Phase-2 Report

Project Title: Enhancing Road Safety with AI-Driven Traffic Accident Analysis and Prediction

GitHub Link

https://github.com/Divagaran-M/Enhancing-Road-Safety-With-AI-Driven-Traffic-Accident-Analysis-And-Prediction-.git

1. *Problem Statement*

In the ever-buzzing ballet of city streets, accidents strike without warning. What if we could forecast the dance of danger before it steps into motion? This project aims to use the predictive power of AI to anticipate traffic accidents based on real-time and historical data - a digital crystal ball to save lives.

2. *Project Objectives*

- Develop machine learning models to predict accident likelihood across locations and timeframes.
- Identify contributing factors like weather, lighting, road conditions, and traffic volume.
- Build heatmaps and risk zones to guide urban planning and emergency readiness.
- Integrate model outputs into a dashboard for intuitive use by traffic management teams.
- Promote proactive safety measures - not reactive responses.

3. *Project Workflow (Flowchart)*

Stages: Data Collection -> Data Cleaning -> EDA -> Feature Engineering -> Model Building -> Evaluation -> Deployment

4. *Data Description*

- Source(s): Open-source datasets (e.g., Kaggle, government transport departments)

- Size: [Insert details]

- Features: Date/time, weather, location, accident severity, vehicle types, etc.

- Target: Accident occurrence / severity

- Nature: Structured, spatio-temporal data

5. *Data Preprocessing*

- Cleaned missing values and inconsistent entries.

- Normalized and encoded features (OneHot for categories, scaling for numerics).

- Geo-coordinates processed for location clustering.

- Merged datasets from different sources with timestamp alignment.

6. *Exploratory Data Analysis (EDA)*

- Accident density by hour, day, and location.

- Correlation between weather and accident frequency.

- Trendlines and anomaly detection.

- Visuals include heatmaps, bar charts, scatter plots.

7. *Feature Engineering*

- Combined time and weather to create 'risk windows'.

- Created location risk scores using historical frequency.

- Binary features for holiday/weekend.

- Lag features for temporal modeling.

8. *Model Building*

- Algorithms: Logistic Regression, Random Forest, XGBoost, and LSTM (for time series).

- Split: 80-20 with stratification.

- Metrics: Accuracy, Precision, Recall, F1-score, AUC.

9. *Results and Visualization*

- ROC curves and confusion matrices.

- Feature importance charts.

- Interactive map of accident hotspots.

- Dashboard prototype (if available).

10. *Tools and Technologies*

- Languages: Python

- Libraries: pandas, scikit-learn, seaborn, TensorFlow/PyTorch, folium

- Deployment: Streamlit/Gradio for demo UI

- Platforms: Jupyter/Colab, GitHub

11. *Team Contributions*

- **Vicknesh V**: Data Collection and Cleaning
- **Srinivasan N**: EDA and Feature Engineering
- **Ashwin G**: Model Building and Evaluation
- **Divagaran M**: Report Writing and Presentation