Air Quality Index (AQI) Prediction Model

Air quality is vital for health and the environment. The Air Quality Index (AQI) reports daily air quality. It measures pollutants like ozone and particle pollution. The EPA (Environmental Protection Agency) is the source for this data.

BY Prashant Kumar

(prashantpacific53@gmail.com)

Learning Objectives

1 Understand AQI

Grasp the concept and importance of the Air Quality Index.

2 ML for Prediction

Discover how machine learning predicts air quality.

3 Explore Techniques

Learn data preprocessing, training, and evaluation methods.

4 Real-World Insights

Gain insights into AQI prediction applications.

Tools and Technology used

Programming Language

The primary programming language used for the project is Python. It facilitates efficient data manipulation and model implementation.

Data Libraries

Key libraries utilized include Pandas for data handling, NumPy for numerical operations, and Scikit-Learn for machine learning. Additionally, Matplotlib and Seaborn are employed for data visualization.

Machine Learning Algorithms

The project implements two machine learning algorithms: Random Forest Regressor and Linear Regression. The Random Forest Regressor is noted for its superior performance in predicting AQI.

Data Source

The dataset used for the analysis is sourced from 'air quality data.csv'. This file contains historical air quality data essential for model training.

Deployment Options

Future deployment options include Flask and Streamlit. These frameworks will enable the model to be accessible as a real-time web application.

Methodology

1

Data Collection

- Collect AQI data from sources like CSV files or APIs.
- Include pollutant levels (PM2.5, PM10, NO_2) and weather factors.

2

Data Preprocessing

- Handle missing values and remove duplicates.
- Normalize and clean data for better analysis.

3

Exploratory Analysis

- Visualize AQI trends using Matplotlib & Seaborn.
- Identify correlations between pollutants and weather.

4

Model Training

- Train regression models for AQI prediction.
- Evaluate performance using RMSE, MAE, and R^2 score.

Problem Statement:

AQI Significance

The Air Quality Index (AQI) indicates air quality levels and their impact on health.

It plays a crucial role in guiding public health policies and decisions.

Monitoring Challenges

Limited monitoring stations hinder comprehensive air quality assessments.

Data accuracy issues and temporal variability complicate effective monitoring.

Predictive Modeling Need

Data-driven predictive models are essential for forecasting air quality trends.

Such models support informed decision-making and enhance public awareness of air quality issues.

Who Is Most at Risk?

- Cardiovascular Diseases
- Children
- Elderly
- Respiratory Diseases

Sensitive groups include children, the elderly, and those with respiratory or cardiovascular diseases. They are at higher risk due to developing lungs or weakened immune systems. Source: American Lung Association, WHO data.

Solution:

ML-Based Model

AQI Prediction Model uses historical data.

Key Pollutants

Analyzes PM2.5, PM10, NO2, CO, SO2, O3.

Tools

Uses Python, Pandas, NumPy, Scikit-Learn.

Algorithms

Implements Linear Regression & Random Forest.

ML-Based AQI Prediction Model

Developed using historical air pollution data to forecast air quality levels.

Utilizes machine learning techniques to enhance prediction accuracy.

Key Pollutants Analyzed

Focuses on PM2.5, PM10, NO2, CO, SO2, and O3 for AQI prediction.

These pollutants are critical for assessing air quality and health impacts.

Technologies and Tools

Implemented using Python programming language for robust data handling.

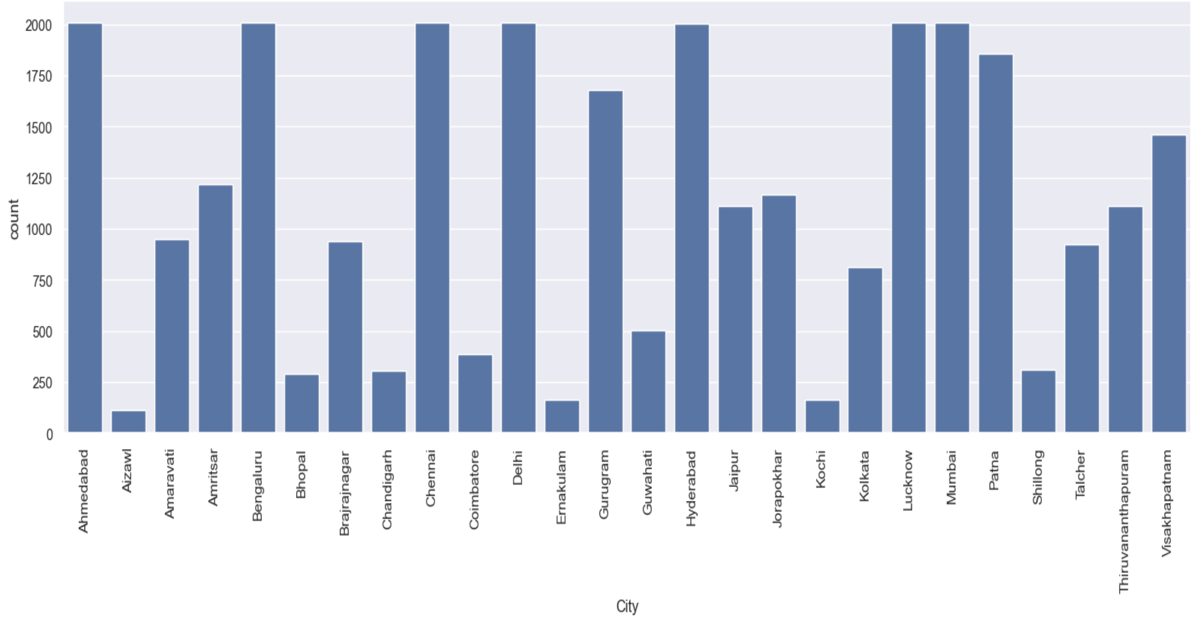
Employs libraries like Pandas, NumPy, and Scikit-Learn for data analysis and model training.

Modeling Techniques

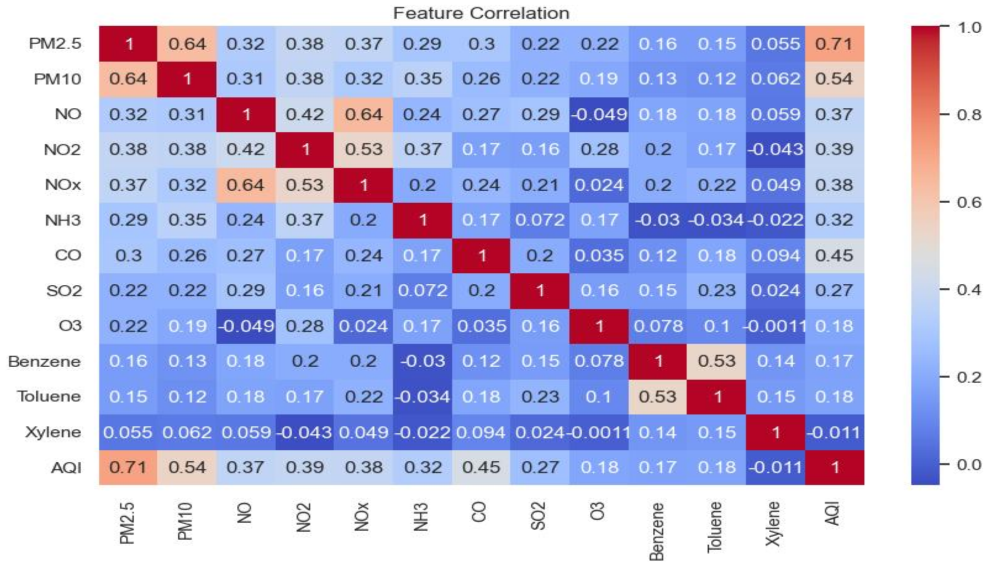
Incorporates Linear Regression and Random Forest Regressor for effective forecasting.

Random Forest Regressor demonstrated superior performance in predictions.

Screenshot of Output:



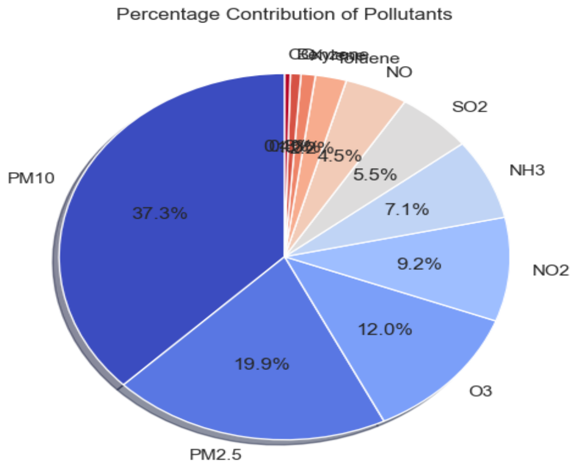
Visualization Of the Data



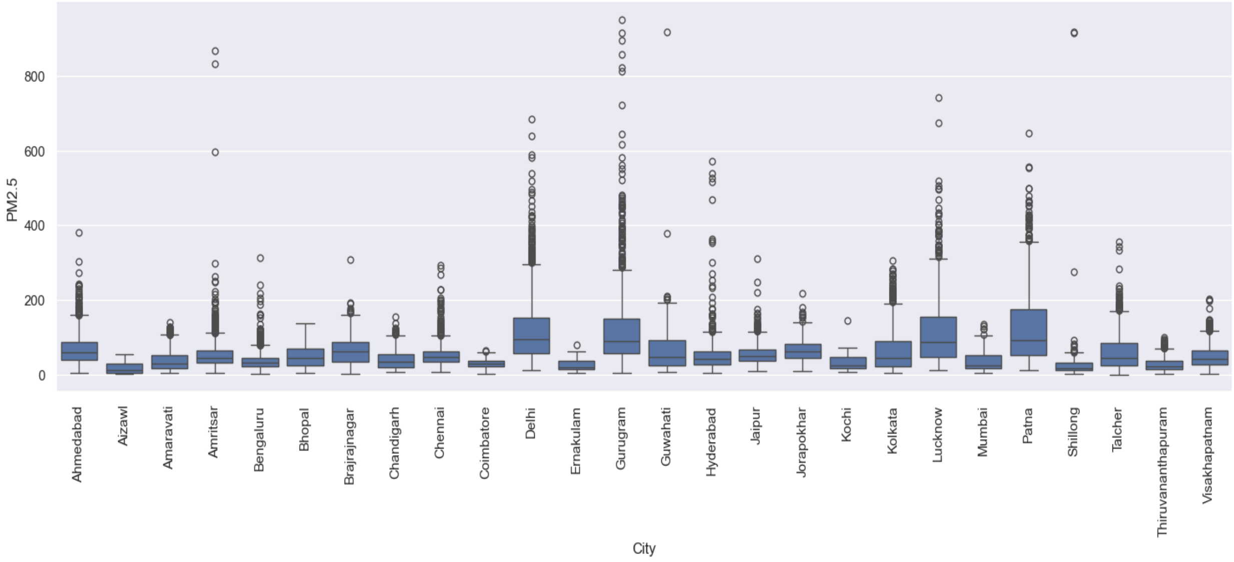
Feature Correlation by Using of Heatmap

```
df.dropna(inplace=True)
df.drop_duplicates(inplace=True)
```

Data Preprocessing Code



Percentage Contribution Of Pollutants



AQI Prediction of PM2.5

Conclusion:

Model Accuracy

The AQI prediction model effectively predicts air quality based on pollutant levels.

The Random Forest Regressor outperformed the Linear Regression model in accuracy.

Algorithm Comparison

The performance of the Random Forest Regressor was significantly better than that of Linear Regression.

This comparison highlights the importance of selecting appropriate algorithms for predictive modeling.

Deep Learning Integration

Future improvements include plans to integrate deep learning techniques into the model.

This integration aims to enhance prediction accuracy and model robustness.

Real-Time Deployment

There is an intent to deploy the model as a real-time web application.

This deployment will facilitate immediate access to AQI predictions for users.

Dataset Expansion

Aiming to expand the dataset will improve the model's accuracy and reliability.

Incorporating more diverse data sources will enhance the model's predictive capabilities.

AQI Level

THANK YOU

Acknowledgments

We extend our sincere gratitude to our mentors, collaborators, and the AICTE, edunet Foundation, & Shell for their support and guidance in this project.