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

**Phase-2**

**Student Name:** HARSHINI J

**Register Number:** 410723104023

**Institution:** DHANALAKSHMI COLLEGE OF ENGINEERING

**Department:** BE COMPUTER SCIENCE & ENGINEERING

**Date of Submission:** 07-05-2025

**Github Repository Link[: https://github.com/Harshini-jk/NM\_Harshini](C:\\Users\\janub\\AppData\\Local\\Microsoft\\Windows\\INetCache\\IE\\TYDCAE49\\__{EDC978D6-4D53-4B2F-A265-5805674BE568}_phase_2_(Harshini)[1]_docx_thumb[1].png)**

### **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.

### **Project Objectives**

✅ 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

### **Flowchart of the Project Workflow**

LOAD DATASET

PREPROCESS DATASET

DATASET PREPROCESSING FOR TRAINING

TRAIN DATASET

TEST DATASET

CONSTRUCT TRAIN MODEL

PREDICT ACCIDENTS

RESULT SUGGESTION

### 

### **4.Data Description**

Here's a possible data description for the project:

Historical Traffic Accident Data:

1. Accident Location: GPS coordinates, street address, intersection, or nearest landmark

2. Time and Date: Timestamp of the accident, including date, time, and day of the week

3. Accident Severity: Severity level (e.g., fatal, injury, property damage)

4. Cause of Accident: Primary cause (e.g., speeding, reckless driving, weather conditions)

5. Number of Vehicles Involved: Number of vehicles involved in the accident

6. Number of Injuries/Fatalities: Number of people injured or killed

Real-time Traffic Data:

1. Traffic Volume: Current traffic volume (number of vehicles per hour)

2. Speed: Average speed of traffic

3. Weather Conditions: Current weather conditions (rain, fog, clear, etc.)

4. Road Conditions: Current road conditions (potholes, construction, etc.)

5. Traffic Signal Status: Status of traffic signals (red, yellow, green)

Additional Data:

1. Road Network Data: Road geometry, lane configuration, and traffic signal locations

2. Demographic Data: Population density, age distribution, and other relevant demographic information

This data will be used to train machine learning models to predict accident likelihood and identify high-risk areas.

### **5. Data Preprocessing**

Data Preprocessing:

1. Handling missing values

2. Data normalization

3. Feature engineering (extracting relevant features)

4. Encoding categorical variables

5. Data transformation (converting data types)

6. Removing duplicates and outliers.

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

*1. Visualize accident distribution: Heatmaps, scatter plots to identify high-risk areas*

*2. Analyze accident patterns: Time-series analysis to identify trends and seasonality*

*3. Examine correlations: Correlation analysis between variables (e.g., weather, traffic volume)*

*4. Identify hotspots: Identify locations with high accident frequency/severity*

*5. Summary statistics: Calculate mean, median, and standard deviation for accident-related variables.*

### 

### **7. Feature Engineering**

Feature Engineering:

1. Extract time-based features (hour, day, month)

2. Create location-based features (latitude, longitude, proximity to intersections)

3. Calculate traffic-related features (traffic volume, speed)

4. Encode categorical variables (weather, road conditions)

5. Derive accident severity features (injury/fatality rates)

### **8. Model Building**

Model Building:

1. Select algorithms: Choose suitable machine learning algorithms (e.g., logistic regression, decision trees, random forest, neural networks)

2. Train models: Train models using the training dataset

3. Hyperparameter tuning: Optimize model hyperparameters for better performance

4. Model evaluation: Evaluate model performance using metrics (accuracy, precision, recall, F1-score)

5. Model selection: Select the best-performing model for deployment.

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

1. Accident hotspots: Visualize high-risk areas on a map

2. Feature importance: Show feature importance scores to understand key factors contributing to accidents

3. Prediction accuracy: Plot model performance metrics (accuracy, precision, recall)

4. Confusion matrix: Visualize true positives, false positives, true negatives, and false negatives

5. ROC-AUC curve: Plot the Receiver Operating Characteristic curve to evaluate model performance.

### **10. Tools and Technologies Used**

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\*: Open Refine, Tri fact

Exploratory Data Analysis (EDA)

1. MATPLOTLIB: Data visualization library (Python)

2. SEABORN: Data visualization library (Python)

3. PLOTY: Interactive data visualization library (Python)

Feature Engineering

1. SCIKIT - learn: Machine learning library (Python)

2. Feature tools: Feature engineering library (Python)

Model Building

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

2. \*Tensor Flow\*: 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

### **11. Team Members and Contributions**

|  |  |  |  |
| --- | --- | --- | --- |
| 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 |