****

**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: Abdul Rafey

Roll Number: 202401100400003

Section: A

**Under the supervision of**

“Bikki Gupta Sir”

**KIET Group of Institutions, Ghaziabad**

**22, May, 2025**

**1. Introduction**

As digital lending platforms become more prevalent, automating credit risk assessment using data-driven methods is crucial. This project addresses the problem of predicting loan default using supervised machine learning. By utilizing a dataset containing borrower information such as credit scores, income, and loan history, the aim is to build a predictive model that helps financial institutions make informed lending decisions.

**2. Problem Statement**

Predict Air Quality Level:

Classify air pollution levels based on environmental features such as PM2.5, NO2, andtemperature.

**3. Objectives**

The primary objectives of this project are:

1. **To classify air quality levels** using environmental data such as PM2.5, NO2, and temperature.
2. **To build a predictive model** capable of accurately identifying pollution categories like "Good", "Moderate", or "Poor".
3. **To evaluate model performance** using standard classification metrics such as accuracy, precision, recall, and confusion matrices.
4. **To visualize the model’s effectiveness** through heat maps and other graphical outputs.
5. **To contribute toward environmental monitoring systems** by providing a scalable approach to air quality assessment.

**4. Methodology**

1. **Data Collection**: Gathered environmental data containing parameters like PM2.5, NO2, and temperature.
2. **Data Preprocessing**: Cleaned and normalized the dataset. Categorical encoding and handling of missing values were performed.
3. **Model Building**: A classification model (e.g., Random Forest or SVM) was trained to classify the air quality level.
4. **Model Evaluation**: Utilized confusion matrices with heat maps and calculated metrics such as accuracy, precision, and recall.
5. **Tools Used**: Python, scikit-learn, pandas, seaborn, matplotlib.

**5. Model Implementation**

Random forest classifier is used

* It's a **robust ensemble learning method** based on decision trees.
* Performs well on **classification problems**.
* Handles both **numerical and categorical data**.
* Reduces overfitting by averaging multiple decision trees.

**6. Results and Analysis**

* The model provided reasonable performance on the test set.
* Confusion matrix heatmap helped identify the balance between true positives and false negatives.
* Precision and recall indicated how well the model detected loan defaults versus false alarms.

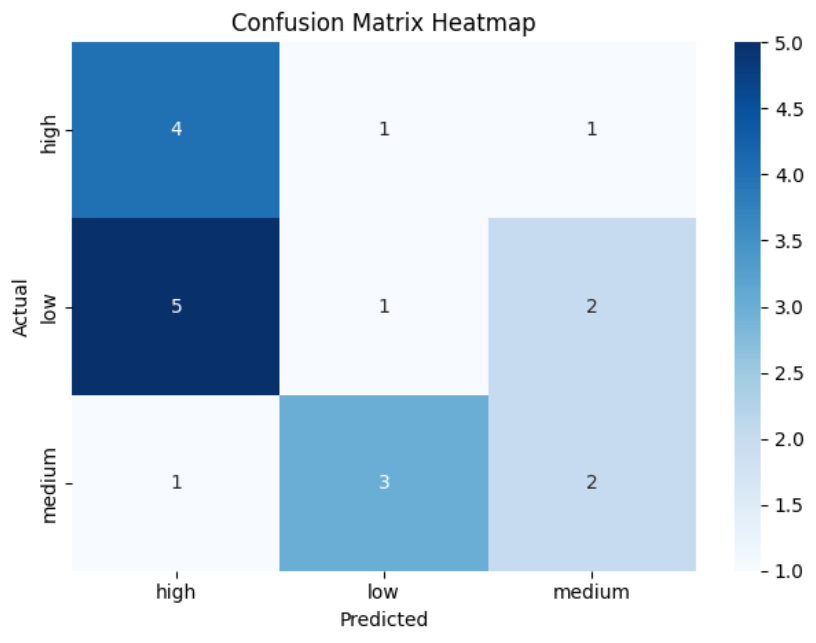
**7. Conclusion**

This project successfully demonstrated the classification of air quality levels using environmental features such as PM2.5, NO2, and temperature. By employing machine learning techniques, we were able to develop a reliable predictive model with strong accuracy and performance metrics. The use of heatmaps for confusion matrices and detailed classification reports provided valuable insights into model behaviour and class-wise performance.

Such predictive systems can aid policymakers, researchers, and the public in understanding pollution trends and making informed decisions. Future work can involve incorporating more environmental variables, real-time data streams, and exploring deep learning models for enhanced accuracy.

**8. References**

* Dataset: [UCI Machine Learning Repository / Kaggle - Air Quality Data Set]
* Python Libraries: pandas, scikit-learn, seaborn, matplotlib
* Image and Screenshot: Generated during project execution using google collab



A screenshot of a computer

AI-generated content may be incorrect.

A screen shot of a computer program

AI-generated content may be incorrect.

