Github : https://github.com/dinesh88845373/NM-DS-Project.git

ENHANCING ROAD SAFETY WITH AI-DRIVEN TRAFFIC ACCIDENT ANALYSIS AND PREDICTIONS

Student Name: k.Dinesh kumar

Register Number:511923104010

Institution:priyadharshini engineering college

Department:CSE

Date of Submission: 08-05-2025

1.Problem Statement:

Understanding the Problem

Road safety is a **global concern**, with millions of accidents occurring annually. Traditional accident analysis depends on historical data, often reactive rather than proactive. AI-driven accident prediction models **use real-time data and machine learning algorithms** to improve safety measures.

Problem Refinement

● The project aims to detect accident trends using big data and AI algorithms. ● It includes factors such as:

- ✧ **Weather conditions** (rain, fog, night driving).
- ✧ **Road type** (highway, urban, rural).
- ✧ **Driver behavior** (speeding, violations, alcohol consumption).

Type of Problem

- **Classification** → Predicting accident severity (minor, major, fatal). ● **Regression** → Estimating accident frequency based on historical trends.
- **Clustering** → Identifying high-risk accident zones.

Why Solving This Matters?

- **Reduced Fatalities** → AI predictions help implement safety measures.
- **Traffic Management** → Authorities can optimize traffic signals and emergency response.
- **Smart Cities** → AI-integrated solutions ensure safer mobility.

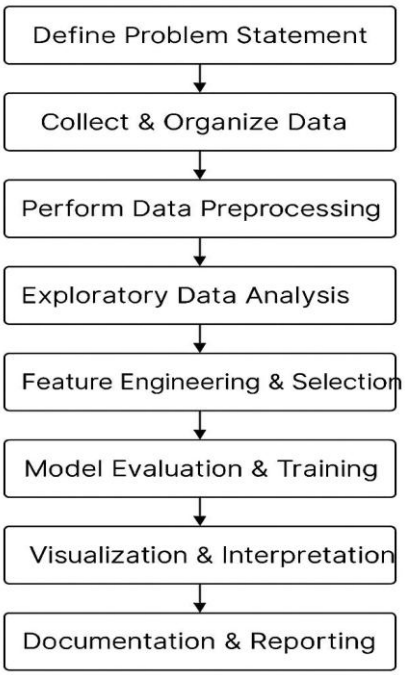
2. Project Objectives:

Key goals include:

- ✓ Develop an AI model that predicts accident likelihood.
- ✓ Analyze risk factors influencing accidents.
- ✓ Provide actionable insights for road planners and safety officers.
- ✓ Enhance model interpretability using statistical insights.
- ✓ Integrate real-time traffic surveillance data for dynamic analysis.

3. Flowchart of the Project Workflow:

Here's the structured process for developing the AI-driven accident prediction system:



is **flowchart** represents the logical steps of **problem identification, AI modeling, evaluation, and actionable insights** for accident prevention

.

4. Data Description:

- **Dataset Source** → Kaggle, UCI ML Repository, official government accident records.
- **Type of Data** → Structured (tabular), time-series, and geographic locationbased datasets.
- **Size of Dataset** → Thousands of records across multiple accident attributes.
  - **Target Variable** → Prediction of accident severity or occurrence.

5. Data Preprocessing:

✓ **Handle Missing Data** → Use mean or median imputation.

✓ **Remove Duplicate Entries** → Ensuring dataset accuracy.

✓ **Detect Outliers** → Using statistical methods like Z-score or IQR.

✓ **Normalize Data** → Standardization for model efficiency.

✓ **Categorical Encoding** → Label encoding for categorical variables.

6. Exploratory Data Analysis (EDA):

EDA helps identify **hidden patterns** and **correlations** in accident data:

- **Univariate Analysis** → Distribution of accident severity using histograms.
- **Bivariate Analysis** → Relationship between speed and accident severity using scatterplots.
- **Correlation Matrix** → Understanding feature dependencies using heatmaps.

✓ **Insights Generation** →

- Accident rates **increase during rainy seasons**.
- **Speeding contributes significantly to fatal crashes**.
- **Highway accidents occur more frequently than urban road crashes**.

7. Feature Engineering:

Transforming raw data into meaningful **predictive variables**:

✓ **Extracting time-based features** (peak accident hours).

✓ **Creating risk score features** for accident-prone areas.

✓ **Dimensionality reduction** to optimize model processing power.

8. Model Building:

Selecting **AI-powered models** for accident prediction:

- ✓ **Classification Models** → Decision Trees, Random Forest, XGBoost.
- ✓ **Regression Models** → Linear Regression, Gradient Boosting.
- ✓ **Deep Learning (Optional)** → Neural Networks for complex feature interactions.
- ✓ **Hyperparameter Tuning** ensures optimal model performance.

9. Visualization of Results&Model Insights:

- ✓ **Graphical analysis helps interpret model effectiveness**:
- ✓ **Confusion Matrix** → Evaluating classification performance.
- ✓ **ROC Curve** → Assessing model accuracy.
- ✓ **Feature Importance Plot** → Identifying key accident risk factors. ✓ **Trend Analysis** → Seasonal accident spikes and high-risk zones.

10. Tools and Technologies Used:

- **Programming Language** → Python
- **Libraries** → pandas, numpy, scikit-learn, seaborn, matplotlib
- **Visualization** → Tableau, Power BI
- **IDE** → Jupyter Notebook, Google Colab

11. Team Members and Contributions:

1.Team Members:

Guna

Annamalai

Guruvarma

Mahesh

Dinesh Kumar

2.Responsibilities:

Data Cleaning:

- *Guna* – Managed handling of missing values, formatting, and outlier detection. ● *Annamalai* – Performed normalization and encoded categorical variables.

Exploratory Data Analysis (EDA):

- *Guruvarma* – Created visualizations and performed initial data inspection. ● *Mahesh* – Conducted statistical analysis and identified data trends.

Feature Engineering:

- *Annamalai* – Created new features and transformed existing ones.
- *Guruvarma* – Performed feature selection and dimensionality reduction.

Model Development:

- *Dinesh Kumar* – Built and trained machine learning models.
- *Guna* – Tuned model parameters and evaluated model performance.

Documentation and Reporting:

- *Mahesh* – Compiled findings into reports and summaries.
- *Dinesh Kumar* – Prepared presentation materials and final documentation.