Air Quality Prediction in Urban Areas

Acharya Institute of Graduate Studies
Amaan Ahmad Ansari
Manish Kumar Paswan
Jayraj Paul

Problem Statement

Air pollution is a major issue in urban areas, affecting health and the environment. Traditional monitoring methods are limited by sensor availability and real-time data gaps. **AI-based models can predict AQI in advance**, helping authorities take preventive actions like traffic control and public health advisories.

Brief Overview

This case study explores how **machine learning** can analyze pollution, weather, and traffic data to predict AQI. AI-powered forecasting enables early warnings, reducing exposure to harmful pollutants. **Delhi's AI-based AQI prediction** is a real-world example of this approach.

Key Objectives

- Develop an AI model to predict AQI.
- Identify key factors affecting air quality.
- Provide early warnings to reduce pollution exposure.
- Highlight AI's role in pollution reduction, especially in developing countries.

Source : www.freepik.com/

Dataset Overview

The dataset used for AQI prediction includes historical air quality, weather, and traffic data collected over three years. It contains pollutant levels (PM2.5, PM10, NO2, CO, SO2, O3), meteorological data (temperature, humidity, wind speed), and traffic congestion metrics.

Source: Public air quality monitoring agencies, weather APIs, and traffic data providers.
Size: Thousands of records covering multiple urban locations.

Key Features:

- Pollutant levels – PM2.5, PM10, NO2, CO, SO2, O3

Methodology

1. Data Collection & Preprocessing

Collected air quality, weather, and traffic data over three years.
Handled missing values using mean imputation.
Extracted date-based features (Year, Month, Day) for trend analysis.

2. Exploratory Data Analysis (EDA)

Analyzed AQI trends and correlations between pollutants.
Visualized data using heatmaps, line charts, and bar plots.

3. Algorithm Used & Model Training

Algorithm Used: Random Forest Regressor
Chosen for its ability to handle missing values and non-linear relationships.
Works by creating multiple decision trees and averaging their predictions.
Feature Selection: Used PM2.5, PM10, NO2, SO2, O3, and time-based features.
Train-Test Split: Data split into 80% training and 20% testing.
Training Process:
The model was trained on historical data to learn patterns.

Source : www.freepik.com/

4. Model Evaluation

Measured performance using:
Mean Absolute Error (MAE)
Mean Squared Error (MSE)
R^2 Score (for accuracy assessment)
Fine-tuned hyperparameters for improved predictions.

5. AQI Prediction & Forecasting

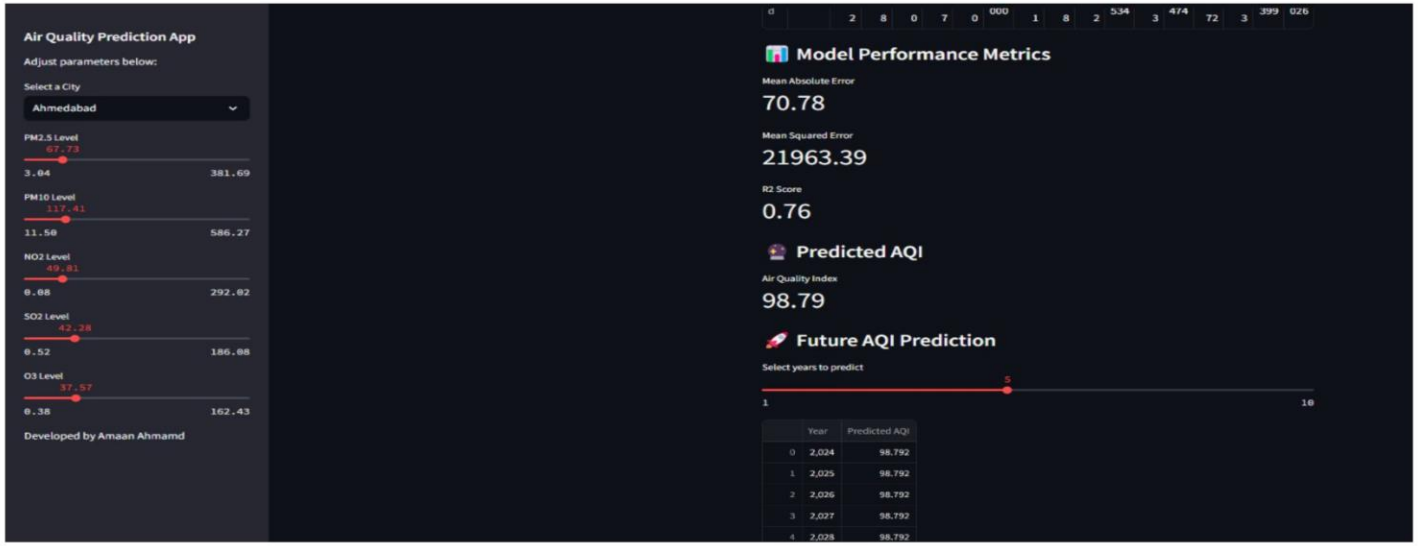
Real-Time Prediction: Model predicts AQI based on user-defined pollutant levels.
Future Forecasting: Predicts AQI trends for upcoming years using trained Random Forest model.

6. Integration with Traffic & Alerts

Can be linked with traffic data to manage congestion.
Provides real-time alerts to the public during high pollution days.

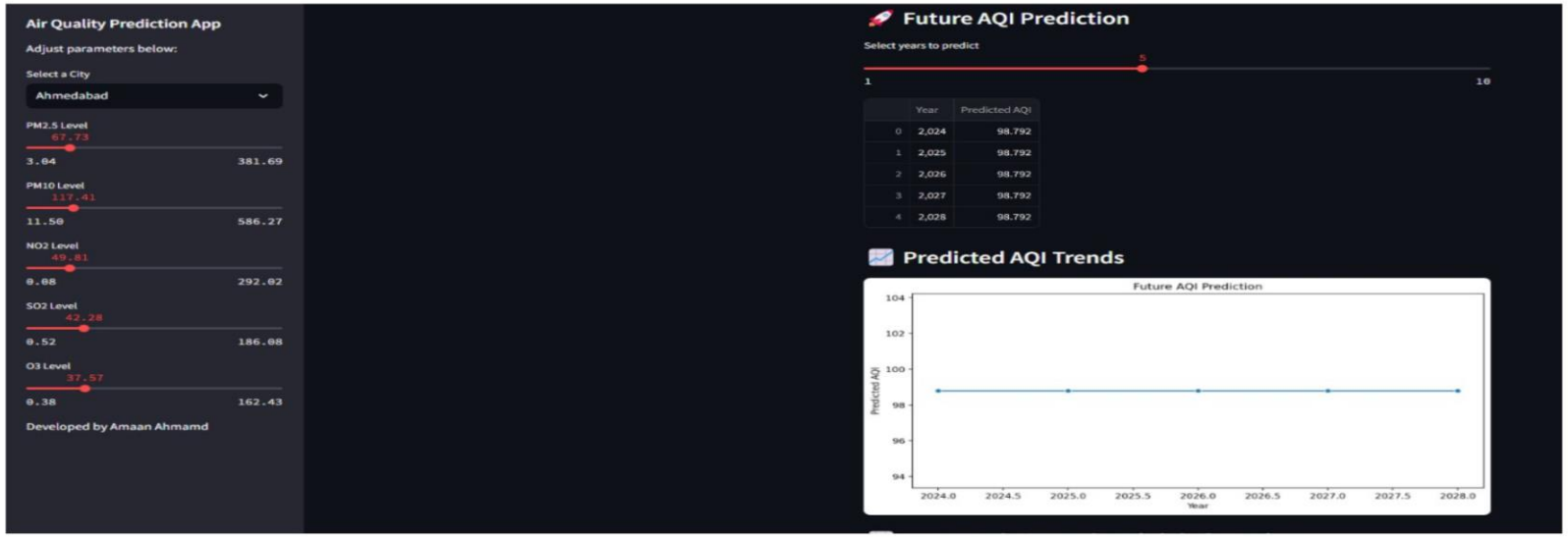
Overview of Data-Set :

Model Performance Metrics :

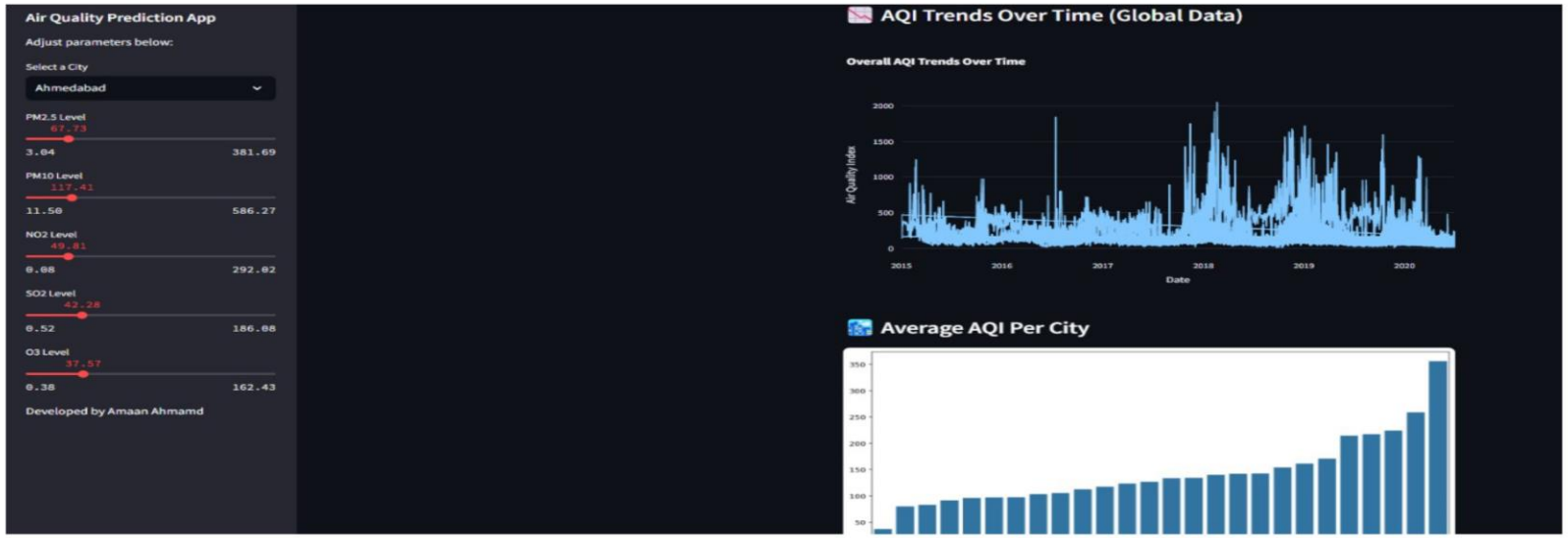


Source : www.freepik.com/

Future Prediction :

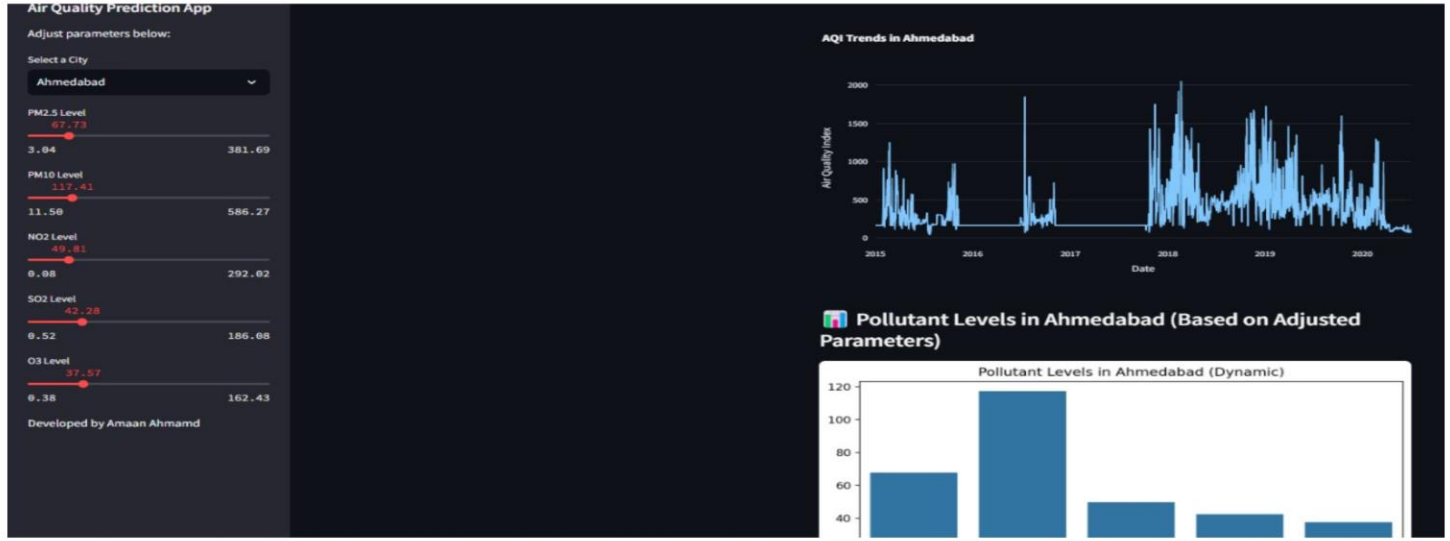


AQI Trends Over Time :

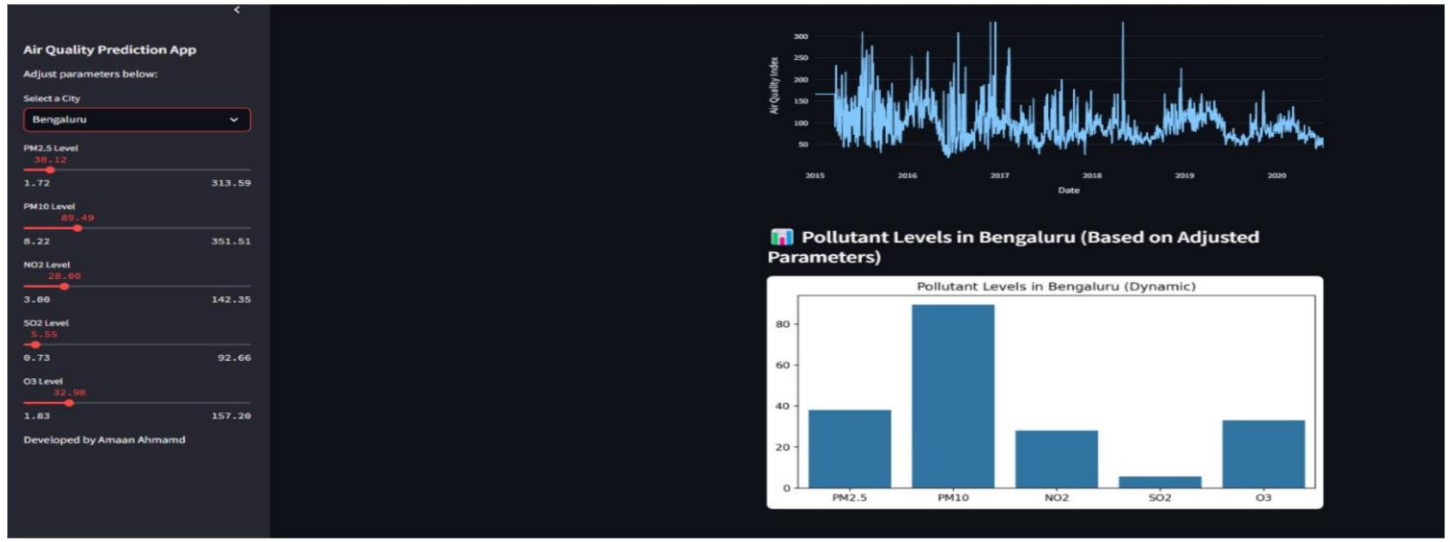


Source : www.freepik.com/

AQI of Particular Area(Ahmedabad) :



AQI of Particular Area(Bangalore) :



Source : www.freepik.com/

QUESTIONS:

Q1. How can machine learning models be used to predict air quality levels?

Ans: Machine learning models like Random Forest Regressor analyze historical air quality data, identifying patterns and correlations between pollutants, weather, and time-based features to predict AQI accurately.

Q2. Which features in the dataset are likely to have the most significant impact on air quality?

Ans: Significant features impacting AQI include PM2.5, PM10, NO2, SO2, O3, and time-based factors (Year, Month, Day), as they strongly correlate with pollution levels.

Q3. Build a model to predict AQI for the next 24 hours.

Ans: A Random Forest Regressor is trained on historical air quality data, using key pollutants and temporal features to predict AQI for the next 24 hours.

Q4. How can this model be integrated with a city's traffic management system to reduce pollution?

Ans: The model can help identify high-pollution periods, enabling traffic control measures like rerouting, vehicle restrictions, and public advisories to reduce congestion and emissions.

Q5. Discuss the role of AI in mitigating air pollution in developing countries.

Ans: AI-powered AQI forecasting helps governments implement proactive measures like early warnings, policy adjustments, and pollution control strategies, benefiting developing countries with limited monitoring resources.

Conclusion

•Air pollution remains a critical challenge in urban areas, affecting public health and the environment.
•Machine learning, specifically the **Random Forest Regressor**, has proven effective in predicting AQI based on historical air quality, weather, and traffic data.
•The model successfully estimates AQI levels and forecasts future trends, aiding in proactive pollution control measures.
•Integration with traffic management and public alert systems can help reduce emissions and minimize exposure to hazardous air quality.
•AI-driven air quality prediction provides **a scalable, data-driven solution** for policymakers to implement better environmental strategies.

Source : www.freepik.com/

References

- *Central Pollution Control Board (CPCB), India - Air Quality Data*
- Breiman, L. (2001). *Random Forests. Machine Learning, 45(1), 5-32.*
- Scikit-Learn Documentation: https://scikit-learn.org/
- Matplotlib & Seaborn Documentation: https://matplotlib.org/
- *Delhi Government's Odd-Even Traffic Policy for Pollution Control (2016)*

Thank You