Team Member	Roles& Responsibilities
Kavya.R	Team Lead& Data Analyst - Oversees project planning, timelines, coordinates between teams, ensure compliance with safety and data privacy regulations.

Nithyapriya.K Data Engineer - Collect, clean, and analyze accident and traffic-related datasets ,Identify trends, patterns, and accident-prone zones.

Divya.N Visualization and Development Specialist - Design and develop intuitive user interfaces (UI) and dashboard

Kaviya.L Machine learning Engineer - Develop and fine-tune machine learning models for accident prediction , Select appropriate algorithms.

Logeshwari.D Documentation and Reporting Specialist - Maintain detailed project documentation, Record development processes, decisions, and updates.