# IMPACT OF AIR QUALITY ON HUMAN HEALTH

## A Machine Learning Approach

VEONA MAINA

# TABLE OF CONTENT

# OVERVIEW

## PURPOSE

To provide an understanding of how air quality impacts human health using data.

## GOAL

To build a predictive model that helps categorize air quality levels and their corresponding health effects.

## WHY IT MATTERS

To with accurate predictions, we can take preventive measures to protect public health, prioritize interventions, and inform policy decisions.

# BUSINESS UNDERSTANDING

**The Problem:** Air pollution contributes to serious health problems, particularly in susceptible populations, and increases healthcare costs.

**Why It Matters:**

- *Human Health:* Poor air quality contributes to respiratory and cardiovascular illness.
- *Economic Consequences:* Medical expenses and loss of productivity.
- *Policy Decisions:* Accurate predictions help further the development of public health interventions and regulations.
-

**The objective:** To develop a model that will predict health impacts from air quality to allow taking the appropriate risk-reduction measures based on proactive insights, and support evidence-based decision-making.

# DATA UNDERSTANDING

**Dataset Overview:**

Used air quality and meteorological data to predict the health impact of air pollution.

***key indicators :***

- Air Quality Index (AQI), Particulate Matter (PM2.5, PM10),Temperature, Humidity, and Wind Speed

***Health Impact Score:*** This score represents the calculated impact of air quality on human health based on the various environmental factors.

***Health Impact Label:*** The health impact score is categorized into five labels:

- Very Low, Low, Moderate, High, Very High

***Goal:*** Accurately predict these health impact labels to inform public health decisions and interventions.

# MODELLING

**In the modeling phase, we use advanced algorithms to analyze and categorize air quality data. Our goal is to predict the impact of air quality on human health by classifying it into distinct categories. This process involves training the model with historical data and refining it to ensure accurate and reliable predictions.**

## PROCESS STEPS

01

### INPUT DATA

Collect and preprocess data including air quality indicators and meteorological factors.

02

### MODEL

Apply classification algorithms to categorize air quality impacts

03

### OUTPUT CATEGORIES

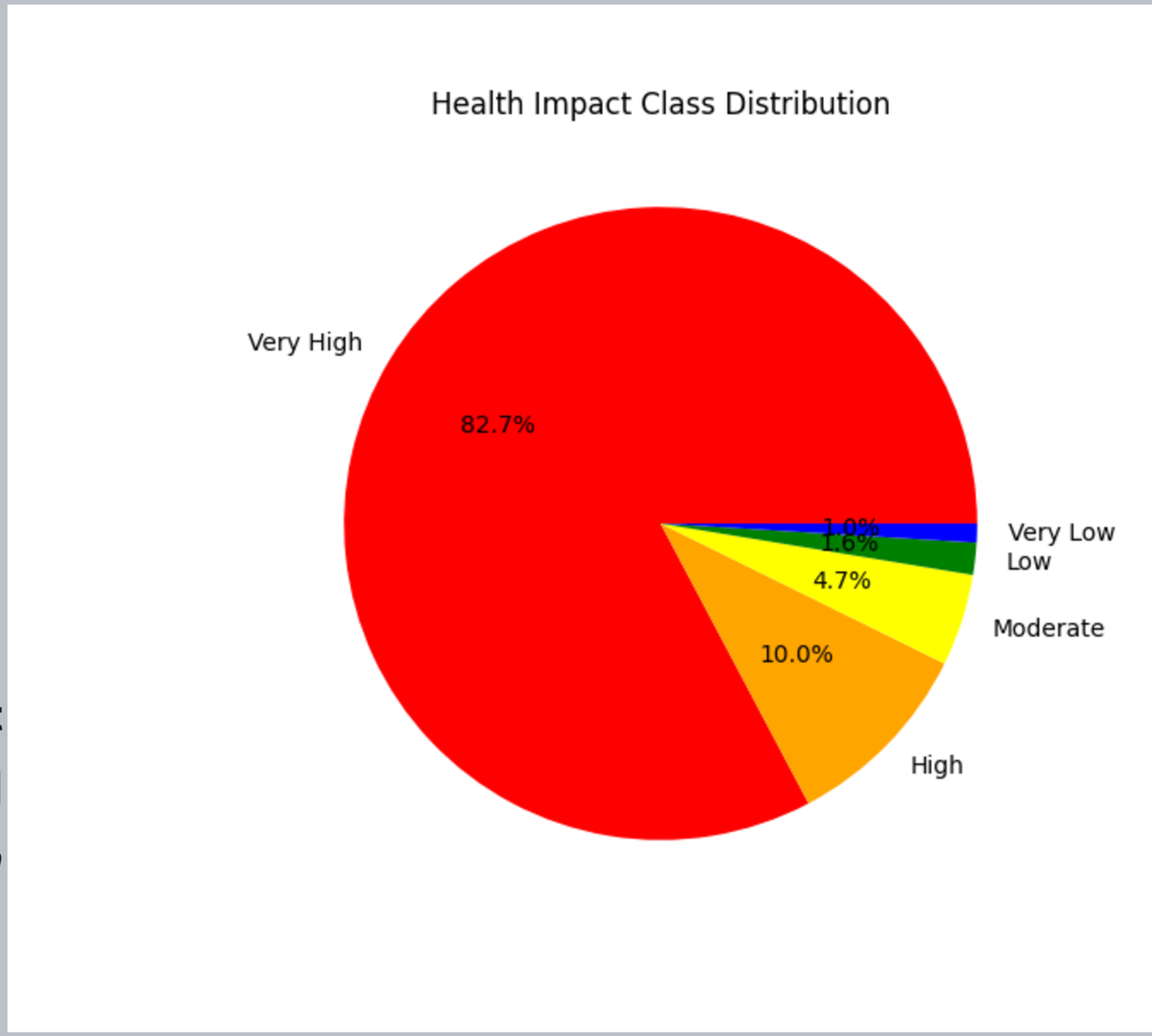
Generate predictions and classify air quality into impact levels: Very Low, Low, Moderate, High, Very High.

# MODELLING

## Why Classification Modeling?

- Classification Models help predict categories.

- In this case, we want to classify the impact of air quality on health into predefined levels. i.e *(Very low, Low, Moderate, High and Very High)*



# MODELLING

- **Model Selection:**

Tested two main models:

  - ***Logistic Regression:*** A simpler model but struggled with the complexity of the data.
  - ***Random Forest:*** A more advanced model that better handled the data and produced more accurate results.

Key Focus:

Ensured the model could effectively differentiate between various health impact categories, even with class imbalances (i.e., some health impact categories were much less frequent than others).

# MODEL EVALUATION

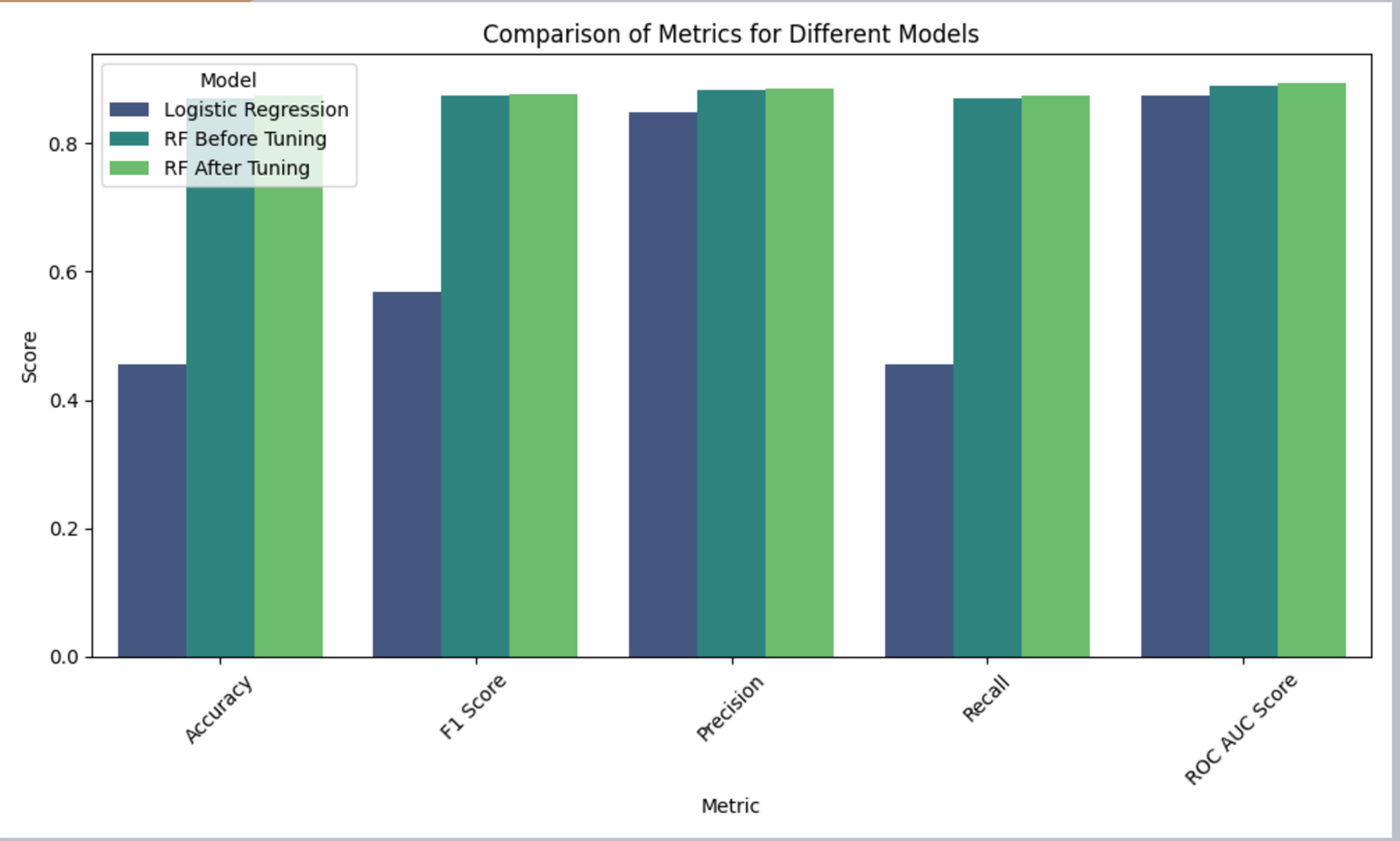
**Performance metrics used to evaluate the two models;**

- ***Accuracy:*** The percentage of correct predictions overall. It shows how often the model's predictions are correct.
- ***F1 Score:*** It’s useful for understanding how well the model handles each class.
- ***Precision:*** It shows how many of the predicted positive cases are truly positive.
- ***Recall:*** It indicates how well the model identifies positive instances.
- ***ROC AUC Score:*** Shows the model’s ability to distinguish between classes. Higher values indicate better performance.

*Model Tuning: The process of adjusting a model's parameters to enhance its performance. By testing various settings, tuning helps identify the optimal configuration to improve accuracy and other metrics, ensuring the model performs at its best.*

# MODEL EVALUATION

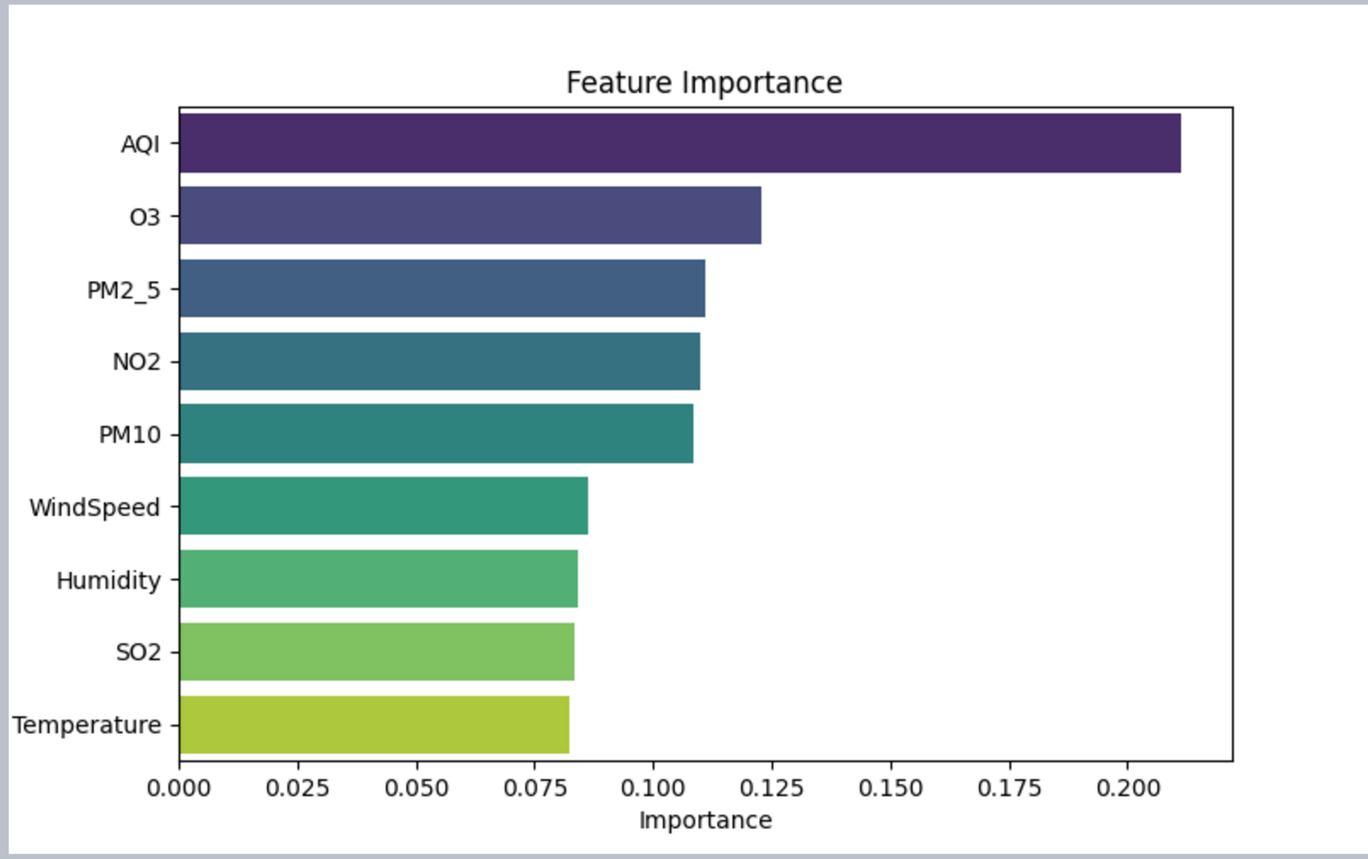
This comparison highlights that the Random Forest model(RF), even before tuning, delivers the best overall results for predicting air quality's impact on health.



# MODEL EVALUATION

**Top Features Driving Predictions:**

- ***Air Quality Index (AQI):*** Higher AQI values indicate more severe health risks.
- ***PM2.5 and PM10:*** Particulate matter that can penetrate deep into the lungs, contributing significantly to health problems.
- ***Meteorological Factors:*** Elements like Temperature, Wind Speed, and Humidity play secondary roles but still contribute to the model's predictions.



# RECOMMENDATIONS

- ***Prioritize Air Quality Monitoring:***

Focus on key air quality indicators such as AQI, PM2.5, and PM10.These factors are most critical in predicting health impacts and should be monitored closely to address potential risks effectively.

- ***Leverage Predictive Insights:***

Utilize the Random Forest model’s predictions to guide health advisories and public awareness campaigns.The model provides reliable insights into how different levels of air quality may affect human health, aiding in proactive health management.

# RECOMMENDATIONS

- ***Invest in Public Health Initiatives:***

Use insights from the model to support policies and programs aimed at improving air quality and protecting public health.Data-driven actions can lead to more effective public health interventions and better health outcomes.

# NEXT STEPS

- ***Monitor and Validate Model Performance:***

The model’s performance will be continuously tracked and updated with new data, to maintain accuracy and relevance in predictions over time.

- ***Explore Additional Data Sources:***

Investigate the incorporation of new datasets that could provide further insights, to enhance the model’s predictive power and coverage.

- ***Refine Model Features and Parameters:***

Continuously analyze the impact of different features and adjust model parameters as needed, to improve the model’s performance and adaptability by ensuring it captures the most relevant factors affecting air quality and health.

# THANK YOU

VEONA MAINA