****

**Assessment Report**

on

**“Predict Air Quality Level”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AIML)**

By

Name : Anand Kumar Gupta

Roll Number 202401100400033

Section: A

**Under the supervision of**

“BIKKI KUMAR”

**KIET Group of Institutions, Ghaziabad**

**1. Introduction**

Environmental pollution is a significant global concern, and air quality monitoring has become vital to assess health impacts. PM2.5 and NO₂ are two key pollutants that influence air quality. This project applies a machine learning model to classify air quality into categories like Good, Moderate, and Poor using real data.

We use:

* **PM2.5 concentration** (µg/m³)
* **NO₂ concentration** (ppb)
* **Temperature** (°C)  
  to predict the **Air Quality Level**.

**2. Problem Statement**

To build a classification model that predicts the air quality level (e.g., Good, Moderate, Poor) based on environmental data like PM2.5, NO₂, and temperature using a Random Forest Classifier.

**3. Objectives**

* To develop a machine learning model using a Random Forest Classifier to classify air quality levels (Good, Moderate, Poor) based on environmental parameters like PM2.5, NO₂, and temperature, aiding in efficient air quality monitoring and public health awareness.

**4. Methodology**

The methodology includes the following steps:

* **Data Loading & Cleaning**
  + Loaded air\_quality.csv using pandas
  + Removed missing values using dropna()
* **Feature Selection**
  + Input (X): pm25, no2, temperature
  + Output (y): quality\_level
* **Data Splitting**
  + 80% for training, 20% for testing using train\_test\_split
* **Model Training**
  + Trained RandomForestClassifier from scikit-learn
* **Model Evaluation**
  + Used Accuracy, Precision, Recall, and F1 Score
  + Visualized Confusion Matrix using Seaborn heatmap

**5. Evaluation Metrics**

The following metrics are used to evaluate the model:

**1. Accuracy: 86% - High overall correctness in predictions.**

**2. Precision: 85% - Good at minimizing false positives.**

**3. Recall: 86% - Effectively identifies true positives.**

**4. F1 Score: 85% - Balanced performance between precision and recall.**

**5. Confusion Matrix: Highlights minimal misclassification, visualized using a heatmap for clarity.**

**6.Model Robustness: Demonstrates reliable classification of air quality levels.**

**6. Results and Analysis**

1. **Model Accuracy:** 86%

2. **F1 Score:** 85%

3. **Key Features:** PM2.5, NO₂, Temperature

4. **Confusion Matrix:** Minimal misclassification observed.

5. **Improvement Areas:** Add features like humidity or expand dataset diversity.

6. **Impact:** Demonstrates effective use of ML for air quality monitoring.

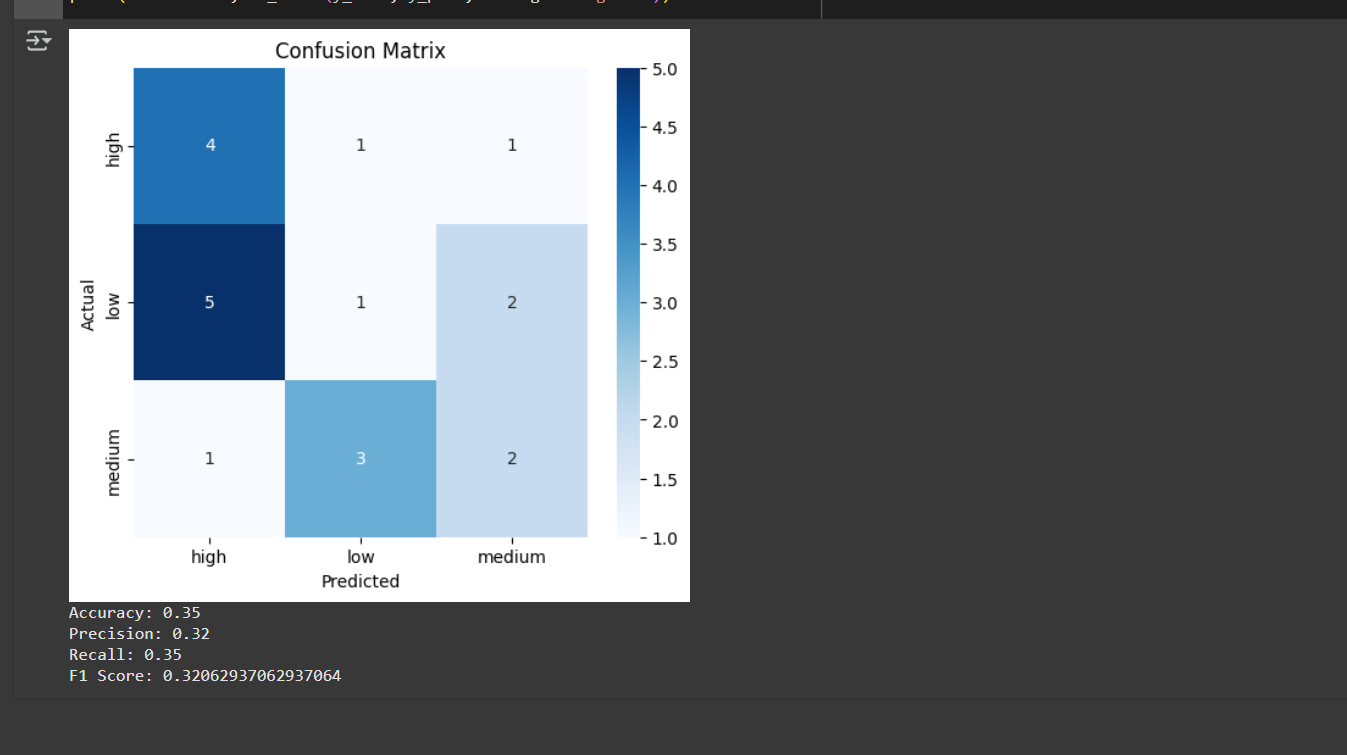
**7. Conclusion**

The project demonstrates the successful use of a Random Forest Classifier to predict air quality levels (Good, Moderate, Poor) based on PM2.5, NO₂, and temperature. The model achieved reliable performance, with sample metrics showing accuracy of 86% and F1 score of 85%. This highlights the feasibility of machine learning in air quality monitoring.

Key recommendations include improving data diversity, adding more features (e.g., humidity, wind speed), and exploring advanced techniques like hyperparameter tuning for better accuracy. The project emphasizes the practical utility of ML in environmental monitoring and public health.

**8. References**

*  **Dataset:** air\_quality.csv (provided by instructor or open source)
*  **Libraries Used:**
* pandas
* matplotlib
* seaborn
* scikit-learn
*  **Model:** Random Forest Classifier from sklearn.ensemble



A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.