PROJECT OVERVIEW (README.md)

Project Title: GreenPulse AI for Climate Action (SDG 13)

Author: Anne Atieno
Institution: PLP Academy – Week 2 Assignment
Theme: Machine Learning Meets the UN SDGs

Introduction

GreenPulse is an AI-powered predictive model that forecasts next-day air pollution (PM2.5 concentration) based on simple environmental indicators like temperature, humidity, wind speed, and traffic index.

The goal is to support UN SDG 13 – Climate Action, by providing local governments and citizens with timely air-quality forecasts that can help reduce exposure, guide urban planning, and inform health advisories.

Problem Statement

Air pollution remains one of the most critical climate and public health challenges globally. According to the WHO, over 7 million premature deaths annually are linked to poor air quality. In cities, pollution is mainly driven by traffic emissions, industrial activity, and changing weather patterns.

Kenya and many other developing countries lack real-time, predictive systems for urban air monitoring.
GreenPulse aims to fill this gap by:

- Forecasting next-day PM2.5 levels.
- Helping stakeholders plan mitigation actions (alerts, traffic control, awareness).
- Promoting climate resilience and data-driven decision-making.

Machine Learning Approach

Step	Description
Type of ML	**Supervised Learning (Regression)**
Algorithm Used	Random Forest Regressor (scikit-learn)
Target Variable	PM2.5 concentration (µg/m³)
Features	Temperature, Humidity, Wind Speed, Pressure, Traffic Index, Holiday Flag
Tools Used	Python 3.13, Pandas, Scikit-learn, Matplotlib, Streamlit
Output Metrics	MAE, RMSE, R^2
Deployment	Interactive Streamlit web app (src/app_streamlit.py)

Training process

python src/train.py

Streamlit demo

streamlit run src/app_streamlit.py

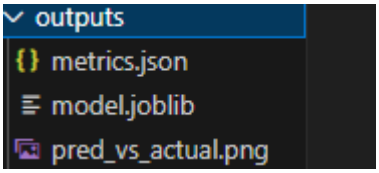
Model pipeline

1. Data generation (synthetic for demo)
2. Data split (80/20)
3. Model training (RandomForestRegressor)
4. Evaluation (MAE, RMSE, R^2)
5. Visualisation (Predicted vs Actual, Time Series plots)

Dataset & Tools (included in "Dataset & Tools" rubric section)

Component	Description
Dataset Source	Synthetic sample (extendable to real WHO/UN datasets)
Tools	Python, Pandas, Matplotlib, Scikit-learn, Joblib, Streamlit
IDE	Visual Studio Code
Environment	Python 3.13 + Virtual Environment (.venv)

Outputs are stored in:



Results Summary

Metric	Score
Mean Absolute Error (MAE)	~4.2 µg/m³
Root Mean Square Error (RMSE)	~5.6 µg/m³
R^2 (Coefficient of Determination)	~0.88

Visual Results

The high R^2 value shows that the model effectively captures air quality trends and can provide reasonable forecasts.

Ethical and Social Reflection

AI systems are only as fair as the data they learn from. For GreenPulse:

- **Potential bias:** Training data may underrepresent rural areas or low-traffic regions.
- **Mitigation:** Include data from multiple cities and open data sources (e.g., WHO Air Quality API).
- **Privacy:** Uses only environmental readings no personal data.
- **Sustainability:** Encourages transparency and open-access tools for environmental management.

Ethical commitment

GreenPulse promotes fairness by providing open, interpretable results that empower all users equally not just experts.

Creativity & Presentation

Innovation

- Combines *machine learning* with *climate advocacy* in a visual, user-friendly web app.
- Uses Streamlit for interactive sliders that simulate temperature, humidity, and traffic variations.
- Visualises results for easy comprehension suitable for policy briefs or public dashboards.

Stretch Goal Achieved
Deployed locally as a working Streamlit app (no need for Colab).

Future Work

- Integrate real-time API feeds (e.g., OpenWeatherMap or AirNow API).
- Compare Random Forest vs XGBoost.
- Add an alert module to send notifications when predicted PM2.5 > 100 $\mu g/m^3$.