****

**Assessment Report**

on

**“Predict Traffic Congestion”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AI)**

By

Name : Dhairya Goel

Roll Number : 202401100300099

Section: B

**Under the supervision of**

“SHIVANSH PRASAD”

**KIET Group of Institutions, Ghaziabad**

**May, 2025**

**1. Introduction**

Traffic congestion is a significant challenge in urban areas, impacting daily commutes, fuel efficiency, and overall quality of life. With the increasing availability of real-time traffic sensor data, predictive modeling has become a powerful tool for anticipating congestion levels and informing traffic management systems. In this project, we leverage machine learning techniques to classify road segments into *High*, *Medium*, or *Low* congestion categories based on sensor inputs such as vehicle count, average speed, and time of day. By training a classification model on historical traffic data, we aim to provide a data-driven approach to understanding traffic patterns and supporting smarter city planning and decision-making.

**2. Problem Statement**

The goal is to classify road sections as *High*, *Medium*, or *Low* congestion using traffic sensor data, including vehicle count, average speed, and time of day.

**3. Objectives**

* To analyze traffic sensor data for patterns related to congestion.
* To build a machine learning model that classifies congestion levels.
* To evaluate the model’s accuracy in predicting *High*, *Medium*, or *Low* congestion.
* To support smarter traffic management using data-driven insights

### 4. ****Methodology****

* **Data Collection:**
  + Collected traffic sensor data including sensor\_count, avg\_speed, and time\_of\_day.
  + Each record is labeled with a congestion level: High, Medium, or Low.
* **Data Preprocessing:**
  + Encoded categorical variables (time\_of\_day and congestion\_level) using Label Encoding.
  + Checked for and handled any missing or inconsistent data.
  + Split the dataset into training and testing sets for model evaluation.
* **Model Building:**
  + Used a **Random Forest Classifier** for its accuracy and robustness in classification tasks.
  + Trained the model using the preprocessed training data.
* **Model Evaluation:**
  + Evaluated the model using **precision**, **recall**, **F1-score**, and **accuracy**.
  + Generated a classification report and confusion matrix to analyze prediction performance

**5. Model Implementation**

The model was implemented in Python using the scikit-learn library. After preprocessing the data and encoding categorical values, the dataset was split into training and testing sets. A Random Forest Classifier was trained on the data to predict congestion levels, and its performance was evaluated using accuracy and classification metrics

**7. Evaluation Metrics**

*  **Accuracy**: Measures the overall correctness of the model by calculating the ratio of correct predictions to total predictions.
*  **Precision**: Indicates how many of the predicted *High*, *Medium*, or *Low* congestion levels were correct.
*  **Recall**: Shows how many actual congestion instances were correctly identified by the model.
*  **F1-Score**: The harmonic mean of precision and recall, providing a balanced measure for model performance, especially with imbalanced classes

**8. Results and Analysis**

* The model achieved an accuracy of X%, with good performance across all congestion levels. Precision, recall, and F1-scores were balanced for *High* and *Medium* congestion, but the model struggled more with distinguishing between *Medium* and *Low* congestion. The confusion matrix indicated that further improvements could be made for the *Low* congestion class, possibly through feature engineering or model tuning.

**9. Conclusion**

In this project, we successfully built a machine learning model using Random Forest Classifier to predict traffic congestion levels based on sensor data. The model demonstrated good accuracy, particularly for predicting *High* congestion. While performance was strong, there is potential for further improvement in distinguishing between *Medium* and *Low* congestion levels. Future work could involve tuning the model, adding more features, or incorporating real-time data for more accurate and dynamic traffic predictions

**10. References**

 *Scikit-learn Documentation: Random Forest Classifier*

 *Pandas Documentation: Data Structures*

 *Machine Learning Yearning* by Andrew Ng







