Predictive Modeling of Urban Air Quality Using Machine Learning and Satellite Data

Authors: Johan Carlo Ilagan, Emily Nguyen, David Chen
Journal: Environmental Data Science Journal
DOI: 10.1016/j.envds.2025.0023
Publication Date: March 15, 2025

Abstract

Air pollution remains one of the leading environmental health threats in urban areas. This paper introduces a predictive model for estimating daily air quality indices (AQI) across Australian cities using satellite imagery and machine learning techniques. The model integrates aerosol optical depth (AOD), meteorological data, and urban density metrics, achieving an R^2 of 0.89 in validation. These results demonstrate the value of data fusion for near–real-time environmental analytics.

---

1. Introduction

Accurate air quality forecasting is essential for public health, urban planning, and environmental policy. Traditional ground-based sensors provide limited coverage, especially in developing regions or rural zones. To address this gap, we propose a hybrid modeling approach combining remote-sensing and local atmospheric data, supported by machine learning algorithms capable of nonlinear pattern detection.

---

2. Methods

2.1 Data Collection

Satellite-derived AOD data were obtained from NASA's MODIS platform and paired with meteorological datasets from the Bureau of Meteorology (BoM).

2.2 Model Development

We implemented Gradient Boosted Regression Trees (GBRT) using Python's *scikit-learn* library. The model was trained on three years of data (2020–2022) and validated on an independent 2023 dataset.

2.3 Feature Engineering

Key features included temperature, humidity, wind speed, traffic density, and vegetation indices (NDVI). Feature importance analysis highlighted AOD and wind speed as top predictors.

---

3. Results

The proposed model achieved an R^2 of 0.89 and a mean absolute error of 4.2 AQI points. Spatial interpolation revealed distinct air quality gradients around major cities, particularly Sydney and Melbourne. The visualization dashboard provides intuitive insights into pollution hotspots and temporal fluctuations.

---

4. Discussion and Conclusion

This research illustrates the potential of machine learning for environmental forecasting. Future improvements will incorporate deep learning architectures (e.g., LSTM networks) for temporal prediction.
Our findings support the integration of satellite data analytics into national air monitoring frameworks for scalable, cost-effective, and high-resolution AQI estimation.

Acknowledgements

We thank the Australian Space Data Analysis Facility for computational support and the Bureau of Meteorology for access to climate data.