**Phase-1 Submission Template**

**Student Name: J HARSHINI**

**Register Number: 410723104023**

**Institution:** DHANALAKSHMI COLLEGE OF ENGINEERING

**Department:** BE COMPUTER SCIENCE AND ENGINEERING

**Date of Submission:** [28.4.2025]

# 1.Problem Statement

Enhancing road safety with AI-driven traffic accident analysis and prediction -- With the continuous rise in urbanization and vehicle usage, road traffic accidents remain a leading cause of injury and death globally. The need for innovative, data-driven solutions to mitigate these incidents is more pressing than ever. This project aims to leverage Artificial Intelligence (AI) and Machine Learning (ML) techniques to analyze historical traffic data, identify high-risk zones and conditions, and develop predictive models that can forecast the likelihood of accidents before they occur.

# 2.Objectives of the Project

So the main objectives of this road safety task is:

#### [✅ **1. In-depth Traffic Accident Analysis**

* Analyze historical traffic accident data to identify common contributing factors such as time of day, weather conditions, road types, traffic volume, and driver behavior.
* Uncover patterns and trends that reveal high-risk scenarios and accident hotspots.

#### ✅ **2. Development of Predictive Models**

* Based on real-time and environmental scenarios we going to design and train a machine learning model capable of forecasting or predicting of traffic accidents before situation.
* Evaluate multiple algorithms (e.g., decision trees, random forests, neural networks) to determine the most accurate and efficient model for accident prediction.

#### ✅ **3. Identification of High-Risk Zones**

* Use geospatial data and clustering techniques to identify accident-prone locations (black spots) or (indicating red dots).
* Generate risk heat maps that visually highlight dangerous areas within a city or region.

#### ✅ **4. Real-Time Accident Risk Assessment**

* Integrate predictive models into a dynamic system capable of assessing accident risk in real time.
* Simulate real-world traffic scenarios to validate the model’s performance under varying conditions.

#### ✅ **5. Decision Support for Authorities**

* Offer data-driven recommendations such as optimized traffic control strategies, infrastructure upgrades, and targeted law enforcement.
* Severe actions had to be taken (below 18 should not be permitted to drive any kind of transport.
* Same law for elders who are also above 70 of age.

#### ✅ **6. Public Awareness and Preventive Insights**

* Propose user-friendly interfaces or dashboards for the public to access risk alerts and safety tips.
* Safety precautions has to be executed by the government.
* Empower drivers with information that helps them make safer travel decision

Top of Form

Bottom of Form

# 3.Scope of the Project

1. **Data Collection & Integration**
   * Gather historical accident data from traffic departments, GPS logs, and open datasets.
   * Collect data including weather conditions, road types, traffic volumes, and temporal information (time of day, day of week, holidays).
   * Integrate data from multiple sources including traffic cameras, and mobile applications.
2. **Data Preprocessing & Feature Engineering**
   * Clean and normalize datasets.
   * Handle imbalanced data via oversampling, under sampling, or synthetic data generation (e.g., SMOTE).
3. **Exploratory Data Analysis (EDA)**
   * Identify accident-prone zones and high-risk conditions through statistical and visual analytics.
   * Analyze temporal and spatial accident trends.
4. **Model Development & Evaluation**
   * Develop predictive models using classification/regression techniques (e.g., Random Forest, Gradient Boosting, Neural Networks).
   * Train models to predict accident likelihood based on real-time or near real-time input.
   * Evaluate model performance using metrics like accuracy, precision, recall, F1-score, and AUC-ROC.
5. **Risk Mapping & Visualization**
   * Generate accident heat map and interactive dashboards for high-risk area identification.
   * Visualize model predictions on a GIS platform or web-based interface.
6. **Real-World Application Scenarios**
   * Provide accident risk alerts and recommendations to drivers via mobile apps or navigation systems.
   * Offer decision-support tools for urban traffic planners and safety agencies.
7. **Ethical & Privacy Considerations**
   * Ensure compliance with data privacy regulations (e.g., GDPR).
   * Incorporate fairness and transparency in model decision-making.

# 4.Data Sources

DATA SOURCE NAME: kaggle

DATASET NAME:UK road safety- traffic accidents and vehicles

LOCATION: Accident\_information.csv

FORMAT: CSV format

ACCESS METHOD: Download link

# 5.High-Level Methodology

1. Data Collection

- Gather UK Road Safety dataset from a reliable source (e.g., Kaggle)

- Ensure data is relevant, accurate, and comprehensive

2. Data Cleaning

- Handle missing values, duplicates, and inconsistencies

- Transform data into suitable formats for analysis

- Remove irrelevant or redundant data

3. Exploratory Data Analysis (EDA)

- Visualize data distributions, trends, and correlations

- Summarize data statistics and patterns

- Identify potential relationships and insights

4. Feature Engineering

- Extract relevant features from the data

- Create new features that support modeling

- Select most informative features for modeling

5. Model Building

- Choose suitable machine learning or statistical models

- Train models using the prepared data

- Tune hyper parameters for optimal performance

6. Model Evaluation

- Assess model performance using metrics (e.g., accuracy, precision, recall)

- Compare model performance and select the best model

- Identify areas for improvement

7. Visualization and Interpretation

- Visualize model results and insights

- Interpret findings and identify key takeaways

- Communicate results effectively to stakeholders

8. Deployment

- Deploy the model in a suitable environment (e.g., web app, dashboard)

- Monitor model performance and update as needed

- Ensure model is scalable and maintainable

# 6.Tools and Technologies

Data Collection

1. \*Kaggle\*: Dataset repository

2. \*API\*: Data extraction from government websites or other sources

3. \*Web scraping\*: Extracting data from websites

Data Cleaning and Preprocessing

1. \*Pandas: Data manipulation and analysis library (Python)

2. \*NumPy\*: Numerical computing library (Python)

3. \*Data cleaning tools\*: OpenRefine, Trifacta

Exploratory Data Analysis (EDA)

1. Matplotlib: Data visualization library (Python)

2. Seaborn: Data visualization library (Python)

3. Plotly: Interactive data visualization library (Python)

Feature Engineering

1. \*Scikit-learn\*: Machine learning library (Python)

2. \*Featuretools\*: Feature engineering library (Python)

Model Building

1. \*Scikit-learn\*: Machine learning library (Python)

2. \*TensorFlow\*: Deep learning library (Python)

3. \*PyTorch\*: Deep learning library (Python)

Model Evaluation

1. \*Scikit-learn\*: Model evaluation metrics (Python)

2. \*Metrics\*: Accuracy, precision, recall, F1-score

Visualization and Interpretation

1. \*Matplotlib\*: Data visualization library (Python)

2. \*Seaborn\*: Data visualization library (Python)

3. \*Tableau\*: Data visualization tool

Deployment

1. \*Flask\*: Web framework (Python)

2. \*Django\*: Web framework (Python)

3. \*Docker\*: Containerization platform

# 7.Team Members and Roles

|  |  |  |  |
| --- | --- | --- | --- |
| SN NO | NAMES | ROLES | RESPONSIBILITY |
| 1. | DEEPALAKSHMI.A | TEAM MEMBER | DATA COLLECTION &DATA PREPROCESSING |
| 2. | ISHWARYA.S | TEAM MEMBER | EXPLORATORY DATA ANALYSIS &FEATURE ENGINEERING |
| 3. | HARSHINI.J | TEAM MEMBER | MODEL BUILDING &MODEL EVALUATION |
| 4. | LOGAPRIYA | TEAM LEADER | VISUALISATION INTERPRETION,DEPLOYMENT |