Nature Presentation

AIR QUALITY INDEX (AQI) PREDICTION

xt

Using Machine Learning on Indian Environmental Data

Presented by
Mohd Saif
Mohd Fahim
Pawan Kumar

www.aqifinder.com

Introduction

Air Quality Index (AQI)

A standardized metric to report daily air quality status.

Importance

Vital for public health, environmental policies, and awareness.

Project Objective

Predict AQI using pollution data from Indian cities.

Goal

Enhance accuracy and reliability of AQI forecasting.

Dataset Overview

Data Source & Location

- Source: Kaggle, Air Quality Data in India
- Combined 5 CSV datasets containing:
- City-wise air quality measurements
- Pollutants: PM2.5, PM10, NO, NO2, NOx, CO, SO2, O3, Benzene, Toluene, Xylene
- Target: AQI (Air Quality Index)

Includes extensive spatial coverage from urban to rural areas.

Methodology

1. **Data Cleaning**

   Handled missing values and removed outliers carefully.

2. **Feature Engineering**

   Created new features to improve model inputs.

3. **Model Selection**

   Tested XGBoost, Random Forest, and LSTM algorithms.

4. **Training & Validation**

   Used an 80/20 split for training and validation datasets.

5. **Evaluation Metric**

   Root Mean Squared Error (RMSE) for accurate comparison.

6. **Experiment Tracking**

   Utilized MLflow for model management and tracking.

Model Training and Evaluation

XGBoost

Lowest RMSE of 25.7, best predictive accuracy.

Random Forest

RMSE of 28.1, strong baseline performance.

LSTM

RMSE of 31.5; needs more data for improvement.

Feature Importance

PM2.5 and PM10 are the most impactful variables.

RESULTS AND VISUALIZATION

Actual vs Predicted AQI plotted for multiple cities.

Time series charts display trends over time clearly.

Model	RMSE
XGBoost	25.7
Random Forest	28.1
LSTM	31.5

Statistical testing confirms significance of results.

Challenges and Solutions

Data Quality

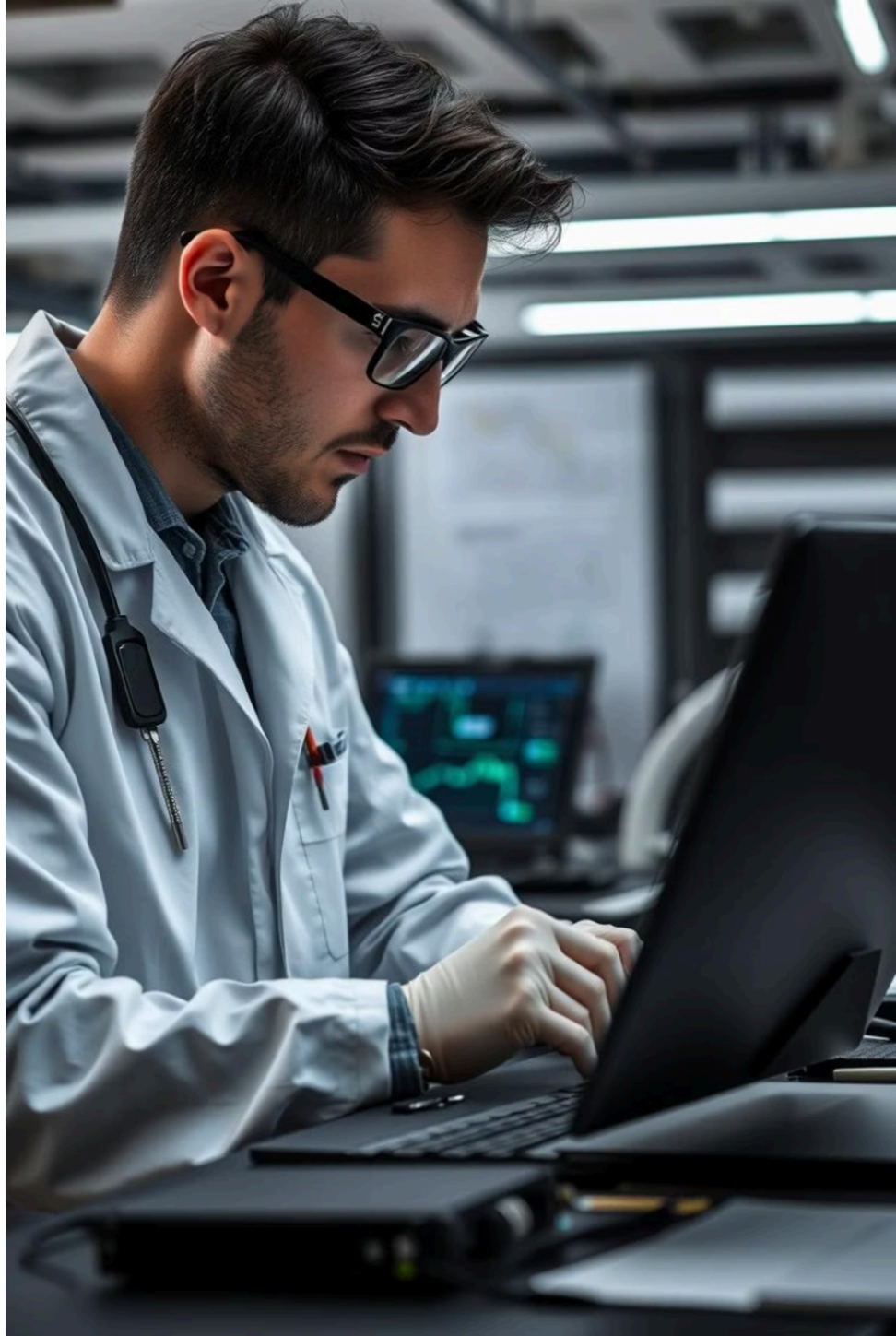
Missing data and sensor noise addressed by imputation and outlier removal.

Model Complexity

Mitigated overfitting using regularization and reduced dimensions.

Interpretability

Analyzed feature importance and SHAP values for transparency.



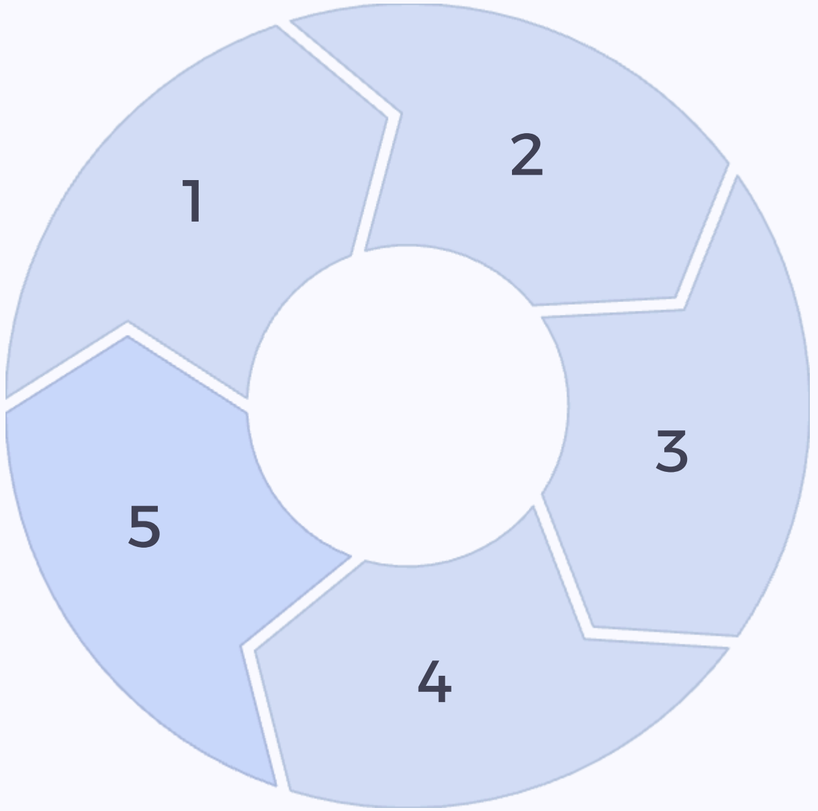
Conclusion and Future Work

Summary

Accurate AQI predictions are achievable with ML methods.

Deployment

Plan to build a web application for public access.



Limitations

Data gaps and limited spatial coverage remain challenges.

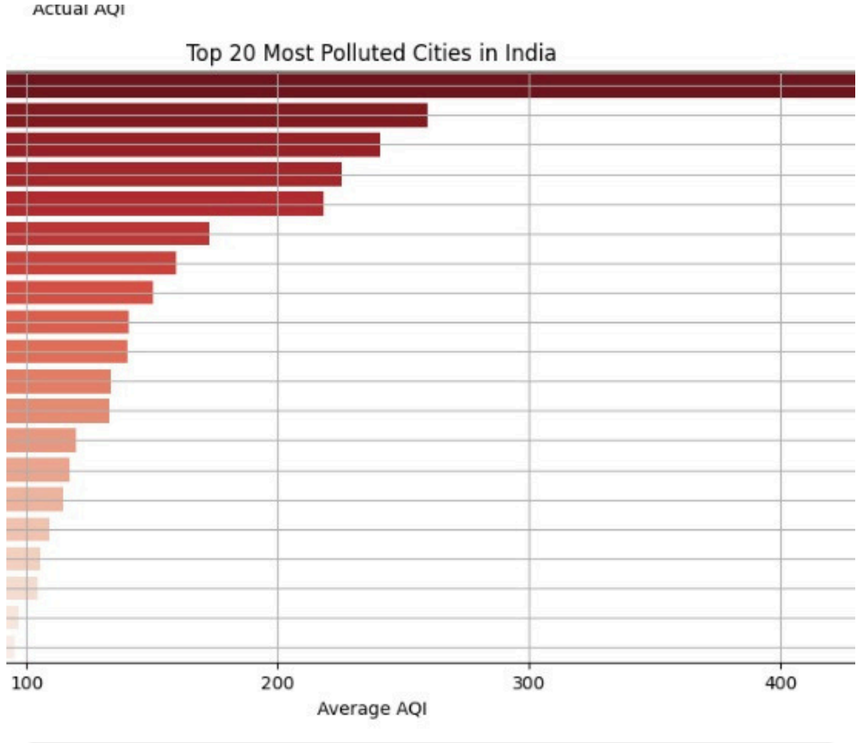
Future Work

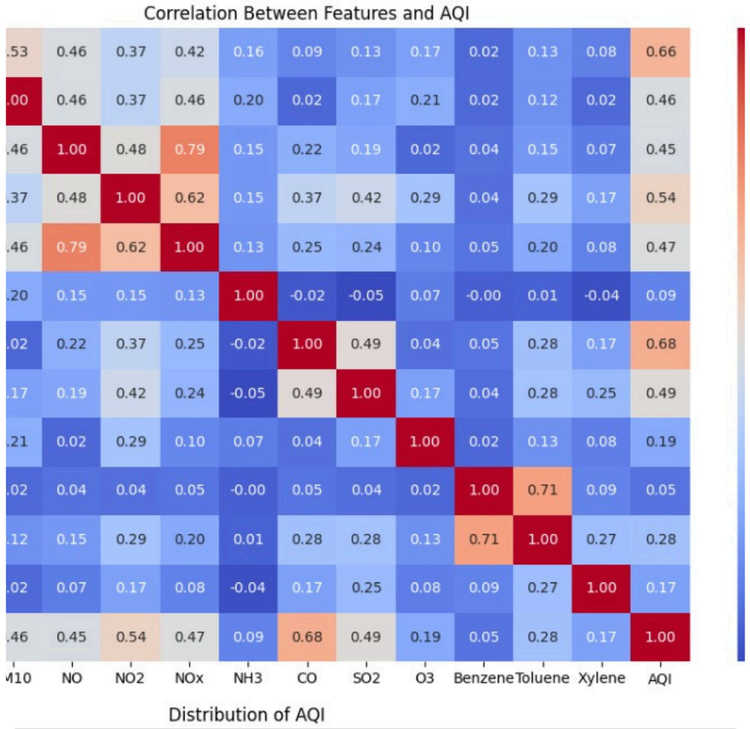
Integrate weather data and satellite imagery for enhancement.

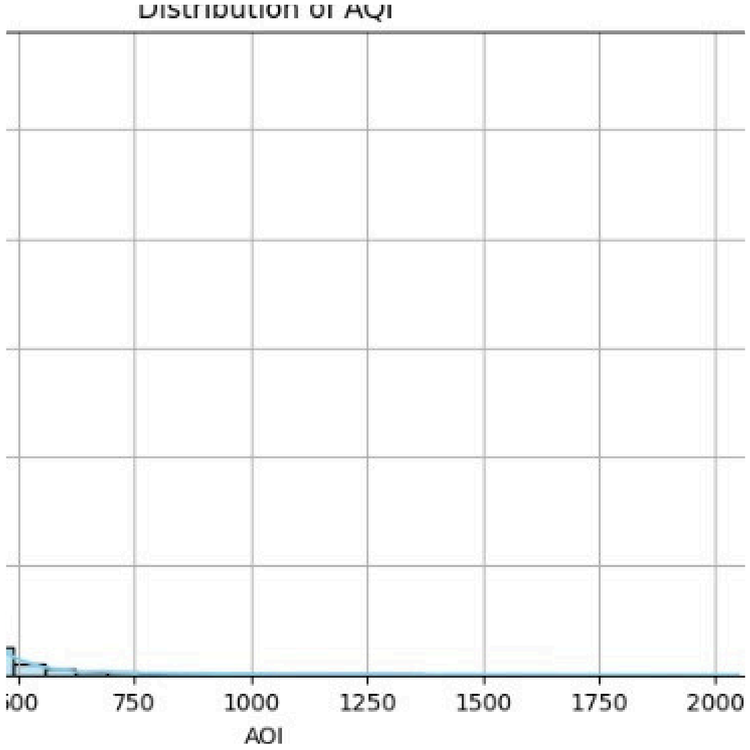
Potential Impact

Enable real-time forecasting and public alert systems.

Output/Result & References.







top 20 Polluted city