Air Quality Prediction Using Machine Learning

Sheela S Maharajpet *[1], Likhitha S †[2], and Kiran T ‡[3]

[1]Department of MCA, Acharya Institute of Technology, Bangalore
[2]Department of MCA, Acharya Institute of Technology, Bangalore
[3]Assistant Professor. Department of MCA, Acharya Institute of Technology, Bangalore

Abstract

Evaluating the nature of the air has become essential for inhabitants in numerous modern and metropolitan regions these days. Air quality is significantly impacted by pollution from fuel consumption, transportation, and electricity generation. The build-up of toxic gasses has a negative impact on smart city residents' quality of life. We require productive air quality observing and forecast models in order to combat the escalating levels of air contamination. These models measure local air contamination and collect data on pollutant concentrations. Particulate matter is made up of tiny solid or liquid particles that can have a serious negative effect on health if inhaled. As a result, assessing and it is turning out to be more significant to anticipate air quality.

Keywords: NSGAII optimized neural network. SOM neural networkair quality prediction..

*Email: sheelamaharajpet4@gmail.com Corresponding Author
†Email: likhitha.22.mcav@acharya.ac.in
‡Email: kiran2810@gmail.com

1 Introduction

Evaluating air quality has become an important concern for inhabitants in modern industrial and urban contexts. Transportation, power, and fuel use all contribute to pollution, which severely deteriorates air quality and affects community health and well-being. The build-up of dangerous gasses presents a serious risk to people's quality of life in smart cities. Strong air quality monitoring, in-depth air contamination information, and forecast frameworks are essential. These systems collect data on pollutant concentrations and provide localized assessments, enabling proactive measures to mitigate pollution's effects. Particulate matter, consisting of microscopic solid or liquid droplets, is particularly concerning due to its severe health implications when inhaled.Further, the swift advancement of industrialization and urbanization has led to a growing consciousness among people regarding air quality. Air quality prediction is an essential application field that may give real-time information on air pollution, which is useful for both government environmental protection departments and regular citizens. (Liu, Cui, & Liu, 2024). Thus the forecast of the air pollution index could be useful for directing traffic and locating harmful pollutants.(Maltare & Vahora, 2023).

In this context, the Random Forest technique is a promising tool for predicting and measuring pollution levels in large urban centers. By incorporating real-time traffic data, weather information, and road data, this technique offers a comprehensive approach to data training and prediction, improving the precision of projections for air quality Our proposed machine learning-based model for air quality prediction has demonstrated impressive training and testing accuracies of 99% and 90%, respectively, indicating its effectiveness in forecasting air pollution concentrations. Additionally, the versatility of this approach allows it to be applied across various fields and databases, highlighting its potential to address air quality challenges on a broad scale.

2 Literature Review

In recent years,"Assessment and prediction of air quality using fuzzy logic and autoregressive models" (2012) stated AI has been used to address environmental issues, including air quality. This work presents two models for air quality assessment. The first model uses a Sigma operator to statistically analyze historical data and evaluate toxicity. The second model employs a fuzzy inference system to classify air quality into five stages: excellent, good, regular, bad, and danger. With the growing awareness of the negative health effects of air pollution, air quality forecasting has become essential in environmental research.(Kumar & Goyal, 2011). The goal of this research is to create a model that will help with daily Air Quality Index (AQI) prediction and decision-making. For pollutants including RSPM, SO2, NO2, and SPM, the US Environmental Protection Agency (USEPA)

Editors: Ratnakirti Roy, Manish Kumar Thakur, and Anushree Raj
DOI:10.48001/978-81-966500-7-0-9 | ISBN: 978-81-966500-7-0

provides a method for estimating the AQI that is modified for Indian regulations.

Singh et al.'s (2012) study predicted urban air quality in Lucknow, India, using linear and nonlinear models, including PLSR, MPR, and ANN approaches. Data from 2005-2009 were used to predict RSPM, SO2, and NO2. Three ANN models (MLPN, RBFN, GRNN) were developed and compared with other models using statistical criteria like correlation coefficient (R) and root mean squared error (RMSE). Nonlinear models, especially ANNs, performed better than linear PLSR and low-order nonlinear MPR models. Jamshedpur, a steel city in eastern India, faces rising air pollution due to industrial activities.(Singh et al., 2012) . The Industrial Source Complex Short-Term gaussian dispersion model estimated NOx emissions from industries, vehicles, and domestic sources, contributing 53%, 40%, and 7%, respectively. Statistical analysis showed the model accurately predicted NOx concentrations with about 68% accuracy.

Gokhale and Raokhande's (2008) evaluated roadside air quality in Guwahati, three models—M-GFLSM, CALINE3, and CAL3QHC—were compared. These models were assessed using vehicle-derived PM10 and PM2.5 emissions, meteorological data, and daily average concentrations during winter. Statistical evaluation aimed to identify the best-performing model under local conditions. In their 2018 review, Fu, Li, and Chen's (2023) examined large information and AI approaches for air quality prediction. They draw attention to the benefits of machine learning models over conventional statistical techniques for managing intricate, substantial datasets. The study examines several algorithms, evaluates how well they function, and points out problems and directions for further investigation. Applications to public health and policy-making highlight how crucial precise air quality forecasts are.

Machine learning regression strategies for PM2.5 prediction are surveyed in the work "Forecasting air pollution particulate matter (PM2.5) using machine learning regression models" by Harishkumar, Yogesh, Gad, et al.'s (2020). It evaluates the performance of models such as gradient boosting, SVR, random forest, and linear regression by comparing them and using measures like RMSE and MAE. The paper discusses feature selection, preprocessing techniques, difficulties, and public health and urban planning applications. The paper "Air temperature forecasting using machine learning techniques: a review" by Cifuentes et al.'s (2020) provides an extensive analysis of machine learning techniques for air temperature prediction. It evaluates regression models, artificial neural networks, support vector machines, and ensemble methods like random forests and gradient boosting. The study discusses feature selection, preprocessing techniques, challenges, and applications in energy management and climate studies.

Lei et al.'s (2022) covers regression models, neural networks, SVMs, and ensemble methods applied to predict PM2.5, PM10, NO2, SO2, and O3 concentrations. The study compares model performance using metrics like MAE and RMSE, discusses challenges, and

Editors: Ratnakirti Roy, Manish Kumar Thakur, and Anushree Raj
DOI:10.48001/978-81-966500-7-0-9 | ISBN: 978-81-966500-7-0

suggests future research directions. Applications in urban planning and public health underscore the significance of accurate air quality predictions. "Machine learning algorithms to forecast air quality: a survey" (2023) covers regression models, neural networks, SVMs, decision trees, and ensemble methods applied to predict PM2.5, PM10, NO2, and O3 concentrations. The study compares model performance using metrics like MAE, RMSE, and R-squared, discusses feature selection, preprocessing techniques, challenges, and suggests future research directions. Applications in urban planning and public health underscore the significance of precise forecasts for air quality.

3 Proposed System

This paper aims in order to deal with the pressing issue of air pollution by utilizing the Random Forest algorithm to predict the presence of air pollutants. (see figure 1).The Air Quality Index (AQI) serves as a vital indicator of air quality, with it's range from 0 to 500 signifying various levels of pollution severity, as defined by the Indian Government (CPCB). Eight major pollutants, including particulate matter (PM 10 and PM 2.5), carbon monoxide (CO), ozone (O3), nitrogen dioxide (NO2), sulphur dioxide (SO2), ammonia (NH3), and lead (Pb), contribute to AQI calculations. The proposed framework use AI techniques, specifically the Random Forest Classifier, to analyze publicly available datasets containing environmental features and predict air quality class based on pollutant concentrations. This system, implemented in Python, processes input datasets through the Random Forest Classifier to generate predictions, thus facilitating proactive measures for air quality management.

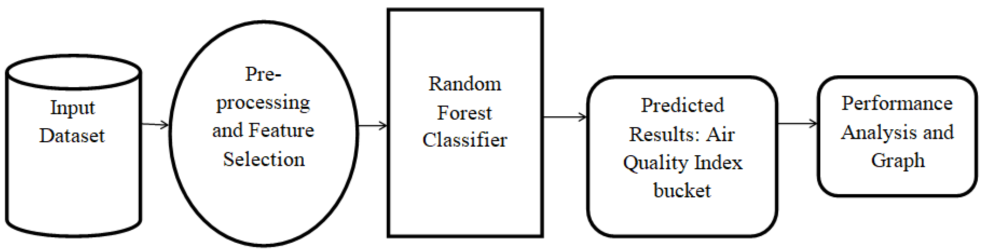

Figure 1. System Architecture

Editors: Ratnakirti Roy, Manish Kumar Thakur, and Anushree Raj
DOI:10.48001/978-81-966500-7-0-9 | ISBN: 978-81-966500-7-0

4 Module Description

- Data Collection: Collecting information is the significant beginning move toward fostering an AI model, as it significantly influences the model's performance. Various techniques, such as web scraping or manual interventions, can be employed for data collection. In this project, the dataset was sourced from the popular repository Kaggle and is stored in the model folder. Accessing high-quality and abundant data is paramount for enhancing the model's predictive capabilities.
- Model Selection: The Random Forest Classifier technique was chosen for this project due to its remarkable performance, achieving a 99% training accuracy. This algorithm's resilience and ability to handle complicated datasets well made it appropriate for implementation.
- The Random Forest Algorithm: To grasp the Random Forest algorithm in simpler terms, imagine planning a trip to a destination you'll enjoy. You could browse online, read reviews on travel platforms, or seek recommendations from friends. Let's say you opt to consult your friends, each of whom shares their past travel experiences and suggests various places. You compile a list of these recommendations and then ask your friends to vote for the best destination from this list. The destination with the highest number of votes becomes your final choice for the trip. Similar to this, the Random Forest method produces reliable and accurate results by combining predictions from several decision trees into a single final forecast.(see figure 2)

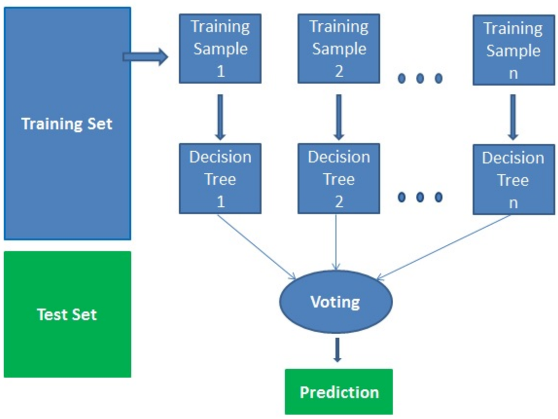

Figure 2. Algorithm

DOI:10.48001/978-81-966500-7-0-9 | ISBN: 978-81-966500-7-0

5 Result

An exceptional component decision marker is furthermore given by unpredictable boondocks. An additional variable that shows the overall significance or responsibility of every component to the figure is outfitted by Scikit-advance close by the model. During the planning stage, it normally concludes every part's significance score. From there on out, the relevance is diminished until the total score is 1. While encouraging a model, this score will help you in picking the most fundamental characteristics and shedding the less pressing ones.(see figure 3)

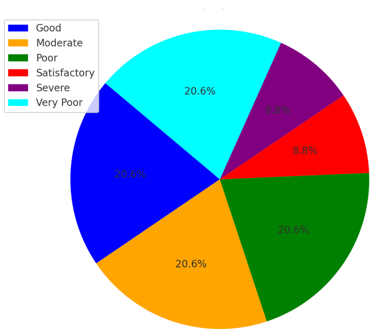

Figure 3. Chart

6 Conclusion

Based on available data, the model described in this study has the important feature of being able to predict air quality under a variety of scenarios. Appropriate management measures can be put in place to improve ambient air quality by predicting conceivable air contamination situations. As a valuable aide for impending examinations in this field, the recommended air quality expectation model shows considerable expansions in expectation precision. The Irregular Backwoods Classifier shows guarantee as a procedure for foreseeing air quality since it can create exact and steady outcomes.

It is a suitable option because of its ability to manage a large number of input variables and reduce overfitting. Nonetheless, it is critical to recognize that the Random Forest Classifier's performance can be impacted by a number of factors, including the sort and volume of preparing information, the decision of hyperparameters, and the specific issue climate. To completely understand the capability of this methodology for air quality expectation, more examination and improvement are vital.

DOI:10.48001/978-81-966500-7-0-9 | ISBN: 978-81-966500-7-0

References

Assessment and prediction of air quality using fuzzy logic and autoregressive models. (2012). Atmospheric Environment, 60, 37–50. https://doi.org/10.1016/j.atmosenv.2012.06.004

Cifuentes, J., Marulanda, G., Bello, A., & Reneses, J. (2020). Air temperature forecasting using machine learning techniques: A review. Energies, 13(6). https://doi.org/10.3390/en13164215

Fu, L., Li, J., & Chen, Y. (2023). An innovative decision making method for air quality monitoring based on big data-assisted artificial intelligence technique. Journal of Innovation and Knowledge, 8(2). https://doi.org/10.1016/j.jik.2022.100294

Gokhale, S., & Raokhande, N. (2008). Performance evaluation of air quality models for predicting PM10 and PM2.5 concentrations at urban traffic intersection during winter period. Science of the Total Environment, 394(1), 9–24. https://doi.org/10.1016/j.scitotenv.2008.01.020

Harishkumar, K. S., Yogesh, K. M., Gad, I., et al. (2020). Forecasting air pollution particulate matter (PM2. 5) using machine learning regression models. Procedia Computer Science, 171, 2057–2066.

Kumar, A., & Goyal, P. (2011). Forecasting of daily air quality index in Delhi. Science of the Total Environment, 409(24), 5517–5523. https://doi.org/10.1016/j.scitotenv.2011.08.069

Lei, T. M., Siu, S. W., Monjardino, J., Mendes, L., & Ferreira, F. (2022). Using Machine Learning Methods to Forecast Air Quality: A Case Study in Macao. Atmosphere, 13(9). https://doi.org/10.3390/atmos13091412

Liu, Q., Cui, B., & Liu, Z. (2024). Air Quality Class Prediction Using Machine Learning Methods Based on Monitoring Data and Secondary Modeling. Atmosphere, 15(5). https://doi.org/10.3390/atmos15050553

Machine learning algorithms to forecast air quality: a survey. (2023). Artificial Intelligence Review, 56(9), 10031–10066. https://doi.org/10.1007/s10462-023-10424-4

Maltare, N. N., & Vahora, S. (2023). Air Quality Index prediction using machine learning for Ahmedabad city. Digital Chemical Engineering, 7. https://doi.org/10.1016/j.dche.2023.100093

Singh, K. P., Gupta, S., Kumar, A., & Shukla, S. P. (2012). Linear and nonlinear modeling approaches for urban air quality prediction. Science of the Total Environment, 426, 244–255. https://doi.org/10.1016/j.scitotenv.2012.03.076

DOI:10.48001/978-81-966500-7-0-9 | ISBN: 978-81-966500-7-0