****

**Assessment Report**

on

**“Vechicle Emission Prediction”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AIML)**

By

Name : Anushka Jaiswal

Roll Number : 202401100400044

Section: A

**Under the supervision of**

“BIKKI KUMAR”

**KIET Group of Institutions, Ghaziabad**

**May, 2025**

**1. Introduction**

As environmental regulations become stricter, it is essential to classify vehicles based on their engine and fuel characteristics to ensure compliance. This project aims to predict a vehicle's **emission category** using machine learning by analyzing features such as engine size, fuel type, and CO₂ emissions.

**2. Problem Statement**

To build a machine learning model that classifies vehicles into **emission categories (A, B, C, etc.)** based on engine and fuel-related attributes. This can aid manufacturers and regulators in emission compliance checks.

**3. Objectives**

* Analyze and preprocess the given vehicle dataset.
* Encode categorical features and scale numeric values if needed.
* Train a Random Forest classifier for emission prediction.
* Evaluate model performance using standard classification metrics.

**4. Methodology**

**Data Collection**: The dataset is provided as a CSV file with 100 entries, each representing a vehicle with:

* Engine size
* Fuel type
* CO₂ emissions
* Emission category (target)
* **Data Preprocessing**:  
  + **Categorical Encoding**: fuel\_type is label encoded.
  + **Feature Selection**: Used engine size, fuel type, and CO₂ emissions.
  + **Train-Test Split**: 80% for training, 20% for testing.
* **Model Building**:  
  + **Algorithm Used**: Random Forest Classifier
  + Trained the model on the training data using Scikit-learn.
* **Model Evaluation**:  
  + Evaluating accuracy, precision, recall, and F1-score.
  + Classification report.

**5. Data Preprocessing**

The dataset had no missing values. The fuel\_type column (categorical) was encoded using LabelEncoder. The dataset was then split for training and evaluation.

**6. Model Implementation**

Random Forest Classifier was chosen due to its high accuracy. It builds multiple decision trees and uses voting to make final predictions. The model was trained on 80 samples and tested on 20 samples

**7. Evaluation Metrics**

The following metrics are used to evaluate the model:

| **Metric** | **Description** |
| --- | --- |
| **Accuracy** | Percentage of correct predictions |
| **Precision** | Correctness of predicted categories |
| **Recall** | Ability to detect all correct categories |
| **F1 Score** | Harmonic mean of precision and recall |

**8. Results and Analysis**

* The model showed good performance on the test set with high accuracy.
* The classification report showed balanced precision and recall for each category.
* The Random Forest model was effective in distinguishing emission categories based on engine and fuel features.

**9. Conclusion**

This project successfully implemented a Random Forest classifier to predict vehicle emission categories. The model can help in automating emissions classification tasks and can be further improved with more data and advanced algorithms.

**10. References**

* scikit-learn documentation
* pandas documentation
* Vechile emission dataset
* Ensemble Learning Techniques



