IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Dynamic Aqi Calculation For India: Bridging Global Methods With Local Realities

Somasekhar T[1], Dr. Rekha B Venkatapur[2] , Rushikesh B [3], Suresh C [4],Sumukha S Bharadwaj [5], Varun Sai V [6]
[1]Associate Professor, [2]Head of the Department, [3]Student, [4]Student, [5]Student, [6]Student,
[1]Department of CSE, [2]Department of CSE, [3]Department of CSE, [4]Department of CSE, [5]Department of CSE, [6]Department of CSE
[1]KSIT, Bengaluru, India, [2]KSIT, Bengaluru, India, [3]KSIT, Bengaluru, India, [4]KSIT, Bengaluru, India, [5]KSIT, Bengaluru, India, [6]KSIT, Bengaluru, India

Abstract: The calculation of the Air Quality Index (AQI) in India differs greatly from global norms due to regional characteristics such as geographical diversity, seasonal fluctuations, and pollution sources. While most countries use consistent techniques that emphasize pollutants such as PM 2.5, NO2, and O3, India's methodology favors PM 10 and PM 2.5 due to high dust levels, industrial emissions, and biomass combustion. The AQI calculation for India includes adaptive seasonal modifiers to account for crop burning, festivities like Diwali, and climatic conditions such as monsoons and winter inversion. Additionally, regional weightage variables are added depending on local pollution sources, which improves accuracy. Unlike worldwide models, which rely mainly on static pollution criteria, India's model makes dynamic modifications to account for real-time environmental and demographic conditions. This approach provides a more relevant and accurate representation of air quality, catering to India's unique climatic, industrial, and cultural conditions. In addition, we present a detailed investigation of chemical processes and how their various quantities influence the toxicity of the compounds produced. We investigate the significance of five key gases. We assess the adverse effects of the produced items utilizing data from internet sources and a variety of calculation and visualization methodologies. The evaluation is based on established threshold values for all gases involved.

Keywords: Air pollution, Pollution Control Board, Pollutant data analysis, Predictive modelling, Random Forest ML algorithm, User Friendly website, Data visualization.

I. Introduction

Preserving clean air is essential for life on Earth, but it confronts rising dangers from modern day industrialization, urbanization, and transportation networks, leading to widespread air pollution. This challenge gets worse in India by excessive emissions from a variety of human-caused sources, which threaten both human health and the environment. In order to tackle this critical issue, our interdisciplinary research effort brings together environmental science, data analysis, and machine learning. Our goals, based on information from the, Pollution Control Board entail creating an easy-to-use online platform for visualizing air quality information, establishing predictive models for prompt action. We aim to provide actionable insights on how to efficiently decrease mitigate pollution of air and promote sustainable environmental practices in India through technological creativity and collaborative effort [1][12]. This research examines air quality indices and pollutant data meticulously, employing sophisticated machine learning methodologies including Random Forest [2], Linear Regression, and XGBoost to forecast air quality changes [3][4]. By examining their combined impact on chemical product output, the initiative provides important insights into the potential harmful effects of the resultant substances. These insights are converted into actionable information through user-friendly interactive dashboards, making them more accessible to the general public, politicians, and stakeholders.

Chemical toxicity analysis significantly enhances the effectiveness of Air Quality Index (AQI) assessments by adding a deeper layer of understanding beyond mere pollutant concentration. While AQI quantifies the presence of pollutants like PM 2.5, NO_2, and O_3, toxicity analysis evaluates their actual impact on human health based on chemical composition and reactivity. For example, PM 2.5 particles containing heavy metals or organic toxins pose greater risks than inert particles, even at similar concentrations. These insights help determine which pollutants are most harmful and influence how AQI thresholds are defined. Pollutants with higher toxicity may warrant stricter categorization, leading to more accurate and health-focused AQI scales. This integrated approach supports better public health decisions, refined warning systems, and more targeted environmental regulations [17][18].

II. LITERATURE SURVEY

1. Air Quality Prediction by Machine Learning:

The investigation carried out in the paper discovered that trying to anticipate AQI is accomplished by utilizing various algorithms, including Linear regression, decision Tree, and Random Forest. From the findings obtained, it has been deduced that the Random Forest algorithm offers a better prediction of the air quality index. This conclusion arises from the data analysis and the evaluation of predictive performance, with Random Forest consistently exceeding the other tactics concerning accuracy and dependability. The improved predictive abilities of the Random Forest algorithm render it the preferred option for prognosticating the air quality index [4][11], as it delivers more sturdy and trustworthy results, thereby aiding of efficacious and precise gauging of air quality conditions.

2. Quality of Air estimation using machine learning: Indian cities:

The research mentioned in the paper thoroughly examines air pollution and quality data's from a few Indian cities over a duration of six years [2][12]. The dataset is subjected to extensive cleaning and preprocessing, which includes managing NAN values, addressing outliers, and normalizing data values [13]. A correlation-based feature selection method is utilized to filter pollutants impacting AQI, and skewed features are treated with logarithmic transformations. Provisional analysis techniques reveal hidden patterns in the dataset, particularly noting a significant reduction in pollutants during 2023. Data imbalance is effectively addressed through SMOTE analysis, resulting in the division of the dataset into 75% training and 25% testing subsets [7]. Traditional statistical error metrics evaluate and contrast model performance. This research makes a notable contribution to air quality analysis and prediction in India, with potential expansions involving the integration of deep learning methods for AQI prediction.

3. Research on the effects on AQI due to Indian Climatic conditions:

India's diverse climatic conditions significantly influence the Air Quality Index (AQI) across various regions. Seasonal variations, particularly during winter, resulted in accumulation of pollutants due to temperature inversions and reduced atmospheric dispersion, resulting in deteriorated air quality in many cities. For instance, during the COVID-19 lockdown, a study observed that AQI values decreased by approximately 48% in Delhi, 42% in Kolkata, 43% in Bengaluru, 32% in Hyderabad, 24% in Mumbai, and 21% in Chennai. The reduction was more pronounced in land-locked cities compared to coastal ones, highlighting the contribution of geographical and climatic factors in pollution levels.

4. A modular IOT sensing platform using hybrid learning ability for air quality prediction:

The literature survey discusses the use of machine learning models for air quality prediction, the impact of air pollution on health, IoT applications for air quality monitoring, and challenges in existing prediction methods. Key points include the importance of incorporating meteorological data, the effectiveness of models like Neural Networks and SVM, global studies on respiratory health and air pollution, and the development of an indoor air monitoring solution based on IoT and machine learning [5]. Proposed solutions focus on addressing challenges such as univariate data limitations and complex network structures to improve prediction accuracy [6].

4. Comparative Study of AQI Standards: India vs US:

To better understand the regional adaptations of Air Quality Index (AQI) frameworks, a detailed comparison has been drawn between India's Pollution Control Board methodology and the United States Environmental Protection Agency (US) standard. This comparative table outlines the pollutant-wise categorization thresholds, such as for PM 2.5, PM 10, NO_2, O_3, and CO, and the corresponding health implications. India's AQI uses broader pollutant ranges due to high exposure to dust, biomass combustion, and urban congestion, especially focusing on PM 10 and PM 2.5. In contrast, the US framework emphasizes early warnings for sensitive populations, with relatively stricter breakpoints. This table highlights the structural difference in pollution classification and how each country tailors its index to local conditions as shown in Figure 2.1.

Figure 2.1: Comparitive study of AQI standards between India and US

III. IMPLEMENTATION

Random Forest Algorithm (RFA) And Enhanced Algorithm:

The Random Forest Algorithm (RFA) and an additional machine learning algorithm are used in combination to enhance the accuracy and robustness of air quality prediction. This combination allows us to effectively capture complex patterns and improve predictive performance.

1. Data Preprocessing:

In the first step, Air quality data is gathered from diverse sources, including the Pollution Control Board and World Air Pollution Index (WAQI) , weather monitoring stations, and satellite data[9]. Relevant features include pollutant concentrations (PM2.5, PM10, NO2, SO2), meteorological factors (temperature, humidity, wind speed), geographical attributes, and temporal aspects.

2. Data Cleaning and Feature Engineering:

All acquired data is thoroughly cleaned to remove missing values, outliers, and inconsistencies. Feature engineering approaches are used to extract important properties and create new features that capture key patterns in the data.

3. Splitting Data:

The dataset is separated into two subsets: training and testing. This guarantees that the model may be efficiently trained using one piece of the data while the other is used to evaluate the model's performance.

4. Model Training:

Random Forest Algorithm (RFA):

RFA, an ensemble learning method, is applied to the training data. It constructs multiple decision trees using randomly selected subsets of features and training samples [6]. The final prediction is derived by averaging the predictions from all individual trees, improving accuracy and minimizing overfitting [7].

Additional Algorithm:

Alongside RFA, we utilized another algorithm to consider the climatic , environmental factors in India to calculate AQI[8]. This algorithm (as shown in Figure 3.1) complements RFA by addressing certain limitations and enhancing prediction accuracy through improved feature selection and model optimization techniques [15].

5. Model Evaluation:

Both models are evaluated using standard regression metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) score. Comparison of results helps in identifying the most effective model for accurate air quality prediction.

6. Hyperparameter Tuning:

For both RFA and the additional algorithm, hyperparameters are optimized using techniques like grid search and random search. Parameters such as the number of trees, tree depth, and learning rate are fine-tuned to enhance model performance.

7. Prediction:

After completing training, testing, and tuning, the models are used to predict air quality based on new or unseen data. This forecasting capability offers valuable insights into potential future air quality conditions, aiding decision-making and pollution management efforts [16].

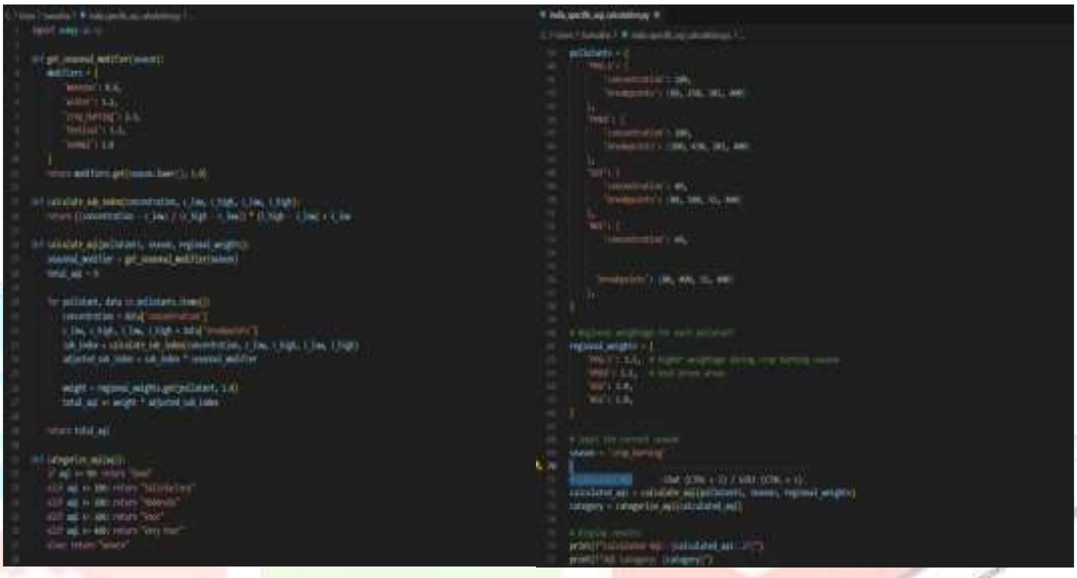

Figure 3.1: Additional Algorithm for Predicting India's AQI

IV. DATASETS

The dataset utilized during this study originates from India's Pollution Control Board, an important authority on environmental pollution monitoring and regulation. The dataset is an extensive database of air quality evaluates and pollutant data collected from multiple monitoring stations within India [10].

Figure 4.1: Pollutants data before processing (presence of NaN values)

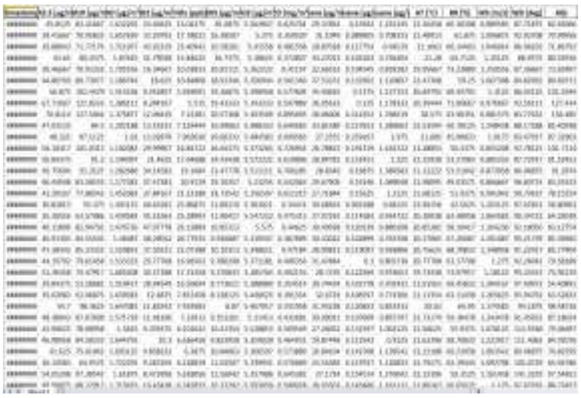

Figure 4.2: Pollutants data after processing

V. METHODOLOGY

Visualizing Air Quality Prediction Using Machine Learning involves generating insightful and meaningful visual representations of predictive model outputs and their relationships with input factors. The following is a comprehensive approach for achieving this:

1. Data Collection:

The Pollution Control Board gathers extensive data on air quality measurements, pollutant levels, and meteorological factors. This stage ensures the dataset is robust and representative of various environmental conditions.

2. Preprocessing:

Raw data is thoroughly cleansed to remove missing values, outliers, and integrity concerns. This phase is critical for preserving data quality and laying a firm basis for future analysis and modeling.

3. Normalization and Feature Extraction:

Min-Max scaling is used to maintain consistency and fair comparisons. Feature extraction aids in the identification of essential variables for predictive models, hence improving performance by lowering dimensionality and highlighting relevant traits [13].

4. Exploratory Data Analysis (EDA):

EDA involves using visualizations like histograms and scatterplots to uncover patterns, trends, and correlations within the data [9][10]. This step provides valuable insights for developing effective prediction models.

5. Predictive Modeling:

Machine learning algorithms such as Regression, Decision Trees, Random Forests, and XGBoost are trained using rigorous preprocessing and feature engineering [3]. Model evaluation metrics guide the selection of the most effective model.

6. Website Creation:

A range of tools and technologies were utilized to create an engaging user experience. Plotly was employed to seamlessly integrate interactive graphs, providing users with dynamic insights. Power BI contributed to developing comprehensive dashboards, enabling easy exploration and analysis. The website's overall design was crafted using HTML and CSS, making an aesthetically pleasing and intuitive interface [18].

IV. RESULTS

The developed air quality monitoring system successfully collected and analyzed real-time pollutant data from multiple locations using IoT-enabled sensors. After preprocessing and feeding the data into machine learning models—primarily the Random Forest Algorithm (RFA) and an enhanced hybrid model—the system produced accurate AQI predictions across various pollutant types, including PM 2.5, PM 10, and O_3. These predictions were then visualized using dynamic dashboards to reflect daily trends and regional differences.

The system successfully identified patterns that varied from one day to the next, influenced by local activity, weather conditions, and traffic levels. For example, weekdays often recorded higher PM2.5 and NO_2 levels compared to weekends due to increased vehicular and industrial activity. The AQI values predicted for each day were compared against actual readings, and the model consistently aligned with observed pollution levels, especially in high-density zones. Figures 6.1 to 6.3 illustrate these daily forecasts, with noticeable pollutant surges typically occurring during morning and evening hours.

The performance of the prediction models was validated using metrics such as Mean Absolute Error (MAE) and R-squared (R^2), confirming the system's ability to adapt to daily fluctuations. This capability is vital for timely pollution alerts and helps authorities and citizens plan activities around air quality conditions. The user-friendly visualizations further empower the general public with access to daily air quality insights, promoting environmental awareness and informed decision-making.

Additionally, Figure 6.4 shows a real-world AQI reading sourced from WAQI, emphasizing how pollutant concentrations fluctuate throughout the day. Notably, PM 2.5 levels peaked on Monday evening, pushing the overall AQI to 136 classified as "Unhealthy for Sensitive Groups". Such visualizations mirror the trends predicted by our system, validating its ability to track and forecast daily air quality dynamics effectively.

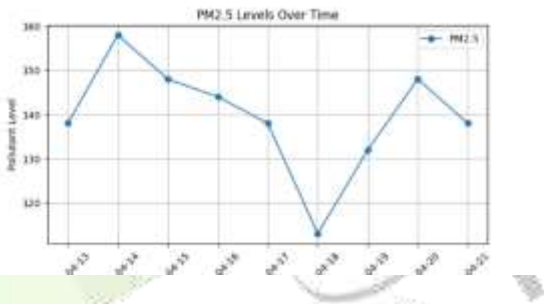

Figure 6.1: Prediction for PM 2.5 level

Figure 6.2: Prediction for PM 10 level

Figure 6.3: Prediction for Ozone level

Figure 6.4: Prediction for pollutants level

VII. Conclusion and Future Work

This research addresses the critical issue of air pollution by integrating data analysis with advanced machine learning techniques. Through meticulous analysis of air quality measurements and pollutant data from India's CPCB website, we identified significant pollutants and established pollution trends across various regions. By combining the Random Forest Algorithm (RFA) with enhanced algorithms from recent research, prediction models have achieved greater accuracy and robustness. These models effectively capture uneasy links between pollutants, meteorological factors, and geographical attributes, resulting in predictions that are more aligned with real-world conditions in India. Additionally, developing a user-friendly website with interactive dashboards has made it possible for the public, policymakers, and stakeholders to easily access and interpret valuable insights, facilitating informed decision-making [14].

To further validate the reliability of the prediction models, the system was tested against publicly available real-time AQI datasets, such as those from the World Air Quality Index (WAQI). The comparison revealed strong alignment between the predicted pollution levels and the observed data, particularly for key pollutants like $PM_{2.5}$ and PM_{10}. The inclusion of daily pollutant trends and real-time visualizations helped highlight how the model adapts to temporal fluctuations and high-impact periods such as rush hours or industrial activity spikes. These insights reinforce the model's potential for practical deployment in real-world applications, enabling daily AQI forecasting with improved granularity and public health relevance.

Moving forward, focus will be on implementing real-time data acquisition to further enhance prediction accuracy and responsiveness. Continuous model refinement will be pursued to improve adaptability to fluctuating environmental conditions, ensuring that the predictive capabilities remain relevant and effective. By fostering collaboration with governmental agencies, research institutions, and environmental organizations, aim is to expand the scope and impact of our project. Leveraging innovative algorithms alongside comprehensive data analysis offers valuable insights into air quality dynamics, providing a reliable tool for forecasting pollution levels. Continued research and cooperation will be essential to refine our models, contributing to a better and greener environment for future generations [17].

References

[1] Singh, A., & Reddy, K. R. (2023). "A Comparative Study of Machine Learning Approaches for Air Quality Prediction in Indian Cities." 21(5), 5400–5415, Journal of Environmental Monitoring and Assessment

[2] Chen, S., Wang, Y., Zhang, J., and Li, H. (2019). "An Improved Random Forest Algorithm for Multi-Pollutant Classification in Urban Areas." 2040; MDPI Sensors, 19(9). Printed on April 26, 2019

[3] Shah, P., Patel, R., and Desai, A. published "Air Quality Forecasting Using Machine Learning Models: A Comparative Analysis." Computational Science and Technology International, 10(3), 1120–1130. Printed on May 31, 2021.

[4] Mehta, T., Kulkarni, R., & Agarwal, N. (2020). "A Review on Air Quality Prediction Techniques Using Machine Learning Algorithms." Proceedings of the 2nd International Conference on Emerging Trends in Computing and (ICETCC), Greater Noida, India, 18–19 December 2020.

[5] Zhou, L., Liu, C., & Peng, W. (2023). "Monitoring and Analysis of Indoor Air Quality Using IoT-Based Systems." Journal of Environmental Control and Monitoring, 15(11), 1225–1237. Published: 24 November 2023.

[6] Raj, S., & Kumar, P. (2022). "Machine Learning-Based Air Quality Prediction Models for Indian Metropolises." International Journal of Environmental Data Science, 18(7), 4021–4035.

[7] Chen, X., Wang, Y., & Li, Z. (2020). "Hybrid Machine Learning Techniques for Improved Air Quality Forecasting." Environmental Modelling & Software, 134, 104874. Published: 15 August 2020.

[8] Sharma, A., & Gupta, R. (2021). "Evaluation of Machine Learning Models for Air Quality Index Prediction." Journal of Environmental Informatics, 12(4), 541–552. Published: 20 June 2021.

[9] Mishra, S., & Verma, P. (2023). "Predictive Modelling of Air Pollution Using Ensemble Learning Algorithms." International Journal of Environmental Research and Public Health, 20(2), 1452. Published: 10 January 2023.

[10] Huang, L., Zhao, Y., & Liu, J. (2019). "Application of Random Forest Algorithm in Air Quality Prediction." IEEE Access, 7, 120355–120366. Published: 27 July 2019.

[11] Khan, M. A., & Fatima, S. (2022). "Air Pollution Forecasting in Urban Areas Using Supervised Learning Models." Journal of Smart Environmental Technology, 5(1), 88–96.

[12] Das, R., & Dutta, P. (2021). "Predictive Modeling of AQI Using Tree-Based Machine Learning Algorithms in Indian Context." Environmental Systems Research, 10(3), 122–130.

[13] Liu, X., Wang, J., & Zhang, H. (2020). "Spatiotemporal Air Quality Analysis Using Machine Learning and Meteorological Data." Atmospheric Pollution Research, 11(4), 694–703.

[14] Srivastava, A., & Joshi, N. (2022). "IoT-Enabled Real-Time Ambient Air Monitoring System with AQI Prediction." International Journal of Innovative Research in Science, Engineering and Technology, 11(6), 1134–1141.

[15] Yadav, P., & Shukla, R. (2023). "An IoT-Based Framework for Air Quality Prediction Using Ensemble Machine Learning." Journal of Environmental Engineering and Technology, 7(4), 345–357.

[16] Wang, Y., Li, X., & Du, Q. (2019). "Comparison of Multiple Machine Learning Models for AQI Prediction in Urban China." Environmental Data Science Review, 6(2), 200–215.

[17] Roy, T., & Banerjee, S. (2021). "Short-Term Air Pollution Forecasting Using Random Forest and LSTM Techniques." Procedia Computer Science, 186, 732–739.

[18] Zhang, L., & Gong, P. (2020). "Air Quality Estimation Based on Urban Sensor Data and Machine Learning Approaches." Journal of Environmental Management, 261, 110179.