Air Quality Index (AQI) Prediction

Authors

Hein Thuya Win
Shuvanga Bikram Hamal
Phyo Thwe Thwe Kyaw

Affiliations

Department of ICT, Rangsit University, Thailand

Air pollution, the contamination of air by harmful substances like gases, particulates, and chemicals, poses a critical threat to health, the environment, and economies worldwide. It contributes to severe health conditions, including respiratory and cardiovascular diseases, and is responsible for millions of premature deaths annually. Addressing air pollution is not just an environmental necessity but a vital step toward ensuring a healthier, sustainable future for all.

1 Introduction

This research focuses on predicting Air Quality Index (AQI) values by analyzing major air pollutants. Accurate AQI predictions are critical for timely interventions, public health advisories, and urban planning. By leveraging advanced predictive models, this study aims to enhance the accuracy of AQI forecasting, which can lead to better environmental management and informed decision-making in urban settings.

2 Objective

- Examine the impact of major air pollutants
- Create accurate models for forecasting AQI values
- Provide timely and reliable AQI predictions to help mitigate health risks and enhance air quality management for policymakers
- Contribute to Literature by adding valuable data and methodologies

3 Methodology

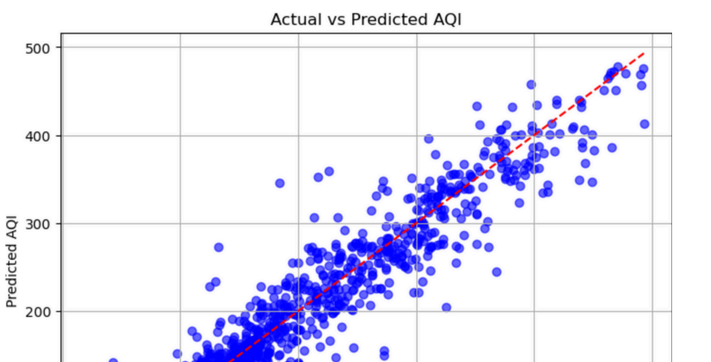
- Data Collection & Data Preprocessing
- Model Selection - Random Forest, Gradient Boosting, XGBoost
- Data Splitting & K-fold Cross-Validation
- Performance Evaluation Metrics (MAE, RMSE, R^2)
- Software & Tools (Python 3.12, Jupyter Notebook)

4 Results

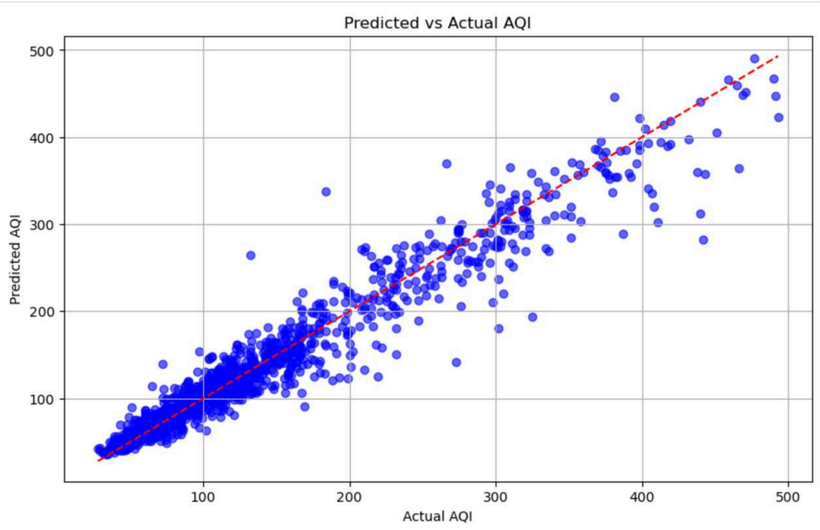
- Random Forest -> MAE: 15.09, RMSE: 22.91, R^2: 0.9371
- Gradient Boosting -> MAE: 14.10, RMSE: 22.20, R^2: 0.9388
- XGBoost -> MAE: 13.91, RMSE: 21.31, R^2: 0.9436

5 Analysis

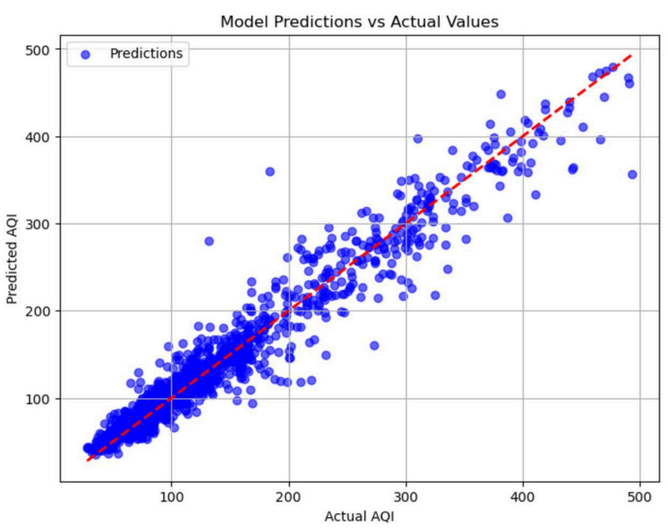
- To complement the model performance evaluation, several visualizations were created to provide deeper insights into the data and prediction results.



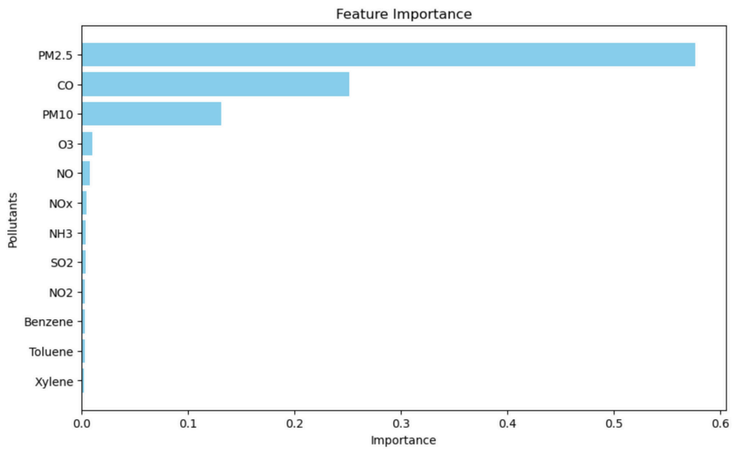
Random Forest



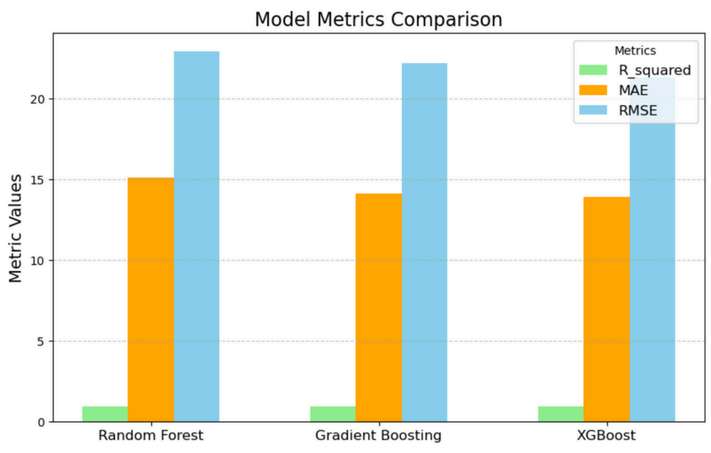
Gradient Boosting



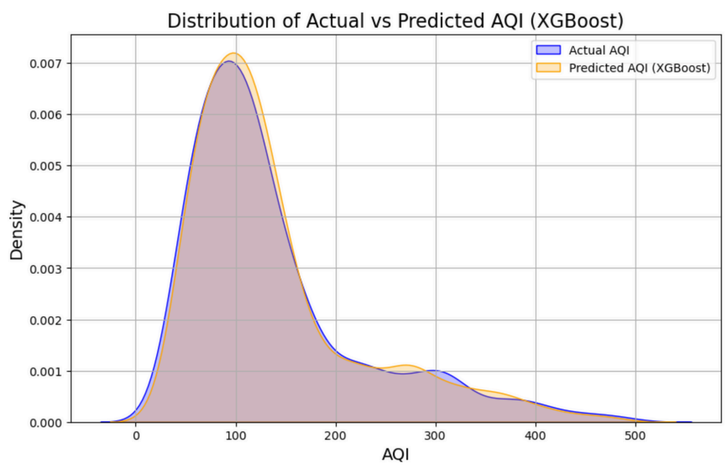
XGBoost



Feature Importance Plot



Metric Evaluation



KDE Distribution Plot

6 Conclusion

- XGBoost demonstrated the best performance with the lowest RMSE (21.31) and highest R^2 (0.9436), making it the most reliable model for AQI prediction in this study.
- Machine learning models demonstrated significant potential for real-time urban air quality monitoring, aiding in data-driven policymaking and public health interventions.
- Future Directions: Incorporating meteorological data and advanced AI techniques can enhance prediction accuracy and scalability, enabling more robust environmental management solutions.