Phase-3

Student Name: POOJA S

Register Number: 620123106078

Institution: AVS engineering college

Department: ECE

Date of Submission: 18.05.2025

Github Repository Link: https://github.com/Pooja-s123/Webproject.git

---

1. Problem Statement

[Road accidents are a major cause of death and injury globally. The traditional methods of analyzingaccident-prone areas and predicting risks are reactive and slow. The need for real-time, data-driven insightsis critical to enhance road safety. This project aims to utilize Artificial Intelligence to analyze historical trafficaccident data and predict potential accident hotspots. This is primarily a classification problem where thesystem predicts whether a location/time is likely to experience an accident.

2. Abstract

This project focuses on enhancing road safety using AI-driven traffic accident analysis and prediction. Theobjective is to reduce road casualties by proactively identifying accident-prone areas through data analysis.
Historical accident datasets are used to train classification models that predict accident likelihood based on
variables such as time, weather, and location

3. System Requirements

Hardware:
- Minimum 8GB RAM- Intel i5 or equivalent processo

Software:- Python 3.9+- Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, XGBoost, Streamli

4. Objectives

- To build a predictive model for identifying potential accident-prone areas- To derive meaningful insights from historical accident data- To aid traffic management and policy-making with AI-driven insights- To deploy the model in an accessible and interactive format

5. Flowchart of Project Workflow

Insert image here from your chosen tool (e.g., draw.io or Canva):
Data Collection -> Data Preprocessing -> EDA -> Feature Engineering -> Model Building -> Model Evaluation-> Deployment

6. Dataset Description

- Source: Kaggle (e.g., US Traffic Accident Dataset)- Type: Public- Size: ~3 million rows, 40+ columns- Structure: Includes datetime, location, weather, severity

7. Data Preprocessing

- Source: Kaggle (e.g., US Traffic Accident Dataset)- Type: Public- Size: ~3 million rows, 40+ columns- Structure: Includes datetime, location, weather, severity

8. Exploratory Data Analysis (EDA)

- Histograms for accident counts by hour/day- Heatmap for feature correlation- Boxplots for accident severity vs. weather

9. Feature Engineering

- *Derived features: is_peak_hour, weekend_flag- Feature selection using correlation and recursive elimination- Log-transformation for skewed data*

10. Model Building

- Baseline: Logistic Regression- Advanced: Random Forest, XGBoost- XGBoost performed best on validation metrics

11. Model Evaluation

Accuracy: 86%,

F1-score: 0.84- Confusion matrix and ROC curves included

12. Deployment

- Platform: Streamlit Cloud- Public Link: [Insert link here]- UI: Simple input form for location, time, weather- Sample Output: "High risk of accident in this location at the selected time.

13. Source code

All source code including preprocessing, modeling, and deployment scripts are available in the linked GitHubrepository.

14. Future scope

Integration with real-time traffic APIs for live prediction- Use of geospatial data and satellite imagery- Predictive alerts via SMS/email to nearby drivers

13. Team Members and Roles

- S.Pooja - Project Lead, Model Building, Report Preparation- B.Poornasri - Data Preprocessing, EDA, Deployment- R.Sowmiya - Feature Engineering, Evaluation, Documentation