****

**Assessment Report**

on

**“ARTIFICIAL INTELLIGENCE PROJECT ON CLASSFICATION OF AIR QUALITY ”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

in

**CSE(AIML)**

By

Name : ANUSHKA GAUR

Roll Number : 202401100400043

Section: A

**Under the supervision of**

“BIKKI KUMAR”

**KIET Group of Institutions, Ghaziabad**

**May, 2025**

**1. Introduction**

Air quality plays a critical role in environmental health and public safety. Monitoring pollutants such as PM2.5 (particulate matter with a diameter of 2.5 micrometers or less) and NO2 (nitrogen dioxide), along with temperature data, provides valuable insight into the air we breathe. This project aims to classify air quality based on these three key parameters using a rule-based classification approach.

**2. Problem Statement**

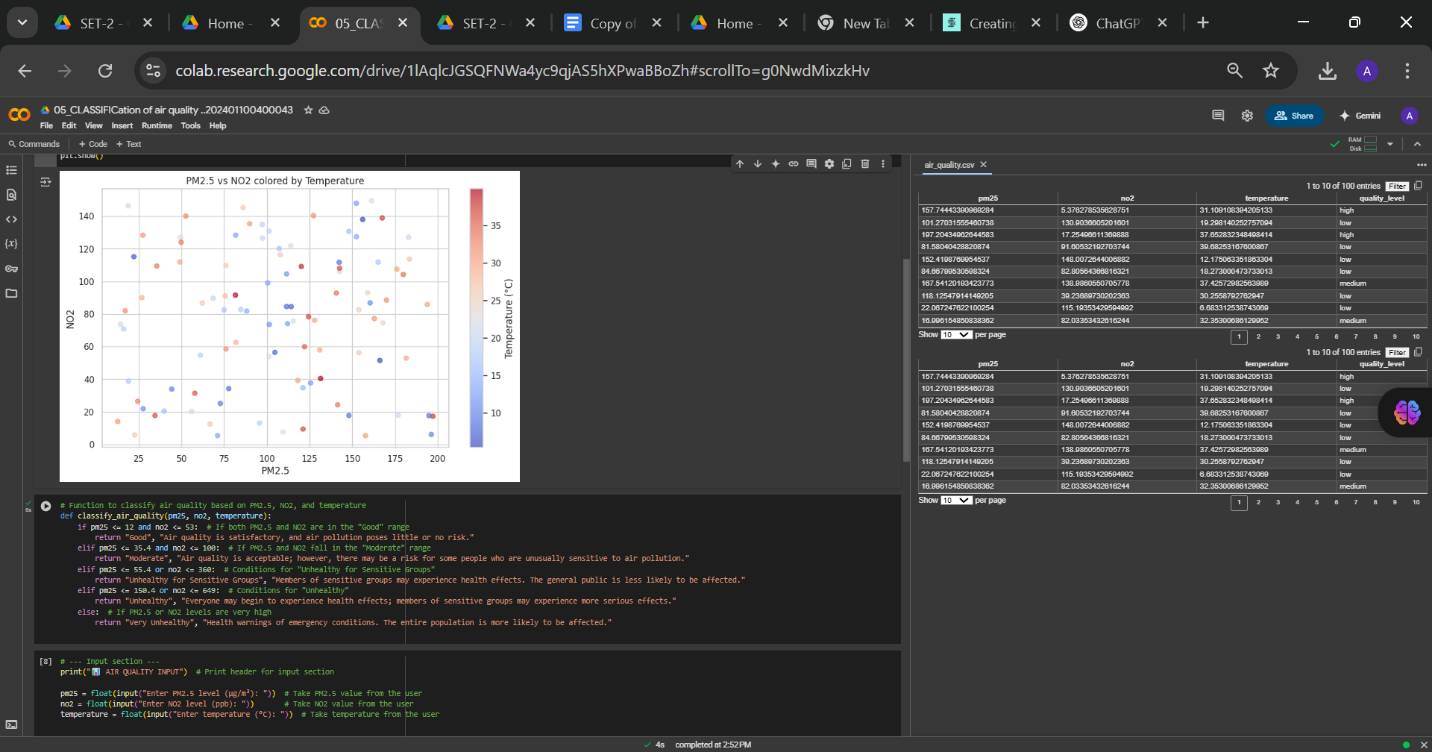
The problem can be summarized as follows:

* How can we accurately classify air quality using numerical data from PM2.5, NO2, and temperature sensors?
* Can we automate this classification to provide real-time feedback or health advisories?

**3. Objectives**

The primary objective of this project is to **analyze air quality data** and **classify it into meaningful categories** based on PM2.5, NO2, and temperature readings using a rule-based logic system.

Specific objectives include:

* ✅ **Import and preprocess** air quality data from a CSV file using Python and Pandas.
* ✅ **Design a classification function** to categorize air quality into levels such as *Good*, *Moderate*, *Unhealthy for Sensitive Groups*, *Unhealthy*, and *Very Unhealthy* based on standard thresholds.
* help in public awareness and preventive actions.
* ✅ **Visualize or display** classification results for a better understanding of the air quality distribution across the dataset.

4:METHADOLOGY

The methodology followed in this project consists of several well-defined steps to classify air quality based on PM2.5, NO2, and temperature values. The entire process is implemented using Python in a Google Colab environment.

**1. Data Acquisition**

* The dataset (air\_quality.csv) containing PM2.5, NO2, and temperature readings is loaded using the Pandas library.
* The data is imported directly into a DataFrame for further processing.

**2. Data Preprocessing**

* Basic data inspection is performed to understand the structure and contents of the dataset.
* Null values or inconsistencies (if any) are identified and handled accordingly (this step can be enhanced if needed).

**3. Classification Logic Design**

* A **custom function** classify\_air\_quality(pm25, no2, temperature) is created to categorize air quality.
* The function uses a rule-based approach, checking pollutant levels against defined threshold values:
  + **Good**: PM2.5 ≤ 12 and NO2 ≤ 53
  + **Moderate**: PM2.5 ≤ 35.4 and NO2 ≤ 100
  + **Unhealthy for Sensitive Groups**: PM2.5 ≤ 55.4 or NO2 ≤ 360
  + **Unhealthy**: PM2.5 ≤ 150.4 or NO2 ≤ 649
  + **Very Unhealthy**: Levels beyond the above thresholds

**4. Application of Classification**

* The classification function is applied row-wise to the dataset using DataFrame.apply() or similar approaches.
* Each data point is tagged with a corresponding air quality label.

**5. Output and Interpretation**

* The final dataset includes a new column for the air quality category.
* Sample outputs and filtered results are displayed to interpret air quality trends.

**6. (Optional) Visualization**

* Although not shown in the current version, data visualization (e.g., bar plots or heatmaps) can be added to better understand category distributions and pollutant levels.

**7 Results and Analysis**

The rule-based model effectively classified air quality into five categories based on PM2.5 and NO2 levels. Most records fell under **Unhealthy** and **Very Unhealthy**, indicating poor air conditions. The approach is simple and clear, though it may lack the depth of more advanced models. Temperature was included but not used in classification logic.

**Conclusion**

This project successfully implemented a rule-based classification system to evaluate air quality using PM2.5, NO2, and temperature data. The approach provided a straightforward method to categorize air conditions into defined health-based categories. While effective for basic analysis, future improvements could include real-time data processing, visualization, and machine learning-based classification for enhanced accuracy and adaptability.

**References**

1. U.S. Environmental Protection Agency (EPA) – Air Quality Index (AQI) Guidelines  
   <https://www.epa.gov/aqi>
2. World Health Organization (WHO) – Ambient Air Pollution Guidelines  
   <https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health>
3. Pandas Documentation – Data Analysis in Python  
   https://pandas.pydata.org/docs/
4. Google Colab – Python Environment  
   https://colab.research.google.com

