# Karachi AQI Prediction System – final Report

## ✓ Project Overview:

This project predicts the **Air Quality Index (AQI)** for Karachi, Pakistan using historical data and real-time readings from the **World Air Quality Index (WAQI) API**.
It automates:

- Daily AQI prediction
- 3-day forecasting
- AQI change rate calculation
- Model retraining via CI/CD pipelines

## ✓ Data Sources:

- **Historical Data:** Stored in historical_aqi.csv (past 15–20 days pollutant readings and AQI values)
- **Real-time Data:** Fetched from **WAQI API** using AQI_API_TOKEN
- **Features Used:** PM2.5, PM10, $O_3$, CO, $NO_2$, $SO_2$, AQI Change Rate

## ✓ Feature Engineering:

- ◆ Extracts pollutant levels from API responses
- ◆ Calculates **AQI Change Rate** for trend tracking
- ◆ Handles missing pollutants (CO, $NO_2$, $SO_2$) with placeholder values
- ◆ Normalizes and cleans data before training

## ✓ Machine Learning Model:

- **Algorithm Used:** RandomForestRegressor
- **Reason for Choice:**

  - ◆ Handles non-linear relationships well
  - ◆ Works with small datasets without heavy preprocessing

# Karachi AQI Prediction System – final Report

- **Training Process:**

  - ◆ Loads features_karachi.csv
  - ◆ Splits into train/test sets
  - ◆ Trains Random Forest model
  - ◆ Saves model as karachi_aqi_model.pkl

## ✓ Prediction & Forecasting:

- **Daily Prediction:**

  - ◆ predict_today.py runs every 3 hours to fetch latest AQI data and store predictions in daily_predictions.csv

- **3-Day Forecasting:**

  - ◆ aqi_forecast.yml generates daily forecasts for the next 3 days and stores them in forecast_3day.csv

- **Dashboard:**

  - ◆ app.py (Streamlit) displays predictions, forecasts, and pollutant trends

## ✓ CI/CD Automation:

The project uses **GitHub Actions** for automation with three workflow files:

- ◆ aqi_pipeline.yml – Retrains the machine learning model daily
- ◆ predict.yml – Runs every 3 hours to make current-day predictions
- ◆ aqi_forecast.yml – Runs daily to generate 3-day AQI forecasts

**Platform:** GitHub Actions (Windows runner)
**Security:** Uses GitHub Secrets (AQI_API_TOKEN, GH_PAT) for authentication

## ✓ Results & Visualizations:

- ◆ Predicted AQI trends stored in daily_predictions.csv
- ◆ Forecast graphs and pollutant trends displayed in dashboard

# Karachi AQI Prediction System – final Report

◆ **Sample Output:**

**Date      AQI**

**2025-08-14  152**

**2025-08-15  147**

**2025-08-16  165**

## Conclusion & Future Work

The **Karachi AQI Prediction System** successfully automates AQI prediction and forecasting using historical and real-time data.
**Future Enhancements:**

◆ Add more pollutants and weather parameters

◆ Improve model accuracy using advanced algorithms like Gradient Boosting

◆ Implement AQI alert system for dangerous levels