*Student Name : ROSHINI. S*

*Register Number :422023104042*

*Institution : Sri Rangapoopathi College Of Engineering*

*Department : Computer Science Engineering*

*Date Of Submission : 05/05/2025*

*Github Repository Link :*

**Lj**

1.Problem Statement

Predicting traffic accident severity is a vital challenge in public safety and urban planning. This project aims to forecast the severity level of a traffic accident using a variety of factors including road conditions, weather, time, and driver-related behaviors.

The problem type is classification, as the target variable represents the severity level of the accident (e.g., minor, serious, fatal).

Timely and accurate prediction can help in dynamic traffic management, resource allocation for emergency services, and improving road safety policies.

2.Project objectives technical

* Develop a classification model to predict the severity of a traffic accident.
* Identify key features influencing accident severity.
* Provide actionable insights for traffic control and policy-making.
* Build an interactive UI using Gradio for real-time predictions.
* Evolve the feature set based on initial EDA findings to enhance model performance.

3.Flowchart of the project overflow

(Include a diagram showing steps: data loading preprocessing EDA feature engineering model training evaluation deployment)

4.Data description

* Dataset Name: Traffic Accident Prediction Dataset
* Source: (Mention source, e.g., public repository or authority)
* Type of Data: Structured tabular data
* Records and Features: [Mention number of rows and columns]
* Target Variable: Accident Severity (e.g., 1, 2, 3)
* Static or Dynamic: Static dataset
* Attributes: Includes road type, weather conditions, driver behavior, vehicle type, day/time, etc.

5.Data processing

* Checked for null/missing values and imputed as necessary.
* Removed low-variance or irrelevant features.
* Categorical features encoded using one-hot or label encoding.
* Scaled numerical values using StandardScaler.
* Detected and handled outliers using IQR and z-score methods.

6.Exploratory data analysis (EDA)

* Univariate: Distribution of accident severity, time of day, weather conditions.
* Bivariate/Multivariate: Correlation heatmaps, severity by road type, weather vs severity.
* Insights:
* Accidents are more frequent and severe in poor visibility and wet road conditions.
* Night-time and weekends show a spike in severe accidents.
* Driver behavior indicators like speeding correlate with higher severity.

7.Feature Engineering

* Combined features like visibility + weather into composite risk index.
* Encoded binary categories (e.g., is\_weekend).
* Removed multicollinear features based on VIF.
* Scaled numerical features for uniformity.

8.Model building

* Algorithms: Logistic Regression, Random Forest, and XGBoost- Train-Test Split: 80:20 split with random\_state for reproducibility
* Evaluation Metrics:
* Accuracy
* Precision, Recall, F1-score
* Confusion Matrix
* ROC-AUC Score

9.Visualization of Results & Model insights

* Feature Importance: Visualized using tree-based model outputs.
* Model Comparison: Bar charts for Accuracy and F1-score across algorithms.
* Confusion Matrix: Show classification performance across severity classes.
* UI Testing: Integrated Gradio app for testing predictions with feature inputs.

10.Tools and Technologies Used

* Language: Python 3
* Environment: Jupyter / Google Colab
* Libraries: pandas, numpy, seaborn, matplotlib, scikit-learn, xgboost, Gradio

11. Team Members and Controbutions

Name Contributions

ROSHINI.S : Data preprocessing and EDA

SATHISHKUMAR.P : Feature engineering and model development

SANJAY. G : Gradio integration and UI

SANTHOSH KUMAR. K : Documentation and reporting