**Phase-3 Submission Template**

**Student Name:** SHANMUGAPRIYA S

**Register Number:** 422223243051

**Institution:** Surya Group of Institutions

**Department:** B.TECH (AI&DS)

**Date of Submission:** 14/05/2025

**Github Repository Link:**

### **1. Problem Statement**

*Road traffic accidents cause significant fatalities, injuries, and economic losses globally. Existing methods based on Static historical analysis are insufficient for proactive intervention. The project aims to use AI and machine learning to Predict accident occurrence and severity, enabling authorities to act early and enhance road safety*

### **2. Abstract**

*This project focuses on predicting the likelihood and severity of road accidents using machine learning techniques Applied to a global road accident dataset. The methodology includes data collection, preprocessing, EDA, model training And evaluation, and visualizations. By identifying patterns and high-risk zones, the model supports authorities in Implementing timely safety measures. The final application is deployed using a web interface for interpretability and access*

### **3. System requirements**

### **Hardware:**

* Minimum: 4GB RAM, Intel i3 or AMD equivalent
* Recommended: 8GB RAM or higher

**Software**:

* Python 3.10+
* IDE: Google Colab or Jupyter Notebook

### **4. Project objectives**

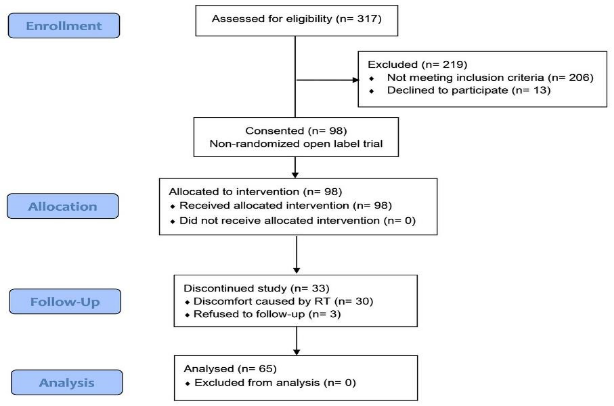
Predict accident severity and likelihood using AI . Identify accident hotspots

. Visualize patterns through dashboards and maps. Support real-world traffic safety planning through decision-support tools

### **5. Flowchart of the work flow**

Data Collection → DataPreprocessing → Exploratory Data Analysis → Feature

Engineering → Model Building → Model Evaluation → Visualization & Insights

**

### **6. Data description**

*Source: Kaggle*

*Format: CSV (Tabular)*

*Records: Thousands*

*Features: Time, location, weather, severity, etc.*

*Target Variables: Accident severity (regression), accident occurrence (classification)*

*Nature: Static, offline dataset*

### **7. Data Preprocessing**

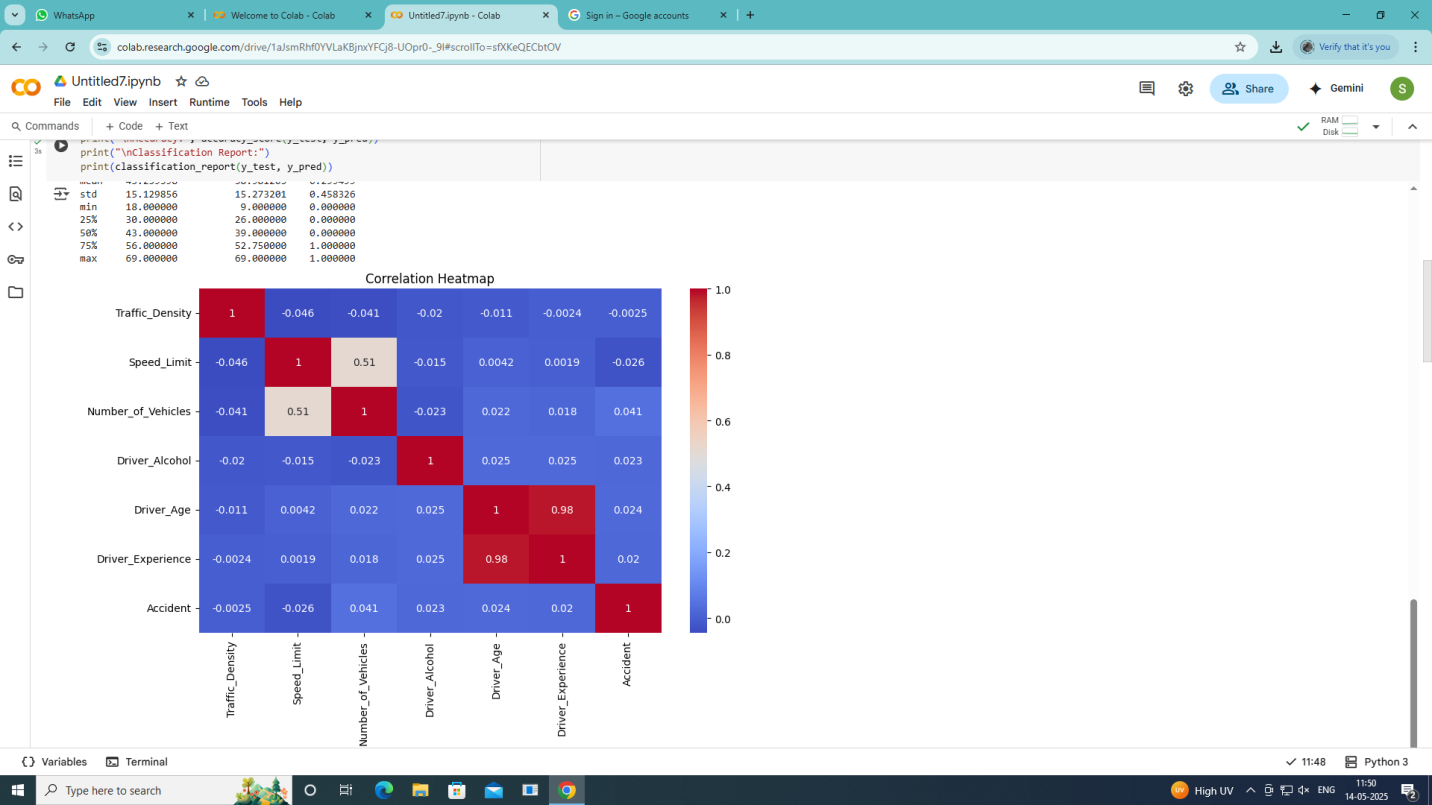
*Missing values: Filled or removed– Outliers: Removed using IQR and Z-scores– Encoding: Label + one-hot encoding– Normalization: Min-Max and Z-score scaling**– Date time & Unit Standardization: Applied to ensure consistent*

### **8. Exploratory data analysis (EDA)**

### *[Build and compare multiple models to solve the defined problem.]*

***Techniques****: Histograms, boxplots, heatmaps, time-series plots, geospatial maps*

*Insights: Most accidents occur during rainy evenings and at intersections. High severity is linked with poor lighting and Wet roads.*

**

### **9. Feature engineering**

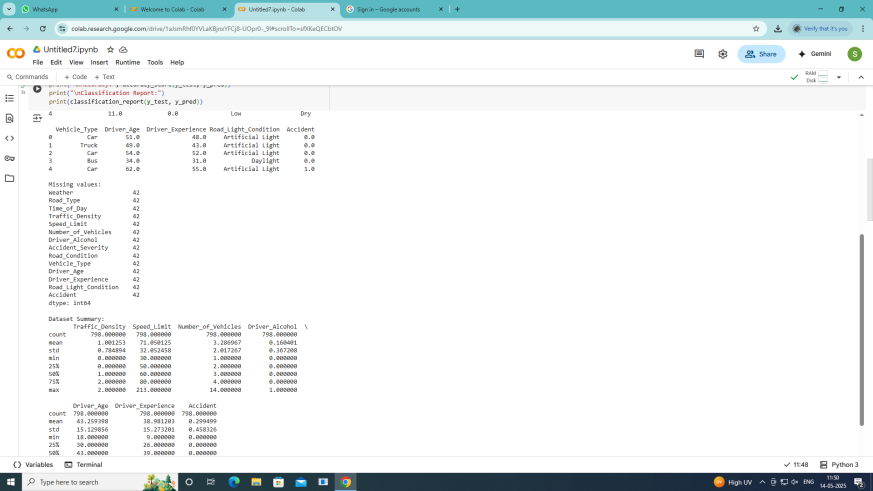
***New Features:*** *Time groupings, weather severity index*

***Transformations****: Log transform, binning*

***Dimensionality*** *Reduction: PCA (optional)*

### **10. Model buliding**

*Classification: Logistic Regression, Random Forest, XGBoost . Regression: Linear Regression, Random Forest Regressed*

* **

### 

### **11. Model evaluation**

### ***Classification:*** *Accuracy, Precision, Recall, F1-score, ROC-AUC*

***Regression****: MAE, RMSE, R^2*

***Visuals****: Confusion matrix, SHAP plots, residuals, risk map*

### **12. Deployment**

**Public link:** [**https://ec9f250af598c09e9f.gradio.live/**](https://ec9f250af598c09e9f.gradio.live/)

**Method**: Stream lit Web App

**Features**: Accepts input, predicts severity/risk, visualizes high-risk zones

**Hosting**: Stream lit Cloud, Firebase, or local demo

### **13. Source code**

<https://colab.research.google.com/drive/1aJsmRhf0YVLaKBjnxYFCj8-UOpr0-_9l#scrollTo=sfXKeQECbtOV&line=9&uniqifier=1>

### **14. Team members and Roles**

Name Role Contributions

1.Sahana A. Project Manager Oversight, coordination ,Model development and evaluation

2.Shanmugapriya S. Data Scientist Data preparation ,EDA, feature engineering

4.Sindhumathi E. ML Engineer Model tuning, deployment, visualization