**Phase-1 Submission**

**Student Name:** Semmozhiyan N S

**Register Number:** 712523104057

**Institution:** PPG Institute of Technology

**Department:** B.E. Computer Science and Engineering

**Date of Submission:** 25/04/2025

**1.Problem Statement**

Predicting air quality levels using advanced machine learning algorithms for environmental insights

1. **Objectives of the Project**
   1. To develop an accurate air quality prediction model
   2. To perform comprehensive data preprocessing and integration
   3. To identify and evaluate key factors affecting air quality
   4. To enable near real-time forecasting capabilities
   5. To design and implement an interactive user interface
   6. To promote environmental awareness through data visualization and insights
   7. To deploy the predictive model and interface for real-world application
2. **Scope of the Project**
   1. **Features to Be Developed or Analyzed**
      1. Analysis of key pollutants such as PM2.5, PM10, CO, NO₂, SO₂, and O₃ that directly influence the Air Quality Index (AQI).
      2. Inclusion of temperature, humidity, wind speed, and atmospheric pressure to examine their correlation with air pollution levels.
      3. Identification of temporal trends, such as daily, monthly, and seasonal variations in air quality.
      4. Ability to input new data and generate predicted AQI levels dynamically
      5. Charts and graphs to enhance user understanding and promote awareness.
   2. **Limitations and Constraints**
      1. Missing or Incomplete Data
      2. The project focuses on machine learning algorithms and does not include deep learning-based time series forecasting (e.g., LSTM), due to time and resource constraints
      3. Generalization Issues
      4. Depending on server and internet speed, there might be slight delays in real-time AQI prediction and response time.
      5. The project will be implemented using Python-based tools (Pandas, Streamlit), with limited scope for integration of advanced IoT or edge devices.

**4.Data Sources**

*Kaggle*

***API:*** *OpenAQ*

**5.High-Level Methodology**

* **Data Collection** –
  + **OpenAQ API** for global and real-time air quality data
  + **Meteorological APIs** (OpenWeatherMap) to obtain temperature, humidity, wind speed, and pressure data
* **Data Cleaning** –
  + Imputation techniques (mean, median) will be applied for missing values
  + Duplicate rows will be removed
  + Data types and formats will be standardized for all features
* **Exploratory Data Analysis (EDA)** –
  + **Time series plots** to visualize seasonal or hourly trends in AQI
  + **Histogram and boxplots** to observe value distributions and outliers
* **Feature Engineering** –
  + **New features** such as pollutant ratios, moving averages, and lag variables may be created
  + **Temporal features** like hour of the day, day of the week, or month will be extracted from timestamps.
* **Model Building** –
  + **Linear Regression** – As a baseline model
  + **Random Forest Regressor** – Robust to outliers and useful for feature importance
* **Model Evaluation** –
  + **Root Mean Squared Error (RMSE)** – Penalizes large errors
  + **R-squared (R²)** – Measures how well the model explains the variance in AQI
* **Visualization & Interpretation** –
  + **Dashboards and interactive charts** will display AQI trends and predictions
  + **Feature importance plots** will help explain model behavior
* **Deployment** –
  + **HTML/CSS + JS** for the frontend interface
  + AWS-
    - EC2-hosting
    - S3-Storage

**6.Tools and Technologies**

* **Programming Language** –
  + - Python
    - HTML
    - CSS
    - JavaScript
* **Notebook/IDE** – Visual Studio Code
* **Libraries** –
  + - Numpy
    - Pandas
    - Matplotlib
    - Flask
    - Streamlit
* **Tools for Deployment** – Docker

**7.Team Members and Roles**

|  |  |  |
| --- | --- | --- |
| Name | Role | Description |
| Sivabalan V | Project Manager | Leads and manages the projectOversees EDA and interprets insights Coordinates all team activities and deliverables |
| Dhyanesh V | Backend & Deployment Developer | Builds backend API (Flask/FastAPI) Integrates model and deploys the application on a cloud platform |
| Semmozhiyan NS | Machine Learning Engineer | Trains, tunes, and evaluates prediction models Selects the best-performing model for deployment |
| Sri Sabarish U | Data Collection & Preprocessing Lead | Sources and integrates AQI/weather data Handles data cleaning, preprocessing,. |
| Chandru M | Frontend Developer + Documentation Lead | Designs the UI and handles user interaction Prepares project documentation, reports, and presentation |