****

# **Assessment Report on Predicting Traffic Congestion**

## **Submitted as partial fulfillment for the award of**

**BACHELOR OF TECHNOLOGY DEGREE**

**SESSION 2024-25**

**CSE(AI)**

**By**

**Name:** Ayushman

**Roll No: 2428CSEAI1102**

**Under the supervision of**

**Sir Shivansh Prasad**

**Institution Name:** K.I.E.T

May, 2025

## **1. Introduction**

With increasing urbanization, traffic congestion is becoming a major challenge for cities worldwide. This project aims to predict **traffic congestion levels** using **sensor data** such as vehicle speed, sensor count, and time of day. By leveraging **machine learning techniques**, transportation authorities can make data-driven decisions to improve traffic flow, reduce bottlenecks, and enhance commuter experiences.

## **2. Problem Statement**

To classify road sections as **High, Medium, or Low congestion** based on real-time **traffic sensor data**. This classification assists city planners in optimizing road networks, improving traffic management, and reducing travel time inefficiencies.

## **3. Objectives**

* **Process sensor data** for effective machine learning modeling.
* **Train a Random Forest classifier** to predict congestion levels.
* **Evaluate model accuracy** using confusion matrix, precision, recall, and F1-score.
* **Visualize traffic congestion trends** through data insights and graphical analysis.

## **4. Methodology**

### **Data Collection:**

The dataset consists of real-time traffic observations, including:

✔ **Sensor Count** - Number of traffic sensors detecting vehicle movement.

✔ **Average Speed** - Speed of vehicles at different times of the day.

✔ **Time of Day** - Categorical variable (morning, afternoon, evening, night).

✔ **Congestion Level** - Target classification (High, Medium, Low).

### **Data Preprocessing:**

✔ Handling missing values using **mean imputation**.

✔ Encoding categorical features (**Time of Day, Congestion Level**) using **LabelEncoder**.

✔ Splitting data into **training (80%)** and **testing (20%)**.

### **Model Building:**

✔ Using **Random Forest Classifier** due to its high accuracy and feature importance.

✔ Training model with optimized hyperparameters (n\_estimators=200, max\_depth=10).

### **Model Evaluation:**

✔ Assessing model **accuracy, precision, recall, and F1-score**.

✔ Generating a **confusion matrix heatmap** to analyze classification errors.

## **5. Data Preprocessing**

* **Handling missing values** in numerical columns using **mean imputation**.
* **Encoding categorical values** in **Time of Day** using **LabelEncoder**.
* **Normalizing data** for consistent feature scaling.
* **Splitting dataset** into **train-test ratio of 80:20** for model training.

## **6. Model Implementation**

The **Random Forest classifier** is chosen for its ability to handle mixed data types and improve classification accuracy. The trained model predicts congestion levels based on sensor inputs.

## **7. Evaluation Metrics**

* **Accuracy**: Percentage of correct congestion predictions.
* **Precision**: How well high congestion instances were correctly predicted.
* **Recall**: How well actual high congestion cases were identified.
* **F1 Score**: Balances precision and recall effectively.
* **Confusion Matrix**: Visual representation of model classification errors.

## **8. Results and Analysis**

✔ The trained **Random Forest model** achieved an accuracy of **~85%**, showing strong predictive capability.

✔ **Heatmap visualization** revealed occasional misclassifications between medium and high congestion.

✔ **Precision and recall scores** indicated reliable congestion detection performance.

## **9. Conclusion**

This project successfully implemented an **AI-based congestion prediction system**, demonstrating its potential for **traffic optimization and urban planning**. Future enhancements could incorporate **weather conditions, road construction data, and advanced deep learning models** to further improve accuracy.

## **10. References**

* **scikit-learn documentation** for model implementation.
* **pandas & seaborn** for data preprocessing and visualization.
* Research papers on **smart city traffic analytics**.







