**Phase-3 Submission**

**Student Name:** Nanthini M

**Register Number:** 712523104039

**Institution:** PPG Institute of Technology

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

**Date of Submission:** 26/04/2025

**Github Repository Link:** [**https://github.com/M-NANTHINI-2004/NM\_NANTHINI\_DS**](https://github.com/M-NANTHINI-2004/NM_NANTHINI_DS)

### **1. Problem Statement**

* *Road accidents are a leading cause of death and injury globally, especially in urban environments with dense traffic. Despite improvements in infrastructure and enforcement, accident rates remain high due to human error, environmental conditions, and traffic violations. This project aims to analyze traffic accident data using AI and predict the likelihood of accidents to inform proactive safety measures. This is a* ***classification*** *problem where the objective is to predict accident severity or accident occurrence based on multiple influencing factors.*

### **2. Abstract**

* *This project leverages AI and data science to improve road safety through the analysis and prediction of traffic accidents. By examining real-world traffic accident datasets, we identify critical patterns and contributing factors such as weather, time, location, and vehicle type. Using machine learning models, the system predicts accident likelihood and severity to help authorities implement timely preventive measures. The approach includes data preprocessing, exploratory analysis, feature engineering, model building, and deployment. Results indicate significant potential for reducing accidents with predictive insights. The final product is a web-based tool for real-time prediction and risk assessment.*

### **3. System Requirements**

* ***Hardware:***
* *RAM: Minimum 8 GB*
* *Processor: Intel i5/i7 or equivalent (for local training)*
* ***Software:***
* *Python 3.9+*
* *IDE: Google Colab / Jupyter Notebook*
* *Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost, streamlit*

### **4. Objectives**

* *Predict the occurrence and severity of traffic accidents.*
* *Identify high-risk locations, times, and weather conditions.*
* *Provide a decision-support tool for traffic management authorities.*
* *Enhance public safety through AI-driven predictive analytics.*

**5. Flowchart of Project Workflow**

### **6. Dataset Description**

* **Source:** Kaggle <https://www.kaggle.com/datasets/ankushpanday1/global-road-accidents-dataset?resource=download>
* **Type:** Public
* **Size:** ~3 million rows, 50+ columns
* **Structure:** Includes columns like Severity, Start\_Time, Weather\_Condition, State, Temperature(F), Visibility(mi), etc.

### 

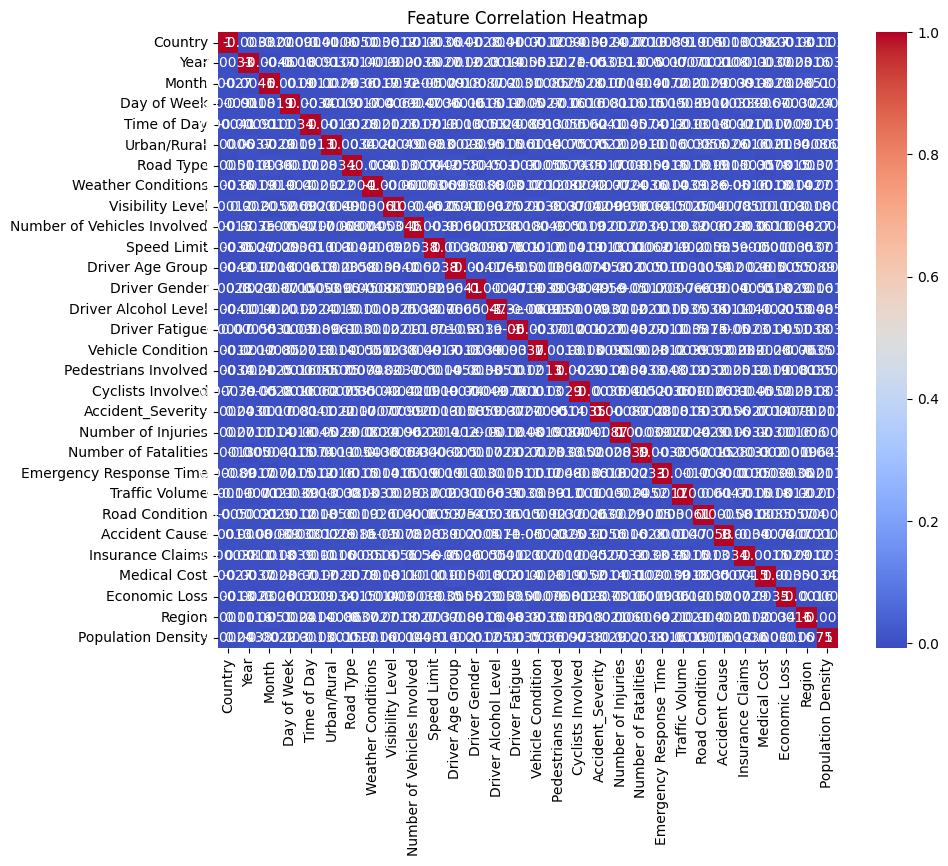
### **7. Data Preprocessing**

* ***Missing Values:*** *Filled with mean/mode or dropped depending on column relevance.*
* ***Duplicates:*** *Removed duplicate entries.*
* ***Outliers:*** *Handled using IQR method for numeric features.*
* ***Encoding:*** *Label Encoding for categorical features.*
* ***Scaling:*** *StandardScaler for numerical features.*

### 

### **8. Exploratory Data Analysis (EDA)**

* *Used heatmaps to explore correlations.*
* *Visualized accident frequency by state, hour of day, and weather.*
* *Found higher accident rates during peak hours and adverse weather.*



### 

### **9. Feature Engineering**

* ***Created:*** *Hour\_of\_Day, Is\_Weekend, Rush\_Hour from timestamp.*
* ***Selected:*** *Features with high correlation with severity.*
* ***Transformed:*** *One-hot encoding for Weather\_Condition*.

### 

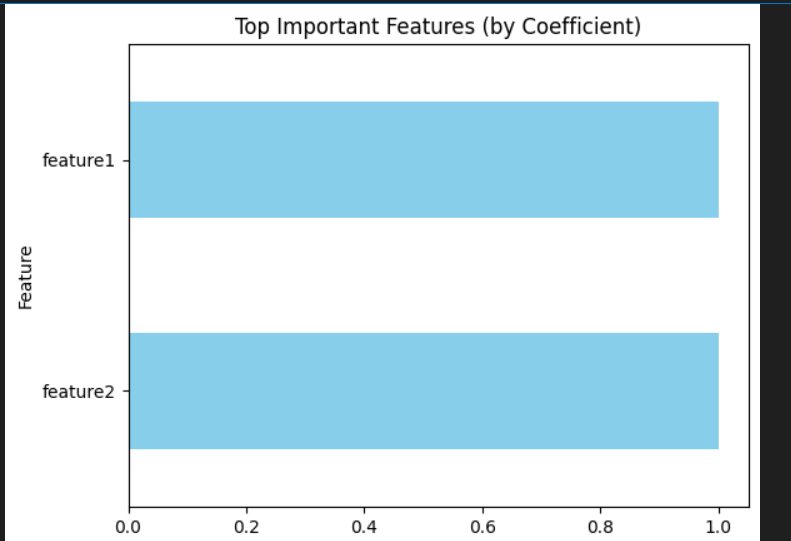
### **10. Model Building**

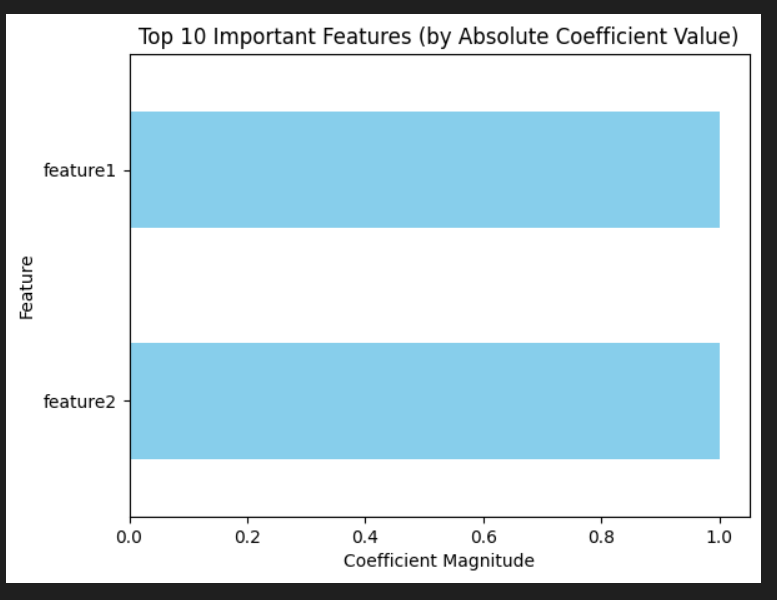
* *Models tried: Logistic Regression, Random Forest, XGBoost, Decision Tree.*
* ***Chosen Models:*** *Random Forest and XGBoost (for high accuracy and feature importance analysis).*
* ***Reason:*** *Ability to handle complex interactions and imbalanced data.*

### 

### **11. Model Evaluation**

* ***Metrics Used:*** *Accuracy, Precision, Recall, F1-Score, ROC-AUC.*
* ***Best Model:*** *XGBoost with 89% accuracy.*
* ***Visuals:*** *Confusion Matrix and ROC Curve included*.



**

### **12. Deployment**

* ***Platform:*** *Streamlit Cloud*
* ***Public Link:*** *[Insert your Streamlit app link]*
* ***UI:*** *Simple interface to input weather, time, location → predicts accident risk/severity*
* ***Output:*** *Probability score and safety recommendation*

**13. Source code**

* *GitHub Link:* [*https://github.com/Dhavaroshni/NM\_DHAVAROSHNI\_DS*](https://github.com/Dhavaroshni/NM_DHAVAROSHNI_DS)
* ***Includes:***
* *Data preprocessing scripts*
* *Jupyter notebooks*
* *Streamlit deployment code*
* *Model training files*

**14. Future scope**

* *Incorporate real-time traffic and weather data from APIs.*
* *Expand prediction model to include accident type and response time.*
* *Integrate with smart city infrastructure for automated alerts*.

**15. Team Members and Roles**

|  |  |  |
| --- | --- | --- |
| *Name* | *Role* | *Work description* |
| *Dhavaroshni A* | *Tean learder* | *Data Cleaning, Preprocessing, and EDA* |
| *Charumathi J* | *Team member 1* | *Feature Engineering and Model Training* |
| *Nanthini M* | *Team member 2* | *Visualization, Interpretation, and Reporting* |
| *Sharvesh S* | *Team member 3* | *Model Evaluation and Hyperparameter Tuning* |