**Phase-1 Submission**

**Student Name:** Dhavaroshni A

**Register Number:** 712523104016

**Institution:** PPG Institute of Technology

**Department:** BE.Computer Science and Engineering

**Date of Submission:** 25/04/2025

**1.Problem Statement**

*Road accidents pose a serious threat to public safety, leading to loss of lives, injuries, and economic impact. Traditional methods of analyzing traffic incidents are reactive and inefficient in predicting and preventing future occurrences. The goal of this project is to use AI and data analytics to analyze historical traffic accident data and build predictive models that help identify high-risk zones and contributing factors, enabling proactive road safety measures.*

**2.Objectives of the Project**

* *Analyse patterns and trends in traffic accident data.*
* *Identify key factors contributing to road accidents (e.g., time of day, weather, location).*
* *Build a predictive model to forecast accident risk levels in specific areas.*
* *Provide visual insights to support decision-making for road safety improvements.*

**3.Scope of the Project**

**Inclusions:**

* *Data cleaning and Preprocessing of traffic accident datasets.*
* *EDA and visualization to uncover key trends. Propose preventive strategies using ADI-driven insights*.
* *Model building using machine learning algorithms.*
* *Prediction of accident hotspots and risk scores.*
* *Interactive dashboard or visualization*.

**Exclusions / Constraints:**

* *Real-time accident prediction is out of scope.*
* *Deployment limited to demo environments (e.g., Streamlit dashboard).*
* *Reliance on publicly available datasets.*

**4.Data Sources**

* ***Dataset:*** *Traffic accident dataset from [e.g., Kaggle/UCI Repository].*
* ***Type:*** *Public dataset.*
* ***Source:*** *Kaggle -* [*https://www.kaggle.com/datasets/ankushpanday1/global-road-accidents-dataset?resource=download*](https://www.kaggle.com/datasets/ankushpanday1/global-road-accidents-dataset?resource=download)
* ***Nature:*** *Static (downloaded once).*

**5.High-Level Methodology**

* ***Data Collection:*** *Dataset will be downloaded from Kaggle.*
* ***Data Cleaning:*** *Handling missing values, removing duplicates, standardizing formats.*
* ***Exploratory Data Analysis (EDA):*** *Use seaborn, matplotlib for heatmaps, trend analysis, correlation matrix.*
* ***Feature Engineering:*** *Time-related features (e.g., hour, day), weather conditions, road types.*
* ***Model Building:*** *Random Forest, XBoost, Logistic Regression (due to their interpretability and performance).*
* ***Model Evaluation:*** *Accuracy, Precision, Recall, F1-Score, Confusion Matrix.*
* ***Visualization & Interpretation:*** *Charts, maps, dashboards using Plotly and seaborn.*
* ***Deployment:*** *Streamlit app or Jupyter Notebook with interactive visualizations*.

**6.Tools and Technologies**

* ***Programming Language:*** *Python*
* ***Notebook/IDE:*** *Jupyter Notebook / Google Colab*
* ***Libraries:***

*Data Handling: pandas, NumPy*

*Visualization: matplotlib, seaborn, plotly*

*Modelling: scikit-learn, XGBoost, Light GBM*

* ***Optional Deployment Tools:*** *Stream lit or Flask*

**7.Team Members and Roles**

|  |  |  |
| --- | --- | --- |
| *Name* | *Role* | *Work description* |
| *Dhavaroshni A* | *Team leader* | *Data Collection, EDA, and Data Cleaning* |
| *Charumathi J* | *Team member 1* | *Feature Engineering and Model Building* |
| *Nanthini M* | *Team member 2* | *Model Evaluation and Visualization* |
| *Sharvesh S* | *Team member 3* | *Dashboard Creation and Report Documentation* |