Pollution Awareness - AI/ML Project

# Overview

This project is part of my AI/ML Internship (Environmental Monitoring & Pollution Control). The goal is to analyze environmental data and build a machine learning model that predicts pollution levels, helping raise awareness and support sustainability.

# Learning Objectives

• Collect and preprocess environmental data

• Apply Machine Learning techniques for predicting pollution levels

• Evaluate the trained model using accuracy and classification metrics

• Visualize results (confusion matrix, feature importance)

• Build awareness through data-driven insights

# Tools and Technology Used

• Programming Language: Python

• Libraries: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Joblib

• Environment: Jupyter Notebook, VS Code

• Version Control: GitHub

• Dataset: Environmental pollution data (Temperature, Humidity, PM2.5, PM10, SO2, NO2, CO, Ozone)

# Methodology

• Data Collection → Environmental dataset with key pollutants

• Data Preprocessing → Handling missing values, encoding, scaling

• Model Training → Random Forest Classifier

• Evaluation → Accuracy, Classification Report, Confusion Matrix

• Visualization → Feature importance & results analysis

# Problem Statement

Rising air pollution levels pose serious health and environmental risks. There is a need to predict pollution levels and create awareness using data-driven AI/ML techniques.

# Solution

• Collected sample dataset of environmental parameters

• Built a Machine Learning model to classify pollution levels into Low, Moderate, High

• Achieved high accuracy (~90%) with Random Forest Classifier

• Visualized pollution factors’ importance (PM2.5, PM10, SO2, etc.)

• Results can support awareness campaigns and decision-making

# Results

The model was trained and evaluated on environmental pollution data. Accuracy was ~90%, and results showed PM2.5, PM10, and SO2 as major contributors to pollution levels. Confusion matrix and feature importance visualizations were generated to support the findings.

# Conclusion

Successfully built an end-to-end ML pipeline for pollution awareness. Random Forest achieved strong prediction performance. The project demonstrates how AI/ML can support environmental monitoring & awareness. Next step: Deploy model as a web app for public use.