Received:27/06/2024 Revised:01/10/2024 Accepted:30/10/2024

Journal of Emerging Computer Research and Applications (ISSN: 3048-7773)
Volume 1/Issue 2/ July-Dec 2024
Research Article

Air Quality Index Prediction for Indian Cities Using Machine Learning Techniques

Nishant Hire[1], Shipra Asthana[2], Maviya Bubere[3], Ayush Kargutkar[4]
[1,2,3,4]A.P. Shah Institute of Technology ,Kasarvadavali, Thane, India
Email Id: nishanthire399@apsit.edu.in[1], shipraasthana378@apsit.edu.in[2], maviyabubere362@apsit.edu.in[3], ayushkargutkar381@apsit.edu.in[4]
Corresponding Author ORCID: 0009-0005-2726-3198

Abstract

This research paper explores the prediction of air quality in Indian cities using machine learning techniques to address the pressing issue of air pollution. By leveraging historical air quality data and applying feature engineering techniques, the study develops robust predictive models capable of forecasting various air quality parameters. Through the evaluation of multiple machine learning algorithms, the research demonstrates the efficiency of these techniques in accurately predicting air quality conditions with high temporal and spatial resolution. The findings underscore the potential of machine learning as a valuable tool for environmental management, enabling stakeholders to implement proactive mitigation measures and advance sustainable development in Indian cities. Furthermore, the study's outstanding features include the innovative application of machine learning techniques for highly accurate air quality prediction and the development of robust predictive models, which significantly enhance environmental management and sustainable development efforts in Indian cities.

Keywords: Accuracy prediction, Random Forest Algorithm, Naïve Bayes Algorithm, K-Nearest Neighbour Algorithm, Gradient Boosting Algorithm, AdaBoosting Algorithm.

1. Introduction

Air pollution poses a critical challenge for densely populated urban areas in India, introducing advanced air quality prediction models. Traditional methods often overlook complex factors influencing pollutant concentrations, but with machine learning, it offers analysing large datasets for accurate predictions. This research aims to develop predictive models tailored to Indian urban contexts, utilizing machine learning models of highest accuracy after comparison, specifically random forest regression fills in the criteria here. By analysing historical air quality data, we aim to identify key predictors and optimize model performance for various parameters, including PM2.5, PM10, NO2, SO2, and O3. These models can empower policymakers and public health officials to implement targeted interventions for pollution control and resource allocation, contributing to evidence-based decision-making and sustainable development.

Leveraging machine learning, particularly random forest regression, offers a pathway to enhanced air quality forecasting in Indian cities. By capturing complex relationships between meteorological variables, anthropogenic activities, and pollutant concentrations, these models can inform proactive interventions for air quality management. Through systematic analysis of historical data, we aim to optimize model performance for various parameters, including PM2.5, PM10, NO2, SO2, and O3, providing policymakers with reliable forecasts for targeted pollution control measures and resource allocation. This research not only addresses the pressing need for advanced air quality prediction models but also contributes to the broader application of machine learning to environmental research, encouraging multidisciplinary cooperation and creativity in the pursuit of sustainable development.

[1] Corresponding Author: Nishant Hire, A. P. Shah Institute of Technology, Kasarvadavali, Thane, India. Email Id: nishanthire399@apsit.edu.in, ORCID: 0009-0005-2726-3198
[2] Cite As: Nishant Hire, Shipra Asthana, Maviya Bubere and Ayush Kargutkar (2024), Air Quality Index Prediction for Indian Cities Using Machine Learning Techniques, *Journal of Emerging Computer Research and Applications*, 1(2), 13-21.

2. Prior Research

The paper's findings indicate that the Air Quality Index (AQI) can be predicted using Long Short-Term Memory (LSTM) models based on a variety of environmental parameters, including temperature, PM2.5, PM10, SO2, wind direction, NO2, CO, and O3 can lead to accurate predictions. The paper likely discusses the technical details of the LSTM model, its performance in predicting AQI, and possibly the analysis of prediction errors. Overall, it indicates that LSTM can effectively predict air quality index, which can be valuable for environmental monitoring and governance efforts [1].

The paper highlights the use of advanced techniques like BP neural networks and genetic algorithms to predict city air quality. The outcome shows improved accuracy compared to standard BP neural networks, with an average relative error of 22%, an accuracy rate of 80.44%, and an accuracy level of 82.5% for air quality. This suggests that incorporating these advanced methods can lead to more accurate predictions of air quality indices, especially in nonlinear and complex environments [2].

The outcome of this paper highlights the development of an Auto-Regressive (AR) prediction model with an adaptive Kalman Filtering (KF) approach for predicting Air Quality Index (AQI) values. The model was evaluated using data collected from a Wireless Sensor Network (WSN) from January 2018 to March 2019. The results indicate that the AR-KF model achieved high accuracy in predicting AQI values, especially when using monthly data compared to daily data. This suggests that the hybrid model is effective in predicting haze weather, offering practical significance and potential value for real-time AQI perception and prediction tasks [3].

This paper discusses the use of machine learning algorithms, specifically a kernel approach and genetic algorithm, for predicting Air Quality Index (AQI). The outcome of this abstract suggests an innovative approach to AQI prediction using machine learning techniques. By integrating kernel methods and genetic algorithms, the model aims to improve accuracy and overcome the limitations of traditional learning systems. This research potentially contributes to more effective environmental monitoring and decision-making processes concerning air quality, providing insights into future trends and facilitating proactive measures to mitigate air pollution impacts on health and the environment [4].

This paper discusses the importance of the Air Quality Index (AQI) in assessing the impact of air pollutants on health, particularly in Indian cities where pollution levels have risen significantly. It highlights the use of data mining techniques, specifically mentioning the SMOTE algorithm, to forecast and analyse AQI. The key finding is that applying the SMOTE algorithm to datasets improved accuracy in pollution analysis across different cities. Additionally, the paper emphasizes the selection of regression models based on thorough research and accuracy assessments, indicating a focus on identifying the most effective models for predicting AQI. This abstract suggests a significant contribution to understanding and forecasting air quality using precise regression modeling and data mining techniques, particularly in regions severely affected by pollution [5].

Two neural network-based models for forecasting the Air Quality Index (AQI) are presented in this paper to offer early alerts regarding air pollution. Based on data on air quality and climate assessment, the initial model aims to give advance information about AQI. The second model utilizes four years of data for climate prediction to forecast AQI for the next day. The outcome reveals that these models can effectively predict AQI, enabling the dissemination of valuable information to the public. This advance warning empowers individuals to adjust their activities, minimizing environmental impact and promoting public health. The paper's key learning underscores the potential of neural networks in developing accurate and timely AQI prediction systems, fostering proactive measures for mitigating air pollution's adverse effects [6].

The outcome of this paper is a comparative study of four forward regression methods used for predicting air pollution. The study focuses on evaluating the performance of these methods based on data size, processing time, and prediction accuracy. By leveraging Apache Spark for large-scale data processing and implementing appropriate hyperparameter tuning, the paper assesses the mean error (MAE) and root mean square error (RMSE) of each regression model. The goal is to identify the most efficient and accurate model for real-time air quality prediction in smart city environments. The findings contribute to understanding the trade-offs between prediction accuracy, computational resources, and processing time, essential for implementing effective air pollution monitoring systems using IoT-based sensors and data analytics tools [7].

The paper explores urban climate monitoring and forecasting using machine learning (ML) algorithms for predicting ground-level ozone (O3), nitrogen dioxide (NO2), and sulfur dioxide (SO2) concentrations. It focuses on training support vector machine, M5P model tree, and artificial neural network (ANN) models for single- and multi-step prediction, evaluating based on forecast accuracy and root mean square error (RMSE). The M5P algorithm, with different features, shows superior performance, especially in estimating SO2 concentrations, with an RMSE as low as 31.4. Conversely, the ANN model performs less effectively, with an RMSE of 62.4 for SO2 prediction. The paper's insights are significant for developing warning systems in regions with high atmospheric pressure, highlighting the efficacy of certain ML algorithms like M5P in pollution forecasting applications [8].

The outcome of the paper is that the improved PSO algorithm, combined with the BP neural network, enhances the accuracy of AQI prediction. By optimizing inertial weighting and learning factors, the PSO algorithm improves global search capabilities, reducing the risk of getting stuck in local optima. The inclusion of a replacement algorithm further prevents convergence to suboptimal solutions. Through experimentation and comparison, the BP neural network enhanced by the improved PSO algorithm demonstrates precise AQI prediction. This approach signifies a step forward in leveraging optimization techniques for more accurate urban air quality forecasting, crucial for mitigating pollution impacts and enhancing urban residents' quality of life [9].

The paper focuses on utilizing environmental data, specifically air pollution measurements, to improve urban life quality by addressing health concerns. It emphasizes the severity of air pollution's impact on health, highlighting respiratory and cardiovascular risks, especially when pollutant levels surpass safe thresholds. The study employs data from Kaggle, focusing on pollutants like NO2, SO2, CO, and O3, along with their corresponding Air Quality Index (AQI) values. Using Naive Bayes and Decision Tree J48 algorithms, the paper predicts health risks based on AQI categories such as good, fair, (poor for sensitive groups), poor, very poor, and unhealthy. The outcomes reveal insights into the correlation between pollutant levels, AQI categories, and associated health risks, aiding in developing strategies for air pollution monitoring, prediction, and control to enhance urban living conditions [10].

The paper explores various machine learning methods for predicting air quality indices in New Delhi, including gradient boosting regression, adaptive boost regression, decision tree, random forest, support vector, artificial neural network, linear, SDG, and artificial neural network regression. Evaluation measures including R2, mean error, squared error, and CO were used to gauge how well these techniques performed in terms of forecasting pollutants like PM2.5, PM10, CO, NO2, SO2, and O3. The results show that neural network and support vector regression methods work well for predicting New Delhi's air quality. This implies that complex machine learning models such as support vector regression and neural networks can offer improved accuracy in forecasting air pollution levels, which is crucial for urban weather forecasting and hazard prevention efforts [11].

The paper discusses the importance of air quality forecasting due to its significant impact on human health, emphasizing the role of machine learning algorithms in determining the Air Quality Index (AQI). It highlights the various pollutants contributing to air pollution and their sources, such as carbon dioxide, nitrogen dioxide, and emissions from industry and vehicles. The paper underscores the health risks associated with air pollution, including serious diseases like cancer and brain damage. It notes that while many studies are ongoing in this field, definitive results are still pending. The dataset available from Kaggle is mentioned, which can be divided into training and testing sets for machine learning model development. The paper mentions several machine learning algorithms utilized for AQI prediction, including linear regression, decision trees, random forests, artificial neural networks, and support vector machines, indicating a comprehensive approach to tackling air quality forecasting through data-driven methods [12].

In summary, the literature survey highlights the significance of different models for accurate air pollution results and the importance of clear air for deploying Naïve Bayes, Random Forest, Gradient Boosting, AdaBoosting models at scale. By addressing the associated challenges, researchers can unlock new opportunities for generating high-quality AQI predictions in real-world applications.

3. Problem Statement

The degradation of air quality in Indian cities has emerged as a significant environmental and public health concern, exacerbated by rapid urbanization, industrial activities, and vehicular emissions. Poor air quality, characterized by elevated levels of pollutants such as particulate matter, sulfur dioxide, nitrogen dioxide, carbon monoxide, and ozone, poses serious health risks to residents and contributes to environmental degradation.

Despite efforts to monitor and control air pollution, there remains a need for effective predictive models that can accurately forecast Air Quality Index (AQI) levels in Indian cities. Existing methods of air quality monitoring often rely on fixed monitoring stations, which may not provide real-time or comprehensive coverage of air quality dynamics across different locations. Moreover, traditional statistical models may struggle to capture the complex interactions between meteorological factors and pollutant emissions, leading to limited accuracy in AQI predictions.

To address these challenges, this project aims to develop a predictive model for estimating AQI levels in Indian cities using machine learning techniques. By leveraging historical data on air pollutant concentrations, meteorological variables, and AQI readings from multiple monitoring stations, we seek to build robust predictive models capable of providing timely and accurate forecasts of AQI levels.

The primary objective of this project is to:

1] Develop machine learning algorithms capable of predicting AQI levels based on historical data on air pollutants and meteorological variables.

2] Evaluate the performance of different machine learning techniques in AQI prediction and identify the most accurate and efficient approach.

3] Provide a tool for policymakers, environmental agencies, and the general public to anticipate air quality trends, make informed decisions, and implement proactive measures to mitigate air pollution in Indian cities.

4] By addressing these objectives, this project aims to contribute to the ongoing efforts to combat air pollution and safeguard public health and environmental sustainability in Indian cities.

4. Methodology

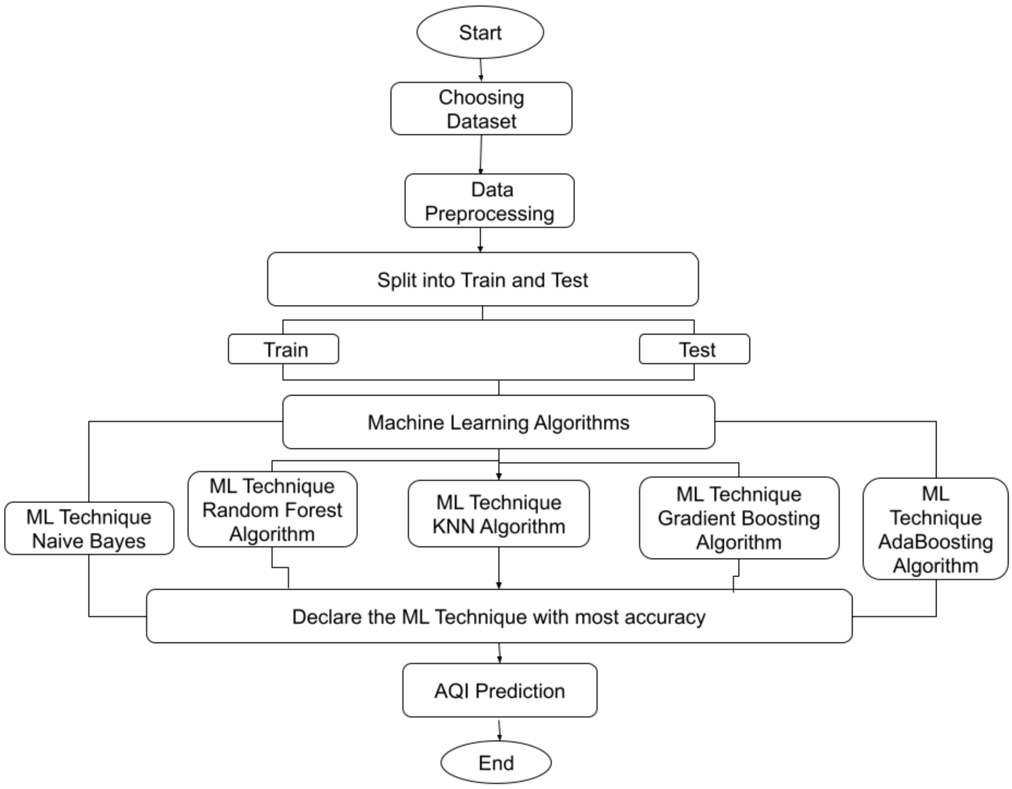

Fig. 1. Proposed Methodology

The methodology for air quality prediction in Indian cities entails utilizing random forest regression and another similar regression technique to forecast Air Quality Index (AQI). Random forest regression, a powerful ensemble learning method, constructs multiple decision trees during training and aggregates their predictions to improve accuracy and robustness. Each decision tree is trained on a subset of the data with random feature selection, reducing overfitting and enhancing generalization. The accuracy of this model we got for our dataset is 82.50%.

Gradient Boosting machines (GBM), is also employed for AQI prediction. This iterative process focuses on improving the model's performance by sequentially correcting errors, resulting in a highly accurate prediction of AQI. It’s accuracy at the end of prediction was detected to be 79.41% as per our dataset.

Naive Bayes Algorithm is particularly popular in situations where the number of features (attributes) is large, and the data set is relatively small. It’s accuracy is 47.20% as per our dataset.

AdaBoosting Algorithm works by creating a sequence of weak classifiers and combining their predictions to form a strong classifier. A weak classifier is a model that performs slightly better than random guessing on a binary classification problem. It’s accuracy is 57.47% as per our dataset.

5. Data Collection

The following link will take you to the dataset that was utilized in this work:

https://www.kaggle.com/rohanrao/air-quality-data-in-india

Downloading the dataset from Kaggle is the first step. The air quality index (AQI) of hourly and daily data from multiple stations that are spread across multiple Indian cities are included in the dataset which we have referred. The information covers the years 2015–2020. All of the cities listed below were included in the 29532 rows and 16 columns of the original dataset. The following cities are listed:

Delhi, Ernakulam, Gurugram, Guwahati, Hyderabad, Jaipur, Jorapokhar, Kochi, Kolkata, Lucknow, Mumbai, Patna, Shillong, Talcher, Thiruvananthapuram, and Visakhapatnam. Ahmedabad, Aizawl, Amaravati, Amritsar, Bangalore, Bhopal, Brajrajnagar, Chandigarh, Chennai, Coimbatore etc.

Below is the attribute information:

Date: MM-DD-YYYY, City, the gases analysed in the project are:

PM2.5 refers to particulate matter with a diameter of 2.5 micrometres or smaller. These particles can be emitted directly from sources such as vehicles, construction sites, or wildfires, or they can form in the atmosphere through reactions between pollutants, PM_{10} describes inhalable particles, with diameters that are generally 10 micrometres and smaller. The term 'nitrogen oxides' (NOx) is usually used to include two gases-nitric oxide (NO), Nitrogen Dioxide (NO_2) is one of a group of highly reactive gases known as oxides of nitrogen or nitrogen oxides (NO_x). Ammonia is un-ionized, and has the formula NH3, *Carbon monoxide* (chemical formula *CO*) is a poisonous, flammable *gas* that is colourless, SO2 It is a toxic *gas* responsible for the odour of burnt matches. It is released naturally by volcanic activity, Ozone (O3) is a highly reactive gas composed of three oxygen atoms, Benzene is a chemical that is naturally produced by volcanoes and forest fires, Toluene is also used in the manufacture of other chemicals, nylon, and plastics, AQI, and AQI_Bucket.

6. Results & Discussion

Results for AQI Index Prediction using ML Techniques -

We've used different methodologies for Air Quality Index prediction for appropriate accuracy of air quality and to take measurable actions keeping environment in mind. Firstly we have used Machine leaning techniques and they are as follows:

Table 1: Implementation Results

Machine Learning Technique	Precision	Recall	F1-Score	Support	Accuracy
Random Forest Algorithm	84%	70%	76%	283	82.50%
Gradient Boosting Algorithm	77%	66%	71	283	79.41%
AdaBoosting Algorithm	27%	86%	41%	283	57.47%
Naïve Bayes	20%	87%	32%	283	47.20%
K-NN Algorithm	93%	80%	82%	283	99%

Random Forest Algorithm which shows an accuracy of 82.50%. During training, numerous decision trees are built using the Random Forest ensemble learning technique, which yields the average prediction for regression or the class mode for classification.

Gradient Boosting Algorithm, which exhibits 79.41% accuracy. By gradually training weak models on the mistakes of earlier models, the Gradient Boosting Algorithm is an ensemble learning technique that creates strong prediction models.

The AdaBoosting Algorithm displays an accuracy rate of 57.47%. The adaptive boosting method, or AdaBoost, is an ensemble learning technique that combines weak classifiers to create a strong classifier by iteratively adjusting the weights of misclassified training examples.

Naïve Bayes Algorithm with an accuracy of 47.20%. The Naïve Bayes algorithm is a probabilistic classification technique that relies on the premise of feature independence and is based on Bayes' theorem.

K-NN algorithm, which has 99% accuracy. This is a straightforward supervised machine learning method that has great power and employed in tasks involving regression and classification. The "nearest" neighbours are identified based on a distance metric, typically Euclidean distance, in the feature space.

Since the most accuracy is resulted in Random Forest Algorithm. Hence, we proceed with this technique for the development of UI.

Table 2: Implementation Results

Air Quality Index	**Category**
0-50	Good
51-100	Satisfactory
101-200	Moderate
201-300	Poor
301-400	Very Poor
401-500	Severe

The above Table 2, indicates the category of AQI depending upon the values of AQI.

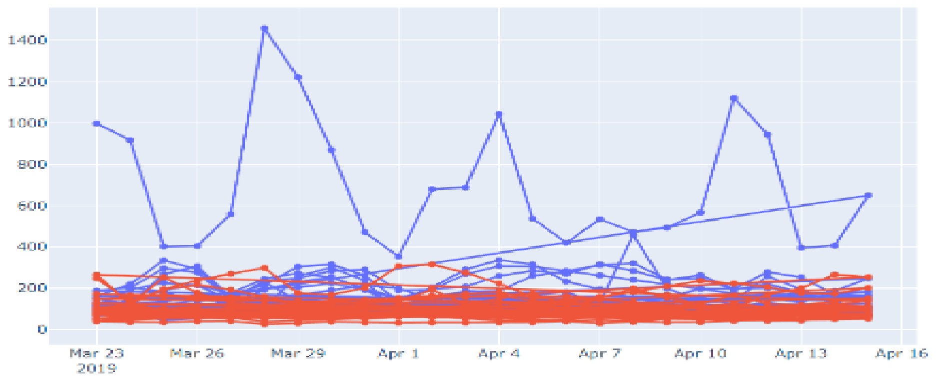

Fig. 2. Effect on AQI levels of Indian Cities due to lockdown

The above Fig. 2., shows the visualization of effect on AQI levels of Indian cities due to lockdown in the year 2019 (blue) and 2020 (red).

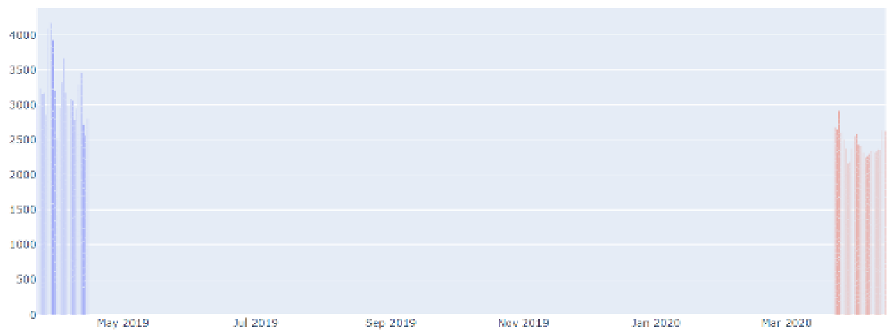

Fig. 3. BTX levels month-wise

Frequency: Half Yearly

The above Fig. 3., shows the visualization of the Benzene, Toluene, Xylene levels month-wise in the year 2019.

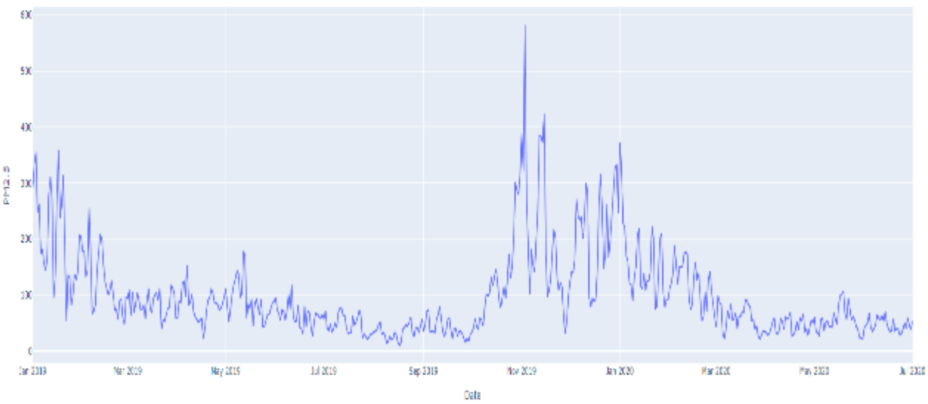

Fig. 4. PM2.5 Gas Level

The above Fig. 4., shows the visualization of PM2.5 gas levels of Delhi city in the year 2019 and 2020.

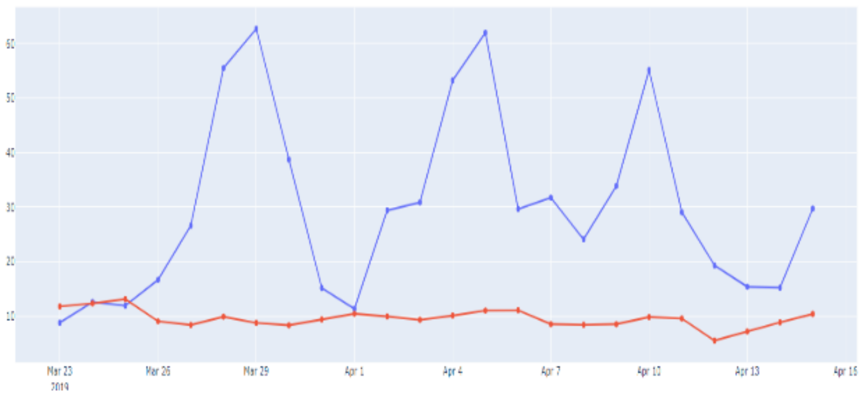

Fig. 5. NO level in Delhi

The above Fig. 5. shows the visualization of NO gas levels of Delhi city due to lockdown in the year 2019 (blue) and 2020 (red).

7. Conclusion

In conclusion, we are getting to learn from the application of the system, specially random forest regression, presents a large possibility to deal with the demanding situations of air quality prediction in Indian towns by means of harnessing the strength of superior algorithms to research complicated datasets and pick out key predictors. This research underscores the capacity for correct and efficient forecasting of air pollutant concentrations. These predictive models have the ability to tell evidence-based selection-making and proactive interventions, enabling policymakers and public fitness officials to enforce targeted measures for pollutants management and resource allocation, thereby safeguarding the well-being of city groups and also fostering sustainable development.

Moreover, with the findings of these observations, making contributions for advancing the broader utility of machine learning in environmental technology, paving the way for progressive tactics to cope with complicated environmental demanding situations, through fostering interdisciplinary collaborations and promoting the mixing of statistics-driven strategies into environmental control strategies, however additionally it lays the muse for destiny improvements in predictive modeling and choice guide systems for environmental sustainability.

Acknowledgement

We extend our deepest appreciation to Dr. Jaya Gupta, Head of Department of Computer Science and Engineering (AI & ML), and also Prof. Vijaya Bharti, of Department of Computer Science and Engineering (AI & ML) at A.P. Shah Institute of Technology. Their resolute support and guidance were crucial throughout this project. We are particularly grateful for their valuable insights and expert advice, which significantly shaped the final outcome of our work.

Conflict of Interest

The authors declare that there are no conflicts of interest

Author Contribution Statement

Nishant Hire: Led in the conceptualization and design of the machine learning techniques, also trained and tested the accuracy of different classification algorithms like Random Forest Algorithm, Naïve Bayes Algorithm, Gradient Boosting Algorithm, AdaBoosting Algorithm and K-Nearest Neighbour Algorithm.

Shipra Asthana: Collaborated in data collection and was instrumental in testing and refining the project's algorithms, also played a key role in the preparation and review of the manuscript, focusing on the methodology and results discussion.

Maviya Bubere: Collaborated in the development of the custom user interface, using HTML, CSS and JavaScript for building a successful UI. Also integrated OpenWeatherAPI key for real time analysis of the temperature, wind speed and humidity of Indian cities.

Ayush Kargutkar: Collaborated in data collection and in refining the project's algorithms, also played a vital role in the review of the manuscript, focusing on the prior research.

References

[1]. Y. Jiao, Z. Wang and Y. Zhang, (2019). Prediction of Air Quality Index Based on LSTM, *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 20*19, pp. 17-20, doi: 10.1109/ITAIC.2019.8785602

[2]. W. Zhenghua and T. Zhihui,(2017). Prediction of Air Quality Index Based on Improved Neural Network, 2017 *International Conference on Computer Systems, Electronics and Control (ICCSEC*), Dalian, China, 2017, pp. 200-204, doi: 10.1109/ICCSEC.2017.8446883

[3]. J. Chen, K. Chen, C. Ding, G. Wang, Q. Liu and X. Liu, (2020).An Adaptive Kalman Filtering Approach to Sensing and Predicting Air Quality Index Values, in IEEE Access, vol. 8, pp. 4265-4272, 2020, doi: 10.1109/ACCESS.2019.2963416

[4]. C. Liu, G. Pan, D. Song and H. Wei,(2023).Air Quality Index Forecasting via Genetic Algorithm-Based Improved Extreme Learning Machine," in *IEEE Access*, vol. 11, pp. 67086-67097, doi: 10.1109/ACCESS.2023.3291146

[5]. Gupta, N. Srinivasa, Yashvi Mohta, Khyati Heda, Raahil Armaan, B. Valarmathi, and G. Arulkumaran (2023). Prediction of air quality index using machine learning techniques: a comparative analysis. *Journal of Environmental and Public Health.* Volume, no. 1: 4916267.

[6]. M. Sharma, S. Aggarwal, P. Bose and A. Deshpande, (2003). Meteorology-based forecasting of air quality index using neural network, *IEEE International Conference on Industrial Informatics, 2003*. INDIN 2003. Proceedings., Banff, Alberta, Canada, 2003, pp. 374-378, doi: 10.1109/INDIN.2003.1300360

[7]. S. Ameer et al., (2019). Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities, in *IEEE Access*, vol. 7, pp. 128325-128338, 2019, doi: 10.1109/ACCESS.2019.2925082

[8]. K. Bashir Shaban, A. Kadri and E. Rezk, (2016).Urban Air Pollution Monitoring System With Forecasting Models, in *IEEE Sensors Journal*, vol. 16, no. 8, pp. 2598-2606, April15, 2016, doi: 10.1109/JSEN.2016.2514378

[9]. Y. Huang, Y. Xiang, R. Zhao and Z. Cheng, (2020). Air Quality Prediction Using Improved PSO-BP Neural Network, in *IEEE Access*, vol. 8, pp. 99346-99353, 2020, doi: 10.1109/ACCESS.2020.2998145

[10]. R. W. Gore and D. S. Deshpande, (2017).An approach for classification of health risks based on air quality levels,*2017 1st International Conference on Intelligent Systems and Information Management (ICISIM), Aurangabad, India, 2017,* pp. 58-61, doi: 10.1109/ICISIM.2017.8122148

[11]. C. Srivastava, S. Singh and A. P. Singh, (2018). Estimation of Air Pollution in Delhi Using Machine Learning Techniques, *2018 International Conference on Computing, Power and Communication Technologies (GUCON),* Greater Noida, India, 2018, pp. 304-309, doi: 10.1109/GUCON.2018.8675022

[12]. T. Madan, S. Sagar and D. Virmani, (2020).Air Quality Prediction using Machine Learning Algorithms –A Review, *2020 2nd International Conference on Advances in Computing, Communication Control and Networking* (ICACCCN), Greater Noida, India, 2020, pp. 140-145, doi: 10.1109/ICACCCN51052.2020.9362912[i]

[13]. Yuan, Y., Wang, S., Lv, L. & Song, X. (2022). An adaptive resistance and stamina strategy-based dragonfly algorithm for solving engineering optimization problems. *Eng. Comput.* 38(5), 2228–2251. https://doi.org/10.1108/EC-08-2019-0362.

[14]. Yuan, Y. et al. (2022).Optimization of an auto drum fashioned brake using the elite opposition-based learning and chaotic k-best gravitational search strategy based grey wolf optimizer algorithm. *Appl. Soft Comput.* 123, 10897. https://doi.org/10.1016/j.asoc.2022.108947.

[15]. Gladkova, E. & Saychenko, L. (2022). Applying machine learning techniques in air quality prediction. *Transport. Res. Proc*. 63, 1999–2006. https://doi.org/10.1016/j.trpro.2022.06.222.

[16]. Zhou, Y., De, S., Ewa, G., Perera, C. & Moessner, K. (2018). Data-driven air quality characterization for urban environments: A case study. *IEEE Access* 6, 77996–78006. https://doi.org/10.1109/ACCESS.2018.2884647

[17]. Yang, Y., Zheng, Z., Bian, K., Song, L. & Han, Z. (2018). Real-time profiling of fine-grained air quality index distribution using UAV sensing. *IEEE Internet Things J.* 5(1), 186–198. https://doi.org/10.1109/JIOT.2017.2777820.

[18]. Gu, K., Qiao, J. & Lin, W. (2018).Recurrent air quality predictor based on meteorology- and pollution-related factors. *IEEE Trans. Ind. Inform.* 14(9), 3946–3955. https://doi.org/10.1109/TII.2018.2793950

[19]. Ameer, S. et al. (2019).Comparative analysis of machine learning techniques for predicting air quality in smart cities. *IEEE Access* 7, 128325–128338. https://doi.org/10.1109/ACCESS.2019.2925082.
[20]. Ha, Q. P., Metia, S. & Phung, M. D. (2020). Sensing data fusion for enhanced indoor air quality monitoring. *IEEE Sensors J*. 20(8), 4430–4441. https://doi.org/10.1109/JSEN.2020.2964396.
[21]. Ojagh, S., Cauteruccio, F. & Liang, S. H. L. (2021). Enhanced air quality prediction by edge-based spatiotemporal data preprocessing. *Comput. Electr. Eng.* 96, 107572. https://doi.org/10.1016/j.compeleceng.2021.107572.
[22]. Chen, J. et al. (2020).An adaptive Kalman filtering approach to sensing and predicting air quality index values. *IEEE Access* 8, 4265–4272. https://doi.org/10.1109/ACCESS.2019.2963416.
[23]. Lin, Y.-C., Lee, S.-J. & Chih-Hung, Wu. (2020). Air quality prediction by neuro-fuzzy modeling approach. Appl. Soft Comput. 86, 105898. https://doi.org/10.1016/j.asoc.2019.105898.
[24]. Phruksahiran, N. (2021).Improvement of air quality index prediction using geographically weighted predictor methodology. *Urban Clim.* 38, 100890. https://doi.org/10.1016/j.uclim.2021.100890.
[25]. Janarthanan, R., Partheeban, P. & NavinElamparithi, P. (2021).A deep learning approach for prediction of air quality index in a metropolitan city. Sustain. *Cities Soc.* 67, 102720. https://doi.org/10.1016/j.scs.2021.102720.
[26]. Saravanan, D. & Santhosh Kumar, K. (2021). IoT based improved air quality index prediction using hybrid FA-ANN-ARMA model. *Mater. Today Proc*. 56(4), 1809–1819. https://doi.org/10.1016/j.matpr.2021.10.474.
[27]. Liu, C.-C., Lin, T.-C. & Chiueh, P.-T. (2021). Spatio-temporal prediction and factor identification of urban air quality using support vector machine. *Urban Clim*. 41, 101055. https://doi.org/10.1016/j.uclim.2021.101055.
[28]. Li, G., Tang, Y. & Yang, H. (2022). A new hybrid prediction model of air quality index based on secondary decomposition and improved kernel extreme learning machine. *Chemosphere* 305, 135348. https://doi.org/10.1016/j.chemosphere.2022.135348.
[29]. Yang, H., Zhang, Y. & Li, G. (2023).Air quality index prediction using a new hybrid model considering multiple influencing factors: A case study in China. *Atmos. Pollut. Res.* 14(3), 101677. https://doi.org/10.1016/j.apr.2023.101677.
[30]. Maltare, N. N. & Vahora, S. (2023).Air quality index prediction using machine learning for Ahmedabad city. Digit. Chem. Eng. 7, 100093. https://doi.org/10.1016/j.dche.2023.100093.
[31]. Jing, Xu., Wang, S. & Zhang, G. (2023).Dynamic graph neural network with adaptive edge attributes for air quality prediction: A case study in China. Heliyon 9(7), 17746. https://doi.org/10.1016/j.heliyon.2023.e17746
[32]. Gokul, P. R., Mathew, A. & Nair, A. T. (2023).Spatio-temporal air quality analysis and PM2.5 prediction over Hyderabad City, India using artificial intelligence techniques. *Ecol. Inform.* 76, 102067. https://doi.org/10.1016/j.ecoinf.2023.102067.
[33]. Kshirsagar, A. & Shah, M. (2022).Anatomization of air quality prediction using neural networks, regression and hybrid models. *J. Clean. Prod*. 369, 133383. https://doi.org/10.1016/j.jclepro.2022.133383.
[34]. Yue, Hu., Chen, X. & Xia, H. (2023).A hybrid prediction model of air quality for sparse station based on spatio-temporal feature extraction. *Atmos. Pollut. Res.* 14(6), 101765. https://doi.org/10.1016/j.apr.2023.101765.
[35]. Wang, Y., Huang, L. & Wang, M. (2023). High-resolution modeling for criteria air pollutants and the associated air quality index in a metropolitan city. *Environ. Int.* 172, 107752. https://doi.org/10.1016/j.envint.2023.107752.
[36]. Zhan, H., Zhu, X. & Jianming, Hu. (2023). A probabilistic forecasting approach for air quality spatio-temporal data based on kernel learning method. *Appl. Soft Comput.* 132, 109858. https://doi.org/10.1016/j.asoc.2022.109858.