Student name: S.Subarthana

Register Number: 623123205043

Institution :VSA Group of Institutions

Department: B.Tech IT information technology

Date of submission:8:05:2025

Github link:

https://github.com/subasubarthu/subarthana.git

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

phase 2

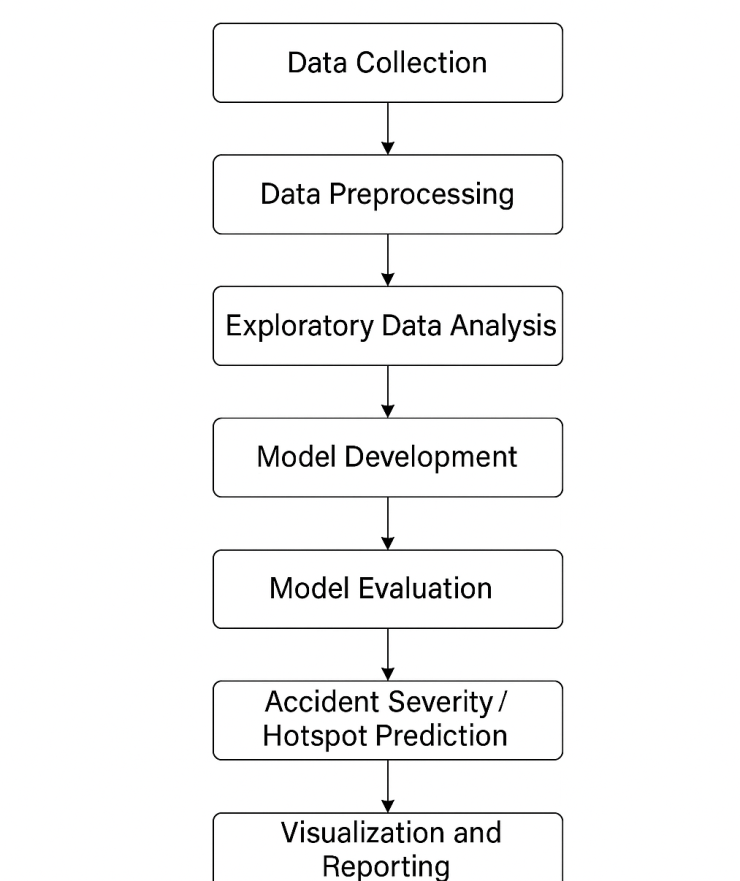
**1. Problem Statement (Phase I)**

* Traffic accidents are a leading cause of death and injury worldwide. With
* increasing urbanization, traffic congestion, and driver distractions, traditional
* safety measures are insufficient. This project aims to address these limitations
* using AI-driven data analysis and predictive modeling.

**2. Project Objectives (Phase I)**

* Analyze historical traffic accident data.
* Identify accident-prone areas and major contributing factors.
* Predict accident severity using machine learning.
* Deliver actionable insights for policymakers and public safety.

**3. Flowchart of the Project Workflow**



**4. Data Description**

● Dataset Name:<https://kaggle.com/competitions/road-traffic-accidents?utm_source>

● Source: UAE - Roads and Transport Authority (RTA)

● Type of Data: Structured tabular data

● Records and Features: 7,000+ accident records with 30+ features (numeric + categorical)

● Target Variable: Accident Severity (e.g., Minor, Major, Fatal)

● Static or Dynamic: Static dataset

**5. Data Preprocessing**

* Removed duplicates, handled missing values using imputation.
* Encoded categorical features using Label and One-Hot encoding.
* Standardized time formats and normalized numeric fields.
* Added derived columns: hour, weekday, severity score, etc.

**6. Exploratory Data Analysis (EDA)**

* Bar charts and histograms for driver age, sex, and experience
* Heatmaps of accident frequency by time of day and day of week
* Pie charts for vehicle types and accident causes
* Correlation plots to find feature impact on severity

**7. Feature Engineering**

* Temporal: Hour, day type (weekend/weekday)
* Spatial: Area code, intersection proximity (if available)
* Environmental: Weather and light condition scores
* Behavioral: Risk profiles based on driver experience + vehicle type

**8. Model Building**

* Random Forest – best for interpretability
* XGBoost – highest accuracy in cross-validation
* Neural Networks – effective for non-linear patterns
* LSTM (optional) – for sequential accident trends (e.g., time series)

**9. Visualization of Results & Model Insights**

* Heatmaps: High-accident zones
* Dashboards: Built using Plotly
* Model Interpretability:
* SHAP values for feature importance
* Confusion matrix, Precision/Recall curves

**10. Tools and Technologies Used**

* Language: Python
* Libraries: Pandas, NumPy, Scikit-learn, XGBoost, TensorFlow, Plotly
* EDA & Visualization: Seaborn, Matplotlib, Plotly
* Deployment: Streamlit, Flask (optional), Docker (optional)

**11. Team Members and Contributions**

Member Role Responsibilitie

subarthana Project Lead Project planning, integration, reporting  
tharun kumar Data Engineer Data cleaning, transformation  
sindhuja Data Scientist Modeling and evaluation  
suriyavani Visualization Expert Dashboards and EDA graphics