10Pearls Shine Internship Program – Final Report

Project Title: AQI Forecast Dashboard

Domain: Data Science

Submitted by: Warda Iftikhar

Abstract:

This project focuses on developing an Air Quality Index (AQI) Forecast Dashboard that predicts air quality levels using machine learning. The system was built using Python, Flask, and React, integrating data visualization for real-time insights. Historical air quality data was analyzed to train the model, *which forecasts future AQI values to help monitor environmental conditions. The dashboard provides users with an interactive interface to view predictions and pollutant trends. The final model was deployed on Railway using CI/CD automation, ensuring smooth and continuous updates.*

Introduction:

Air pollution is one of the most critical environmental challenges impacting public health and sustainability. To address this issue, predictive modeling of air quality has become essential for enabling early awareness and informed decision-making. This project, titled AQI Forecast Dashboard, aims to forecast the Air Quality Index (AQI) using machine learning techniques. The system integrates data preprocessing, model training, and interactive visualization to provide accurate and interpretable predictions of air quality levels. Developed using Python, Flask, and React, the dashboard offers real-time insights through an intuitive interface and has been successfully deployed on the Railway platform for public accessibility.

Methodology:

The AQI Forecast Dashboard project followed a complete machine learning pipeline from data collection to deployment, integrating feature engineering, model training, explainability, and visualization to provide accurate and actionable air quality predictions. Key steps include:

1. **Data Collection:** Retrieved 6 months of hourly air quality (PM2.5, PM10, CO, NO_2, SO_2, O_3) and weather data (temperature, humidity, wind, pressure) using Open-Meteo APIs to build a comprehensive historical dataset.
2. **Data Preprocessing:** Cleaned and prepared the dataset by handling missing values and outliers. Computed Air Quality Index (AQI) for each timestamp and engineered additional features such as time-based features (hour, day, month, weekday), lagged values, rolling averages, and AQI change rates to enhance model performance.
3. **Feature Store:** Leveraged Feast to store processed features, enabling efficient retrieval for model training and real-time predictions. Materialized features to both online and offline stores for seamless integration.
4. **Model Training:** Implemented and evaluated Random Forest and XGBoost regression models using processed features. Models were assessed using MSE, RMSE, MAE, and R^2 Score, with Random Forest achieving high accuracy and stable predictions.
5. **Model Explainability:** Applied SHAP (SHapley Additive exPlanations) to determine the contribution of each pollutant and feature, providing transparency and insight into model decisions. Visualizations highlighted key drivers of AQI.
6. **Dashboard Development:** Developed a full-stack application using Flask (backend API) and React (frontend) to display real-time AQI forecasts, pollutant trends, and historical data, with interactive charts and gauges for user-friendly visualization.
7. **Deployment:** Deployed the dashboard on Railway, providing accessible, real-time predictions and ensuring uptime for continuous monitoring.

8. **CI/CD Pipeline**: Automated data updates, model retraining, and deployment using GitHub Actions, ensuring that the pipeline runs seamlessly without manual intervention and that predictions remain up-to-date.

Findings / Results:

The AQI Forecast Dashboard achieved accurate and interpretable predictions by combining robust data preprocessing, feature engineering, and model training. Key outcomes include:

- **Model Comparison:** Random Forest and XGBoost were trained and evaluated. Random Forest showed better stability, while XGBoost had slightly lower error in some intervals.

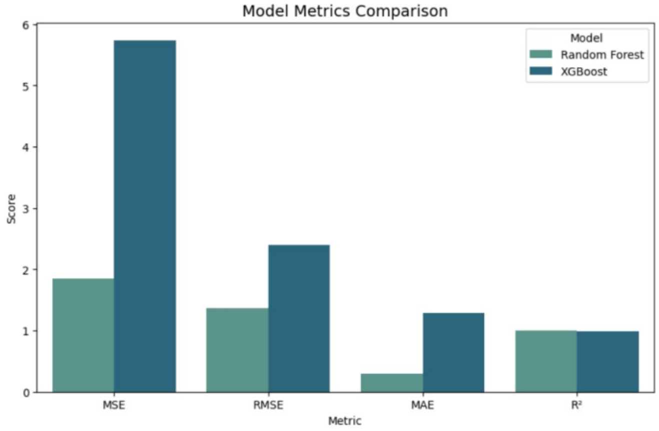
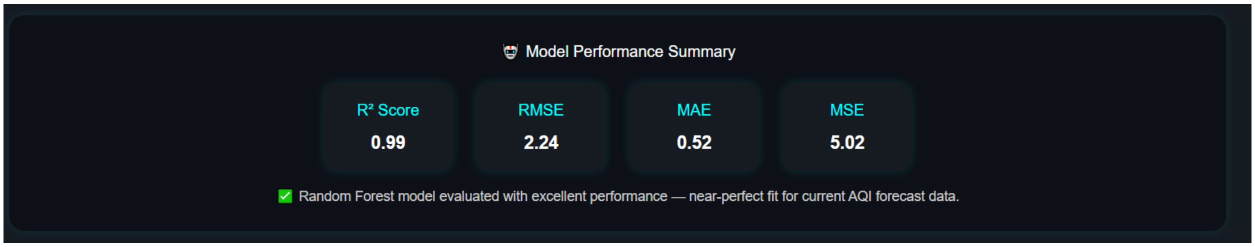


Figure 1: Model metrics comparison



Model Performance Summary

R² Score	RMSE	MAE	MSE
0.99	2.24	0.52	5.02

✅ Random Forest model evaluated with excellent performance — near-perfect fit for current AQI forecast data.

Figure 2: Random Forest Model Summary

- **Feature Importance (SHAP):** SHAP analysis revealed the most influential pollutants and features affecting AQI predictions, helping understand the model's decisions.

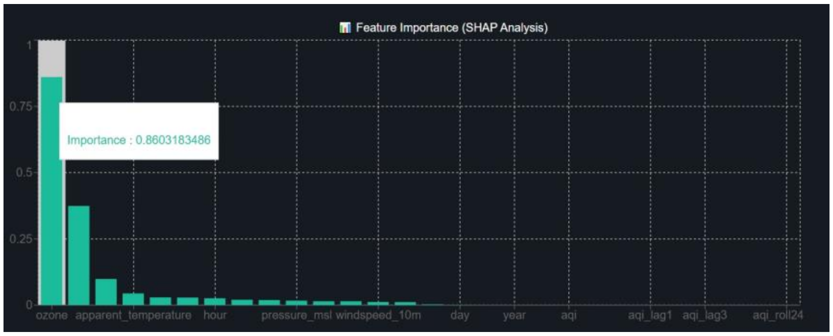


Figure 3: SHAP analysis on dashboard

- **Correlation Analysis:** Examined relationships between pollutants and weather variables, ensuring meaningful features were used in modeling.

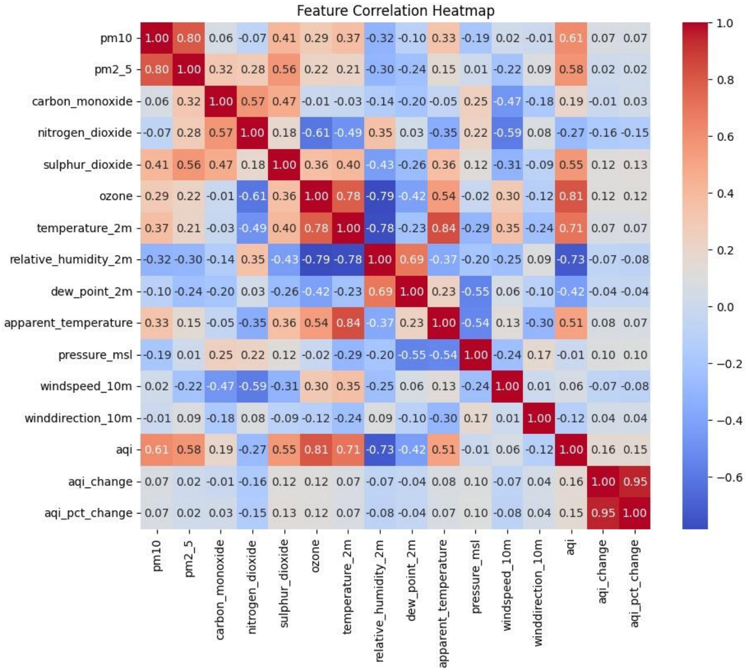


Figure 4: Correlation heatmap of pollutants

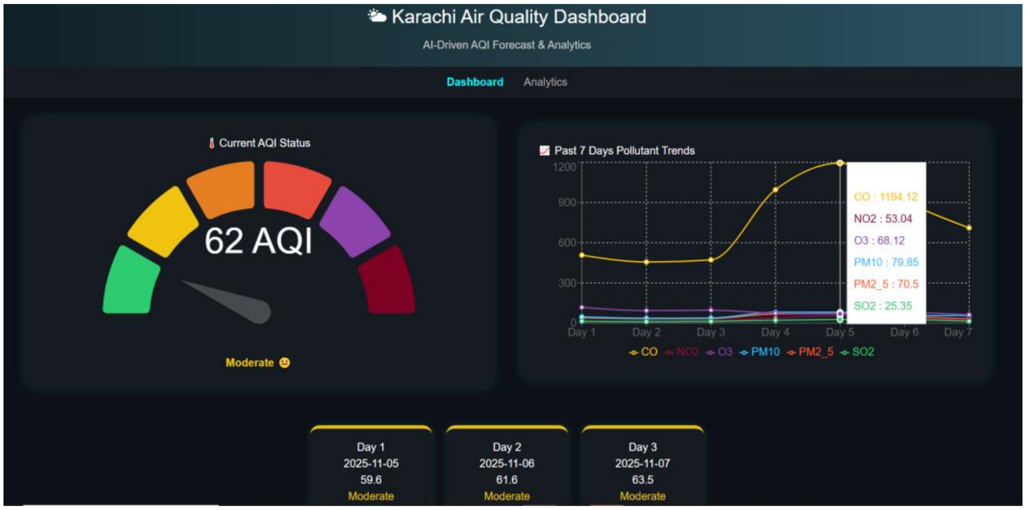
- **Dashboard Insights:** The interactive dashboard visualizes real-time AQI forecasts, pollutant trends, and historical patterns, providing actionable insights for end-users.
- **Deployment & Accessibility:** Deployed on Railway for real-time access. CI/CD via GitHub Actions ensures automatic updates of the data pipeline and model.

Conclusion:

The AQI Forecast Dashboard project successfully integrated data collection, preprocessing, feature engineering, model training, and deployment to provide accurate and actionable air quality predictions. By leveraging Random Forest and XGBoost regressors, and interpreting model outputs with SHAP, the project highlights the most influential pollutants affecting AQI. The interactive dashboard, built with Flask and React, offers end-users clear insights into real-time and historical air quality trends.

Deployment on Railway ensures accessibility, while CI/CD via GitHub Actions allows seamless updates. This project demonstrates a practical application of data science in environmental monitoring, combining predictive modeling, explainability, and user-centric visualization.

Final Dashboard:



Karachi Air Quality Dashboard

AI-Driven AQI Forecast & Analytics

Day 1
2025-11-05
59.6
Moderate

Day 2
2025-11-06
61.6
Moderate

Day 3
2025-11-07
63.5
Moderate

Deployment Link:

Access the AQI Forecast Dashboard : aqi-forecast-dashboard-production.up.railway.app