PROJECT TITLE:ENHANCING ROAD SAFETY WITH AI-DRIVEN TRAFFIC ACCIDENT ANALYSIS AND PREDICTION.

Student Name: M.Saravanan

Register Number: 511923205301

Institution: Priyadharshini Engineering College

Department: B.Tech--IT

Date of Submission: 08/05/2025

Github Repository Link:

1. PROBLEM STATEMENT

Despite advancements in vehicle technology and infrastructure, traffic accidents continue to pose a serious threat to public safety, causing thousands of fatalities and injuries annually. Traditional methods of accident analysis and prevention rely heavily on reactive measures and lack the ability to accurately predict and prevent future incidents. There is a critical need for a proactive, data-driven approach to identify high-risk areas, understand contributing factors, and forecast accident likelihood in real time. Leveraging artificial intelligence (AI) and machine learning can enable the analysis of vast amounts of traffic data to uncover hidden patterns and deliver predictive insights. This project aims to develop an AI-driven system that enhances road safety by analyzing historical accident data and predicting potential accident zones, thereby supporting timely interventions and informed urban planning.

The despite significant efforts to improve road safety, traffic accidents continue to occur at alarming rates, often due to preventable factors. Current systems are predominantly reactive rather than predictive, lacking real-time analytical capabilities. There is a need for an intelligent solution capable of predicting and mitigating accident risks based on a variety of dynamic and historical data points.

2. OBJECTIVES

1. Data Acquisition and Integration

- Collect historical and real-time traffic accident data from various sources (e.g., government databases, sensors, mobile apps, weather services).
- Integrate multiple datasets (e.g., road conditions, traffic density, weather, time of day) to create a comprehensive traffic accident database.

2. Data Processing and Feature Engineering

- Clean, preprocess, and transform raw data to ensure consistency and accuracy.
- Identify and engineer relevant features (e.g., weather conditions, traffic volume, time variables) that significantly affect accident occurrence.

3. Pattern Recognition and Analysis

- Analyze accident data to identify spatial and temporal patterns (e.g., accident hotspots, peak accident times).

- Determine the relationship between environmental, vehicular, and human factors and accident severity.

4. Predictive Modeling

- Develop machine learning models (classification, regression, clustering) to predict the likelihood of accidents in specific locations and times.
- Train, validate, and optimize models for high accuracy and reliability.

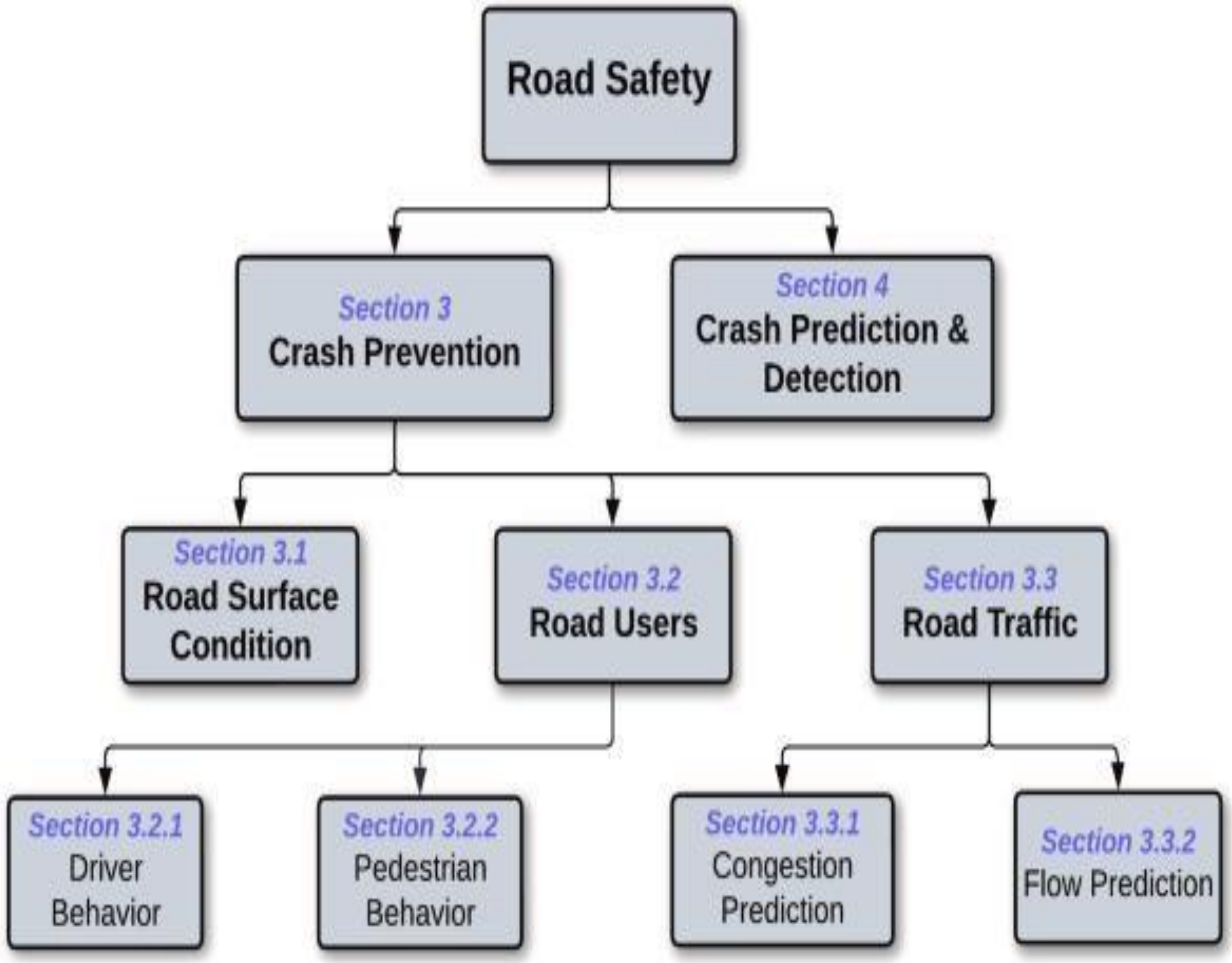
5. Risk Zone Identification

- Use clustering algorithms (like K-means, DBSCAN) to identify and map high-risk areas ("black spots") prone to frequent accidents.
- Generate dynamic risk maps for real-time visualization.

6. Recommendation Systems

- Provide actionable recommendations for improving road safety (e.g., adding warning signs, changing traffic signal timings, adjusting speed limits based on predictions).
- Suggest optimal routes for drivers based on accident risk predictions

3. FLOW CHART OF THE PROJECT WORKFLOW



4. DATA DESCRIPTION

Structured Data: Tables with rows (each row = one accident) and columns (features).

- **Large Volume**: Millions of records over multiple years for better AI model learning.
- **Multi-Source**: Data combined from police reports, weather systems, vehicle records, GPS devices.
- **Real-time Data (Optional for Prediction Systems)**: Live traffic feed, real-time weather updates, dynamic GPS data.
- **Road Type**: Highway, urban street, rural road.
- **Road Hazards**: Construction area, potholes, debris.

- **Speed at Impact**: Vehicle speeds before collision.
- **Driver Distraction**: Phone use, fatigue indicators (if available).
- **Environmental Lighting**: Daylight, twilight, darkness.
- **DATASET LINK:**

5. DATA PREPROCESSING

1. **Integrate** different data sources (accidents, traffic, weather, vehicle data).
2. **Clean** the data by removing duplicates, fixing errors, and handling missing values.
3. **Transform features**: Encode categorical variables, normalize numerical values, and extract useful datetime features.
4. **Process geospatial data**: Use clustering and distance calculations to create meaningful location-based features.
5. **Engineer additional features**: Create risk indices and interaction features to capture hidden relationships.
6. **Split the data** into training, validation, and test sets to train AI models effectively.

---

6. EXPLORATORY DATA ANALYSIS (EDA)

1. Understanding the Data Structure

- **View data types**: Check if columns are numeric, categorical, or datetime.
- **Check basic statistics** (using .describe() in Python):
    - Mean, median, min, max, standard deviation of numerical features (e.g., speed, traffic density).
- **Check missing values**: Identify which columns have missing data and how much.

2. Multivariate Analysis (Three or More Features Together)

- **Accidents by Weather + Time of Day**:
  - Example: Rainy nights may have more fatal accidents than rainy days.
- **Accidents by Road Type + Traffic Volume**:
  - High traffic on highways vs low traffic on rural roads — different accident patterns.
- **Use 3D plots or multi-variable heatmaps** to understand these relationships.

3. **Key Observations and Insights**
   At the end of EDA, summarize:

- Where and when do most accidents happen?
- Under which conditions (weather, road surface) are accidents most severe?
- What are the common factors in fatal accidents (e.g., speeding, night, rain)?
- Are there particular roads or regions that require urgent safety improvements?

7.

Feature Engineering means creating new, meaningful input features from raw data to improve the performance of AI/ML models.
Good feature engineering is critical for accurate accident analysis and prediction.

1. **Datetime Features**

Extract useful information from accident timestamps:

- **Hour of Day** (e.g., rush hours vs late night)
- **Day of Week** (weekday vs weekend accidents)

Example: Create a feature "Night Time" (1 if accident between 10 PM – 5 AM, else 0).

2. **Weather Features**

From weather data at the accident time:

- **Rain Intensity** (none, light, heavy)
- **Visibility Level** (meters/kilometers)
- **Road Surface Condition** (dry, wet, icy)

Example: Create a "Bad Weather" feature: 1 if heavy rain, snow, or fog; else 0.

3. **Traffic Features**

Use traffic flow and congestion data:

- **Traffic Volume** (vehicles/hour)

- **Speed Limit Violation** (actual speed - allowed speed)
- **Average Speed** (on that road segment)
- **Congestion Level** (categorized: low, medium, high)

Example: Feature “OverSpeed” = 1 if actual speed > speed limit by 20%+.

4. **Accident Context Features**

From accident reports:

- **Number of Vehicles Involved**
- **Number of Casualties**
- **Accident Type** (head-on collision, rear-end, rollover)
- **Lighting Conditions** (daylight, darkness with/without streetlights)

Example: Feature "Night_No_Streetlights" = 1 if accident occurred at night with no streetlights.

8. MODEL BUILDING

1. Define the Problem Clearly

- **Classification Task**:
  Predict the **severity** of an accident (e.g., Minor, Major, Fatal).
- **Prediction Task**:
  Predict the **likelihood** of an accident happening given traffic, weather, time, and location features.

2. Splitting the Data

- Split the dataset into:
  - **Training set** (e.g., 70%) — to train the models.
  - **Validation set** (e.g., 15%) — to tune model hyperparameters.
  - **Test set** (e.g., 15%) — to evaluate final model performance.
- Use **Stratified Sampling** if classes (like "fatal accidents") are imbalanced.

3. Model Evaluation

- Evaluate models on the test set.
- Compare model performance across metrics (Accuracy, Precision, Recall, F1).

- Plot **Confusion Matrix** to understand how well the model is classifying each severity class.
- For prediction models, use **Mean Squared Error (MSE)** or **Root Mean Squared Error (RMSE)** if predicting accident risk score.

4. Feature Importance Analysis

- After model training, find which features are most important.
- Helps you explain model predictions and improves trust:
  - E.g., "Rain + Night + High Traffic Volume" may be top predictors of severe accidents.

9. ISUSALIZATION OF RESULTS AND MODEL INSIGHTS.

1. Model Performance Visualization

✅Accuracy and Confusion Matrix

- Plot a **Confusion Matrix** to see how well the model classifies accident severity (e.g., Minor, Serious, Fatal).
- Helps identify **where the model is making mistakes** (e.g., confusing "Serious" and "Fatal" accidents).

Tools: Seaborn heatmaps, Matplotlib.

2. Feature Importance Visualization

✅Top Factors Causing Accidents

- Plot **Feature Importance Bar Charts** showing the most influential factors:
  - Weather conditions
  - Traffic density
  - Time of day
  - Over-speeding

Example:
"Bad Weather" might contribute 30% to accident severity prediction.

- **SHAP (SHapley Additive exPlanations) Values**:
  Visualize **how each feature** positively or negatively influences a specific accident prediction.

Tools: SHAP library, XGBoost feature importance, Random Forest feature importance.

3. Exploratory Analysis Visualization

✅Accident Distribution Charts

- **Bar Charts or Pie Charts** for:
    - Accident count per severity level.
    - Accidents under different weather conditions.
    - Accidents per road type (highway, urban, rural).

4. Geospatial Visualization

✅Accident Hotspot Mapping

- Use **Geospatial Maps** to plot accident locations:
    - Highlight **accident hotspots** in cities or highways.
    - Use **heatmaps** to show accident density.

Tools: Folium, Kepler.gl, Plotly Express.

10. TOOLS AND TECHNOLOGIES

.PROGRAMMING LANGUAGE: PYTHON

.NOTEBOOK ENVIRONMENT: GOOGLE COLAB

11. TEAM MEMBERS :

1) ***M.TAMIL SELVAN: PROBLEM STATEMENT,OBJECTIVES***
2) **R.UKESH:DATA DESCRIPTION,FLOW CHART**
3) **M.SARAVANAN:DATA PREPROCESSING,EXPLORATORY DATA ANALYSIS (EDA)**
4) **S.VIJAY:FEATURE ENGINEERING,MODEL BUILDING,**

VISUSALIZATION OF RESULTS AND MODEL INSIGHTS.