10Pearls

Air Quality Index (AQI) Prediction System – Project Report

Submitted By: Ayesha Jamil

1. Project Overview

Air pollution has become a major concern globally, impacting public health and the environment. This project aims to predict the Air Quality Index (AQI) for Peshawar using machine learning techniques. The system leverages historical and real-time weather and pollution data, providing accurate AQI forecasts and actionable insights for decision-making.

Key objectives:

- Collect and store AQI-related environmental data.
- Train and evaluate machine learning models for AQI prediction.
- Identify the best model and explain predictions using SHAP.
- Provide a user-friendly interface for real-time AQI monitoring.

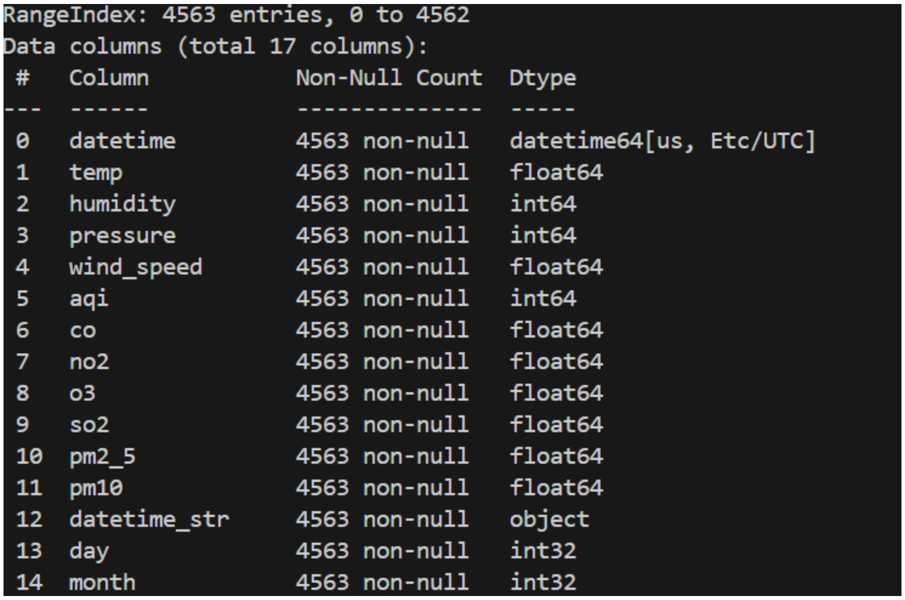
2. Data Collection & Ingestion

Data Sources:

- OpenWeather API for historical and real-time weather data (temperature, humidity, pressure, wind speed).
- OpenWeather Air Pollution API for pollutants (PM2.5, PM10, CO, NO_2, O_3, SO_2).

Process:

The system automatically collects air quality and weather data from the past 180 days and also adds new real-time data every day. After that it connects to the Hopsworks Feature Store to get the saved air quality (AQI) data. It first loads the API key from the .env file to connect securely. After connecting, it downloads the aqi_data feature group. Then, it does some basic data analysis like showing the first few rows, key statistics, missing values, and data types. This helps check if the data is clean, complete, and ready for use in model training or prediction.



```
RangeIndex: 4563 entries, 0 to 4562
Data columns (total 17 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   datetime        4563 non-null   datetime64[us, Etc/UTC]
 1   temp            4563 non-null   float64
 2   humidity        4563 non-null   int64
 3   pressure        4563 non-null   int64
 4   wind_speed      4563 non-null   float64
 5   aqi             4563 non-null   int64
 6   co              4563 non-null   float64
 7   no2             4563 non-null   float64
 8   o3              4563 non-null   float64
 9   so2             4563 non-null   float64
 10  pm2_5           4563 non-null   float64
 11  pm10            4563 non-null   float64
 12  datetime_str    4563 non-null   object
 13  day             4563 non-null   int32
 14  month           4563 non-null   int32
```

Automation:

Daily ingestion is automated through GitHub workflows to fetch, process, and store new data.

3. Machine Learning Models

Four models were trained and evaluated for AQI prediction:

- Random Forest Regressor
- Gradient Boosting Regressor
- Linear Regression
- Ridge Regression

Evaluation Metrics:

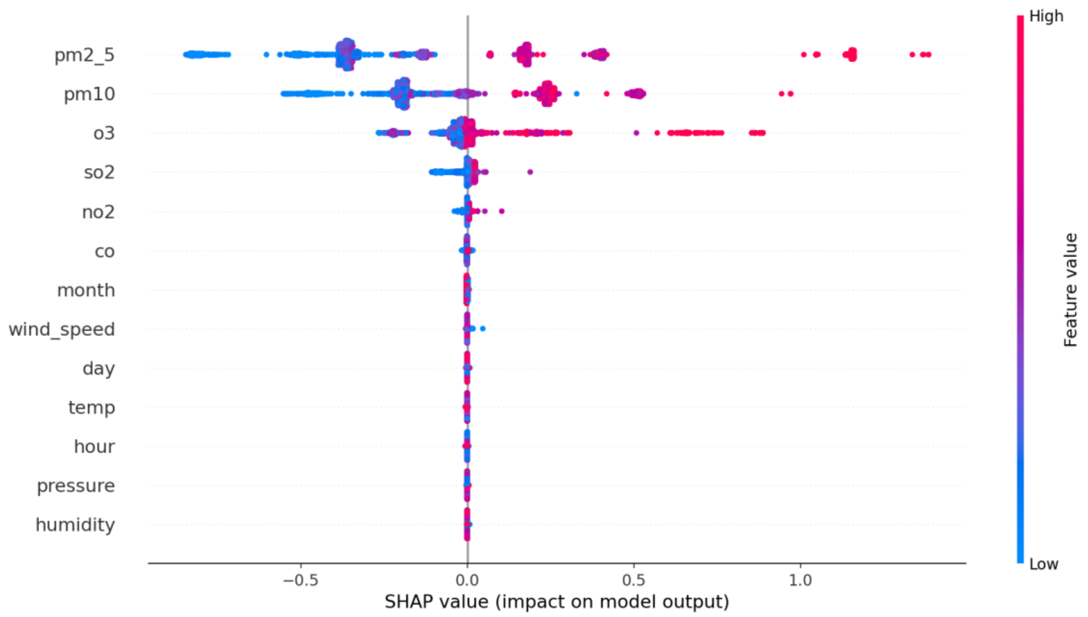
- RMSE (Root Mean Squared Error): Measures prediction accuracy.
- MAE (Mean Absolute Error): Measures average prediction error.
- R^2 Score: Measures variance explained by the model.

Each model was evaluated on historical data split into training and testing sets. The model with the lowest RMSE was selected as the best model, saved as best_model.pkl, and registered in the Hopsworks Model Registry. The best model selected according to the lowest RMSE was Gradient Boosting Model.

4. Model Explainability

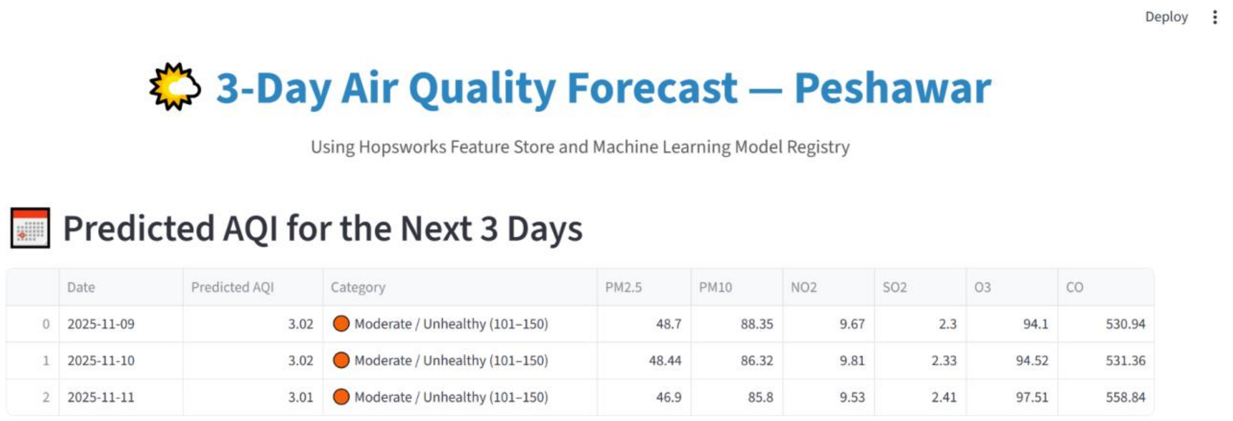
Tool Used: SHAP (SHapley Additive exPlanations)

In SHAP file, it connects to Hopsworks to load the latest AQI data and the best trained model. It then uses SHAP to explain how different features like temperature, humidity, and pollution levels affect the model's AQI predictions. Finally, it creates a summary plot showing which factors have the strongest impact on air quality.



5. Web Application

The Streamlit app connects to the Hopsworks Feature Store to show real-time and predicted Air Quality Index (AQI) for Peshawar. It loads the best ML model from Hopsworks to forecast the next three days AQI and displays results with clear charts and tables. Users can also check historical AQI data and view an AQI category guide.

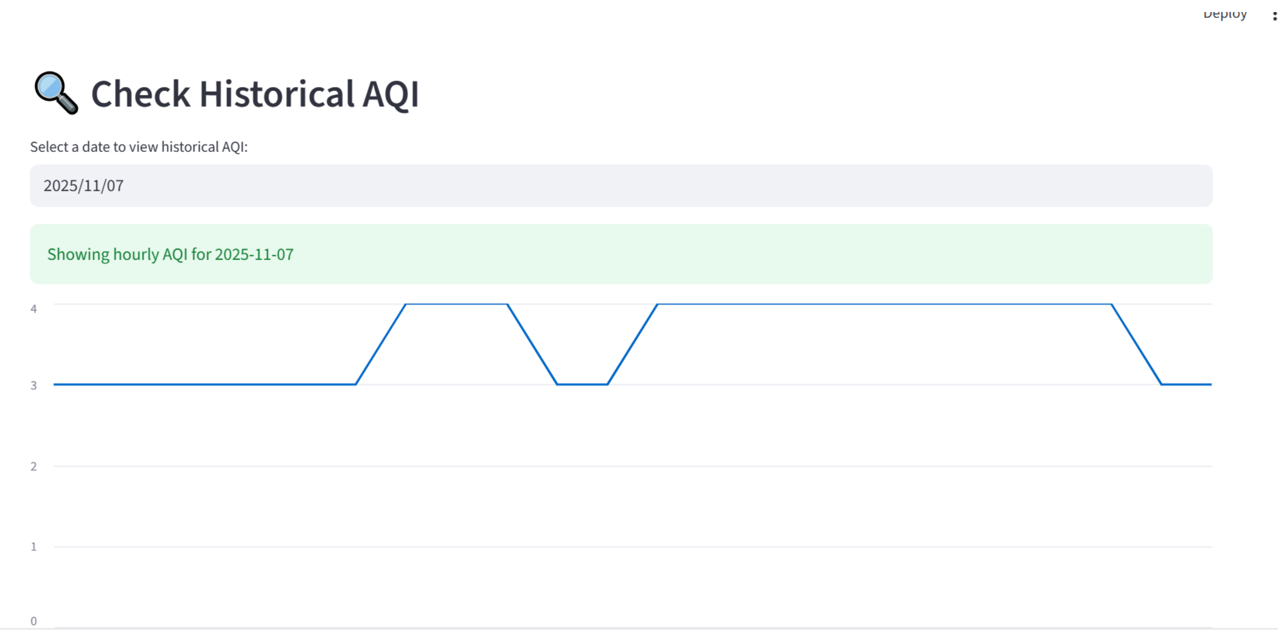


🌤 3-Day Air Quality Forecast — Peshawar

Using Hopsworks Feature Store and Machine Learning Model Registry

📅 Predicted AQI for the Next 3 Days

	Date	Predicted AQI	Category	PM2.5	PM10	NO2	SO2	O3	CO
0	2025-11-09	3.02	🟠 Moderate / Unhealthy (101–150)	48.7	88.35	9.67	2.3	94.1	530.94
1	2025-11-10	3.02	🟠 Moderate / Unhealthy (101–150)	48.44	86.32	9.81	2.33	94.52	531.36
2	2025-11-11	3.01	🟠 Moderate / Unhealthy (101–150)	46.9	85.8	9.53	2.41	97.51	558.84



🔍 Check Historical AQI

Select a date to view historical AQI:

2025/11/07

Showing hourly AQI for 2025-11-07

6. Summary

- Developed a full end-to-end AQI prediction system.
- Selected and deployed the best performing model using robust evaluation.
- Explained model predictions for transparency and trust.
- Provided a real-time web application for monitoring AQI.