Project Title: AQI System for Karachi

Deployed at: Karachi AQI Dashboard

Contents

1. Project Overview 2
2. Data Pipelines 2
2.1 Feature Pipeline (Hourly) 2
2.2 Training Pipeline (Daily) 3
2.3 Prediction Pipeline 3
3. Dashboard Features 4
3.1 Header 4
3.2 AQI Overview Cards 4
3.3 Weather Conditions 4
3.4 AQI Forecast Visualization 4
3.5 Pollutant Concentration & Gauge 4
3.6 Top Model Comparison 4
3.7 AQI Protection Tips 4
4. Feature Analysis & Insights 5
4.1 Time-series Analysis 5
4.2 Hourly & Weekly Patterns 5
4.3 Pollutant Analysis 5
4.4 SHAP-based Feature Importance 5
5. Technical Stack 5
6. Automation & Deployment 6
6.1 GitHub Actions 6
6.2 Deployment 6
7. Strengths of the Project 6
8. Summary 6
9. Screenshots ..... **Error! Bookmark not defined.**

1. Project Overview

The project provides a real-time and forecasted Air Quality Index (AQI) dashboard for Karachi, integrating pollutant data, weather features, and predictive modeling. It leverages MongoDB for data storage, Python ML pipelines for modeling, and Streamlit for interactive visualization.

Key goals:

- Display **current AQI**, **pollutant concentrations**, and **weather metrics**.
- Predict **next 3-day AQI** using a machine learning model.
- Rank **top-performing models** with RMSE, MAE, and R^2.
- Provide actionable **AQI protection tips** to the public.
- Maintain **automated daily training** and **hourly feature updates** via GitHub Actions.

2. Data Pipelines

2.1 Feature Pipeline (Hourly)

Purpose: Fetch latest pollutants and weather features for AQI prediction.
Trigger: Every hour (cron: 0 * * * *)

Steps:

1. Connect to MongoDB (training_features collection).
2. Fetch historical pollutant data from **OpenWeatherMap API** for the last 3 months.
3. Fetch historical weather data from **Open-Meteo API**.
4. Merge pollutant and weather data to create **training-ready features**:
    - Time features: hour, day, month, day_of_week
    - AQI change: aqi_change = aqi.diff()
5. Store merged data into training_features collection.

Outcome: Ensures the model always has updated **pollutants + weather features** for predictions.

2.2 Training Pipeline (Daily)

Purpose: Train AQI prediction models daily and store top-performing models.
Trigger: Daily at 2 AM UTC

Steps:

1. Load **merged feature data** from MongoDB.
2. Split data using **time-series-aware train/test split** (no shuffle).
3. Train multiple models:
    - Ridge Regression
    - Random Forest
    - Gradient Boosting
    - XGBoost (strongest model)
4. Evaluate models using: RMSE, MAE, R^2, and cross-validated R^2.
5. Store **top 2 models** in MongoDB (model_registry) along with pickled binaries and performance metrics.

Outcome: Daily refreshed models ensure predictions remain **accurate and current**.

2.3 Prediction Pipeline

Purpose: Generate real-time AQI forecasts.
Steps:

1. Fetch the **best model** from model_registry.
2. Fetch **latest weather forecast** for 4 days from Open-Meteo API.
3. Repeat latest pollutant values for all forecast hours.
4. Predict **PM2.5 → AQI** using US EPA formula.
5. Classify AQI into categories: Good, Moderate, Unhealthy, etc.
6. Store predictions in MongoDB (predictions) and current weather in weather.

Outcome: Provides **up-to-date AQI forecasts** for the next 3 days.

3. Dashboard Features

The Streamlit dashboard provides:

3.1 Header

- City: Karachi
- Dashboard title: **Karachi AQI Dashboard**
- Last updated date automatically from latest predictions

3.2 AQI Overview Cards

- **Today's AQI** with category & emoji
- **Next 3-day AQI forecast** with cards and hover effects
- Visual design: Dark theme + gradient color + neon hover effects

3.3 Weather Conditions

- Current **temperature, humidity, pressure**
- Automatically updated from latest forecast

3.4 AQI Forecast Visualization

- **Bar chart:** Next 3-day AQI predictions
- **Map:** Karachi city with latest AQI marker

3.5 Pollutant Concentration & Gauge

- **Pie chart:** Shows distribution of pollutants (pm2_5, pm10, no2, so2, co, o3, nh3)
- **Gauge chart:** Real-time AQI indicator with color-coded ranges

3.6 Top Model Comparison

- Shows **RMSE, MAE, R^2** of top models
- Highlights **best model** with red pattern
- Slim bars for compact comparison

3.7 AQI Protection Tips

- Actionable tips like:
    - Wear masks on high AQI days
    - Avoid outdoor activity
    - Use air purifiers
    - Stay hydrated

4. Feature Analysis & Insights

4.1 Time-series Analysis

- Trend, seasonality, and residuals visualized using **seasonal decomposition**
- 24-hour **rolling mean** for AQI stability

4.2 Hourly & Weekly Patterns

- Average AQI per **hour of day** and **weekday**
- Helps identify **peak pollution periods**

4.3 Pollutant Analysis

- Boxplots & histograms for major pollutants
- Correlation heatmap to understand feature importance

4.4 SHAP-based Feature Importance

- SHAP values from the **best model** highlight which features contribute most to AQI predictions
- Interactive plots and force plots for **explainable AI**

5. Technical Stack

Layer	Technology / Library
Dashboard UI	Streamlit, Plotly, HTML/CSS
Data Storage	MongoDB Atlas
ML Modeling	Scikit-learn (Ridge, RF, GB), XGBoost
Feature Engineering	Pandas, Numpy
Forecast API	Open-Meteo (weather), OpenWeatherMap (pollution)
Pipeline Automation	GitHub Actions (hourly + daily triggers)
Visualization	Plotly, SHAP, Matplotlib, Seaborn
Environment	Python 3.10, dotenv, requests, pymongo

6. Automation & Deployment

6.1 GitHub Actions

- **Hourly feature pipeline:** Keeps training_features up-to-date
- **Daily training pipeline:** Retrains models automatically
- **Manual triggers:** Both pipelines can be manually triggered

6.2 Deployment

- Deployed on **Streamlit Cloud**
- Fully responsive layout
- Dark mode dashboard with **interactive charts, hover effects, and glow styles**

7. Strengths of the Project

- Fully **automated ML pipeline** with daily training and prediction
- Real-time and forecasted AQI for public awareness
- **Interactive visualization** (charts, map, gauges)
- Explainable AI with **SHAP feature importance**
- **Data-driven insights** with hourly and weekly trends
- Top-performing model highlighted for transparency

8. Summary

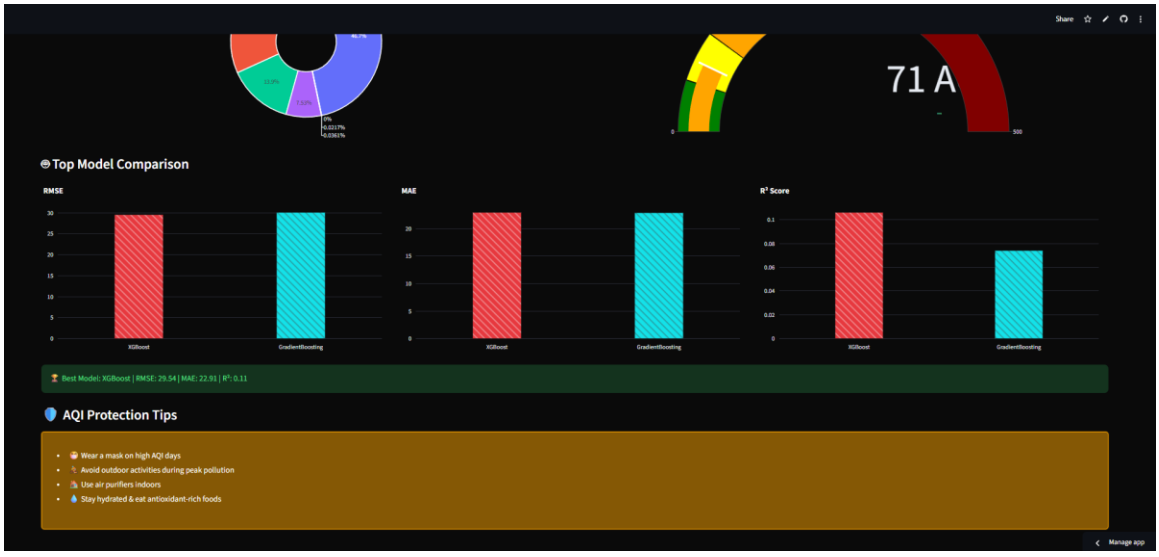
This project demonstrates a **full-stack AQI monitoring system** integrating:

- **Data ingestion** (pollutants & weather)
- **Feature engineering** (time features, merging datasets)
- **ML modeling** (multiple models, automated training, top model selection)
- **Prediction pipelines** (real-time AQI forecast)
- **Interactive visualization** (dashboard with charts, map, tips, gauges)
- **Explainable AI** (SHAP)
- **Automation & deployment** (GitHub Actions + Streamlit Cloud)

9. Dashboard


Karachi AQI Dashboard
AQI Overview
Today
71
Moderate
2026-02-16
88
Moderate
2026-02-17
73
Moderate
2026-02-18
86
Moderate
Weather Conditions
Temp (°C)
20.8
Humidity (%)
92.0
Pressure (hPa)
1014.6
AQI Forecast & Map
AQI Forecast (Bar View)
Predicted AQI
Time
Pollutant Concentration & AQI Gauge


Pollutant Concentration & AQI Gauge
Pollutant Concentration
CO
O3
PM10
PM2_5
SO2
NO2
NH3
Current AQI
71 A
Top Model Comparison
RMSE
MAE
R² Score
XGBoost
GradientBoosting
Best Model: XGBoost | RMSE: 29.54 | MAE: 22.91 | R²: 0.11


71 A
Top Model Comparison
RMSE
MAE
R² Score
XGBoost
GradientBoosting
Best Model: XGBoost | RMSE: 29.54 | MAE: 22.91 | R²: 0.11
AQI Protection Tips
Wear a mask on high AQI days
Avoid outdoor activities during peak pollution
Use air purifiers indoors
Stay hydrated & eat antioxidant-rich foods