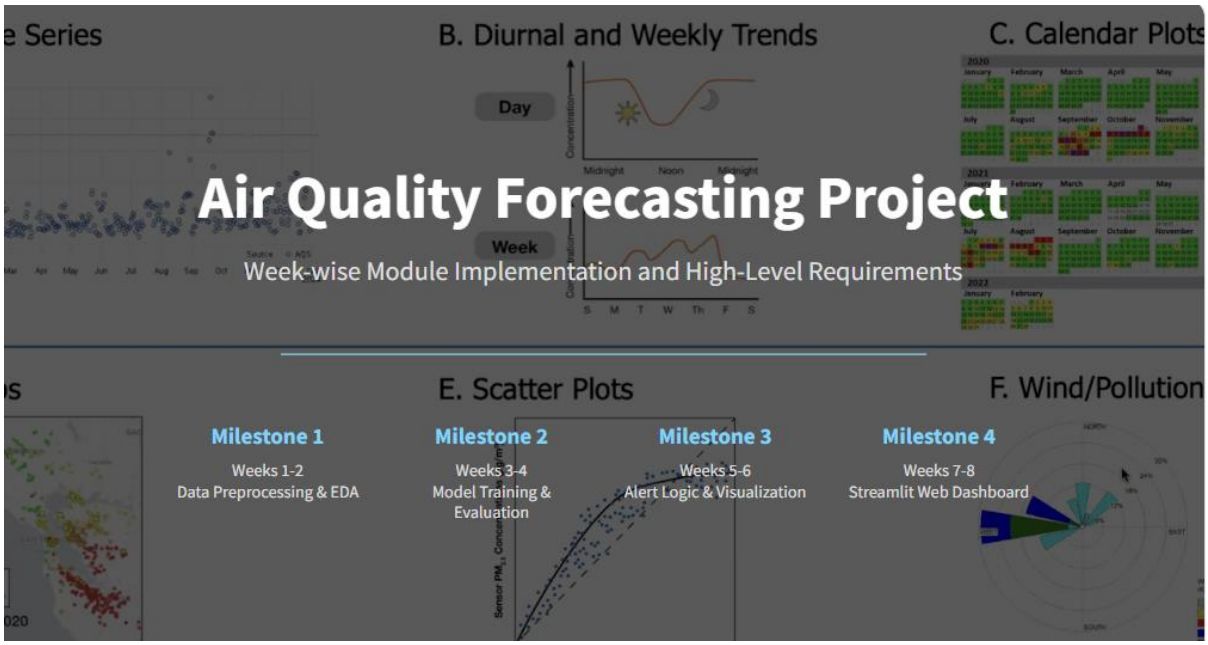
1. Project Title

Forecasting Air Quality Using Historical Pollution Data



---

2. Project Statement

Air pollution is a growing concern in urban environments, directly impacting human health and environmental quality. Accurate forecasting of air quality can help authorities take preventive measures, issue warnings, and inform the public. This project aims to build a time series forecasting system that predicts air quality index (AQI) and pollutant levels using historical pollution data. The goal is to empower environmental agencies, city planners, and the public with timely insights for healthier decision-making.

---

3. Outcomes

- A time series forecasting model to predict AQI and key pollutants (PM2.5, PM10, NO2, etc.).
- Visual dashboard showing historical trends and future air quality predictions.
- Alerts or warnings when predicted AQI crosses safety thresholds.
- Ability to analyze pollutant contributions and seasonal trends.
- Admin functionality to upload new datasets and retrain models for updated forecasting.

---

4. Modules to be Implemented

1. Data Collection & Preprocessing

- Load historical air quality datasets from sources like CPCB, OpenAQ, or Kaggle.
- Handle missing timestamps, resample time series (hourly/daily).

• Normalize pollutant values and handle outliers.
• Feature engineering: day of week, season, temperature if available.

2. Forecasting Model

• Train time series models like ARIMA, Prophet, LSTM, or XGBoost with lag features.
• Compare models using MAE, RMSE, and forecast plots.
• Train per-city or per-station models if data is regionalized.
• Save the best-performing model for inference.

3. Alerting & Trend Analysis

• Calculate AQI category (Good, Moderate, Unhealthy, etc.) from predicted values.
• Highlight days predicted to breach safe air levels.
• Visualize long-term pollutant trends (e.g., seasonal spikes).
• Enable customizable thresholds for alerts.

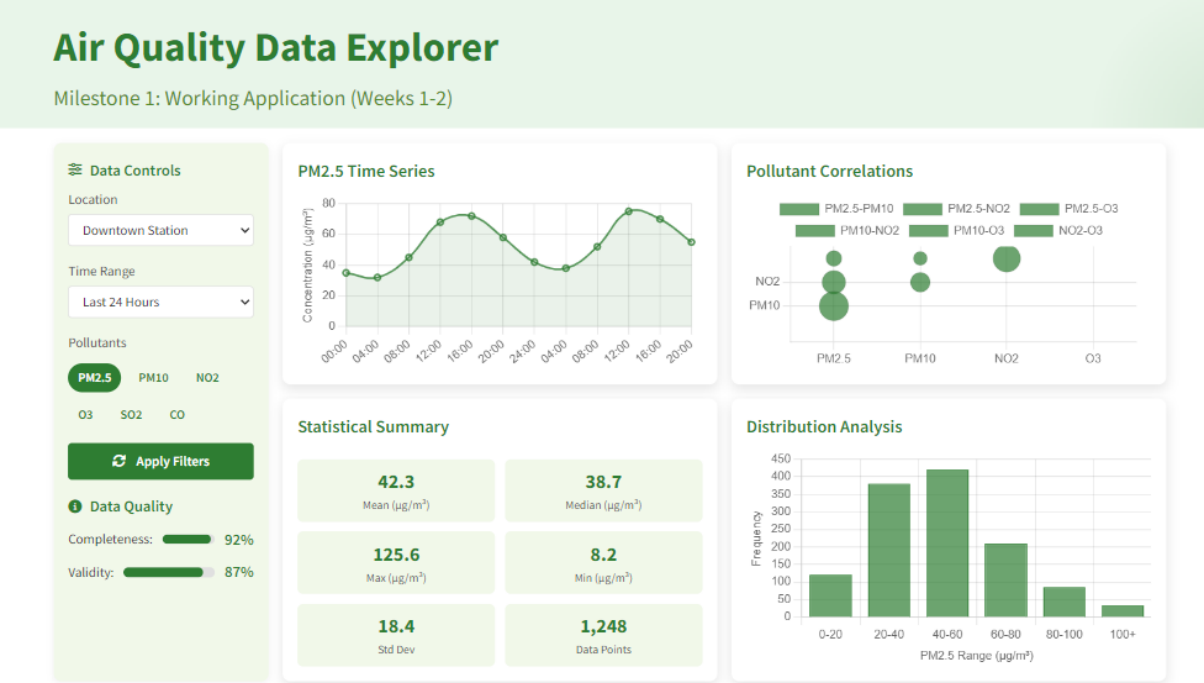
4. Web Interface & Admin Panel

• Develop a Streamlit dashboard for visualizing forecasts and trends.
• Allow users to select city/station and date range.
• Display line plots, AQI gauge meters, and alert banners.
• Admin section for uploading new CSVs and triggering retraining.

---

5. Week-wise Module Implementation and High-Level Requirements

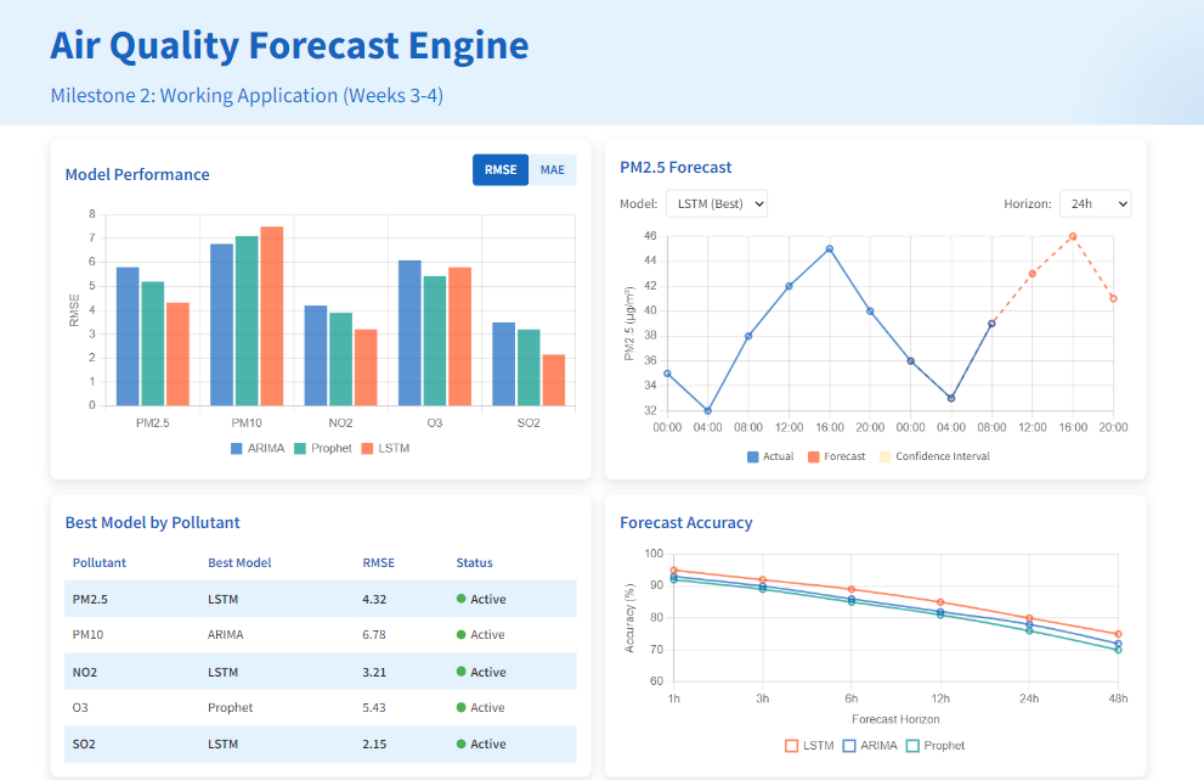
✅ **Milestone 1: Weeks 1–2**
Module: Data Preprocessing & EDA
• Download and preprocess air quality datasets.
• Conduct EDA to understand pollutant trends and correlations.
• Resample data and prepare features for forecasting.



✅ Milestone 2: Weeks 3–4

Module: Model Training & Evaluation

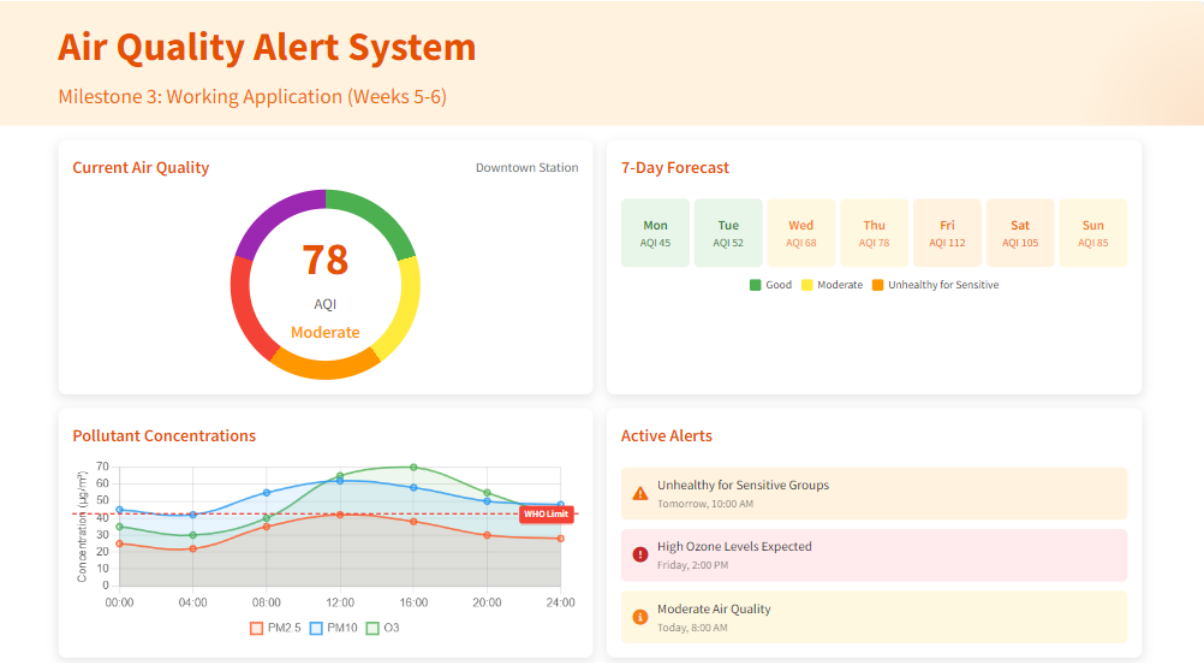
- Train ARIMA/Prophet/LSTM models on historical pollutant values.
- Evaluate predictions with RMSE/MAE.
- Select best-performing model and save for inference.



✅ **Milestone 3: Weeks 5–6**

Module: Alert Logic & Trend Visualization

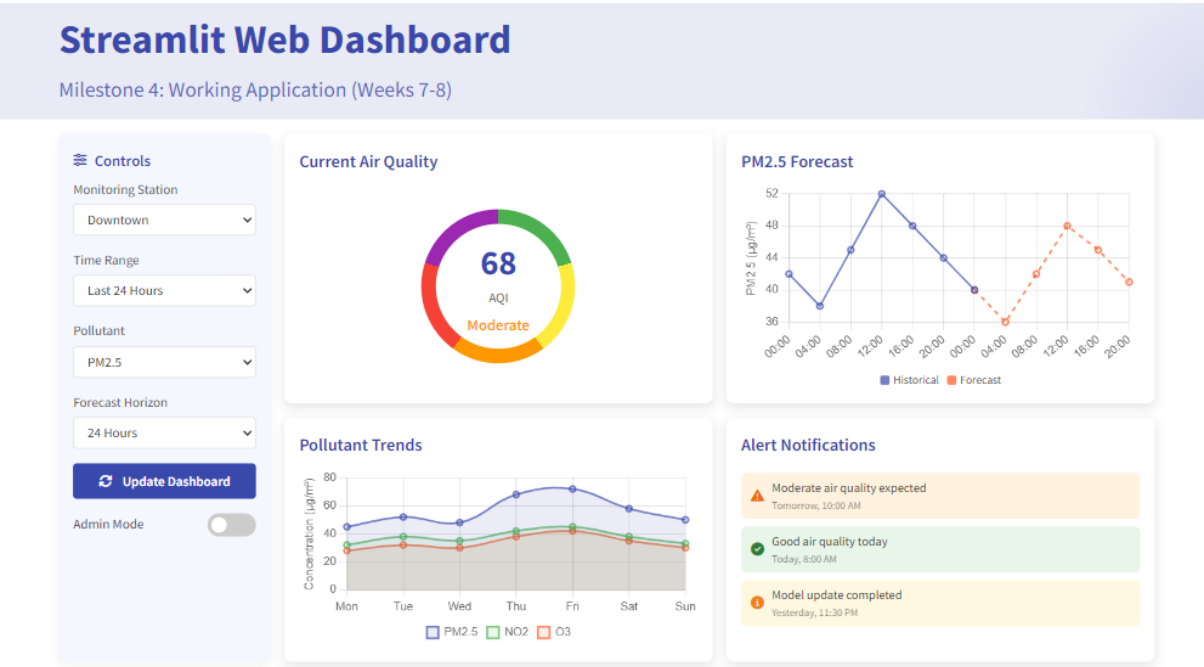
- Define AQI calculation logic and thresholds.
- Highlight high-risk days in forecasts.
- Generate visual trend reports.



✅ **Milestone 4: Weeks 7–8**

Module: Streamlit Web Dashboard

- Interactive UI to select station, timeframe, and forecast variable.
- Display AQI gauge, line plots, and alerts.
- Admin interface to upload new data and retrain models.



---

6. Evaluation Criteria

✅ **Milestone 1 Evaluation (Week 2)**

• Cleaned and structured time series dataset ready.

• EDA insights on seasonal and regional pollution behavior.

✅ **Milestone 2 Evaluation (Week 4)**

• Model RMSE/MAE within acceptable limits for AQI forecasting.

• Model tested across multiple pollutants and timeframes.

✅ **Milestone 3 Evaluation (Week 6)**

• Alerting mechanism functional.

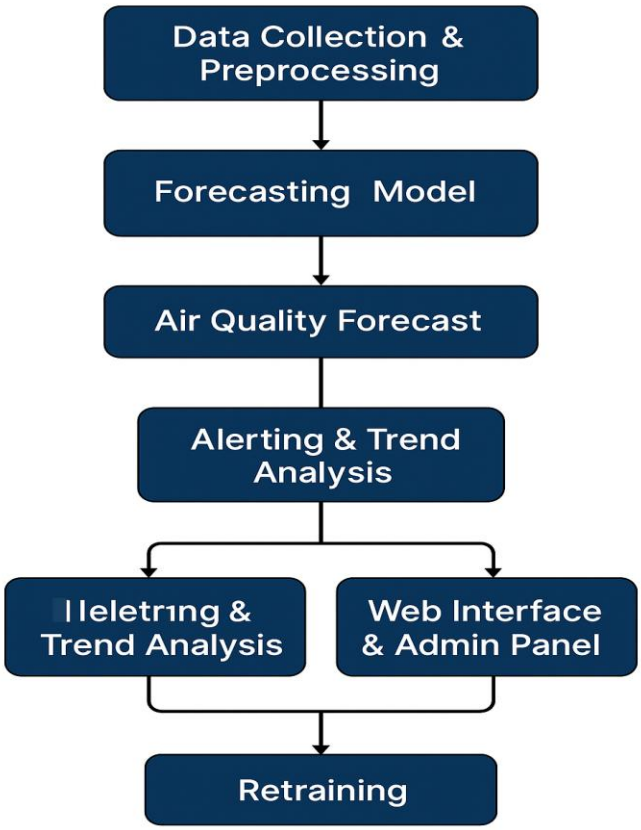
• Forecast reports clearly show high-risk periods and trends.

✅ **Milestone 4 Evaluation (Week 8)**

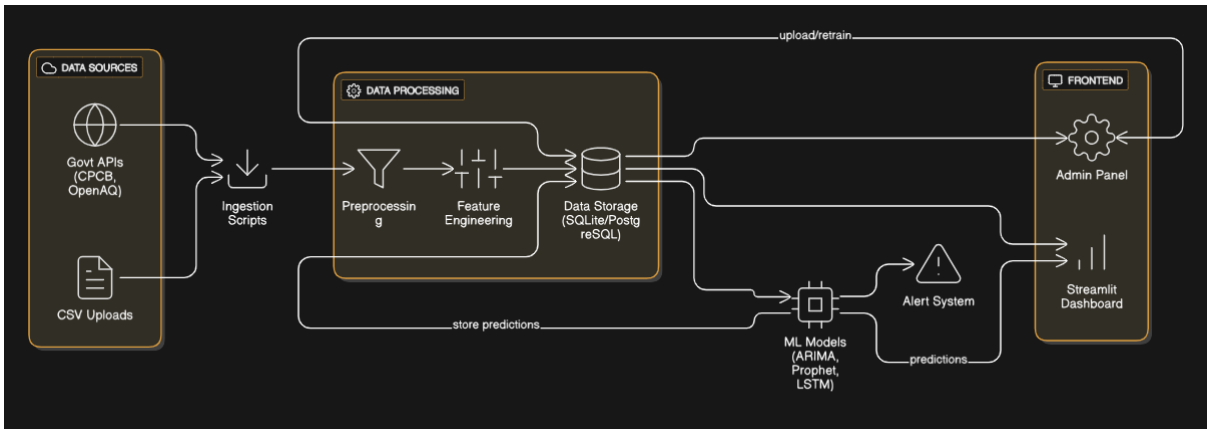
• Streamlit dashboard deployed with forecast visualization.

• Admin features working for dataset upload and model retraining.

7. Workflow Diagram



8. Architecture Diagram



9. Database Schema