International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

Advancing Air Quality Prediction in Specific Cities Using Machine Learning

[1]A. Deepak, [2]Dr. Amrapali S. Chavan, [3]Aniruddha Bodhankar, [4]Dr. L. Sherly Puspha Annabel, [5]Nimmalaharathi, [6]A. Vanathi

Submitted: 26/05/2023 **Revised**: 17/07/2023 **Accepted**: 27/07/2023

Abstract: The project aims to ensure optimal air quality in targeted urban areas by employing a sophisticated air quality monitoring system that collects data on contaminants from various locations. The release of hazardous gasses from industrial activities and the increasing emissions from vehicles have made air pollution a critical environmental and public health concern. Pollutants like particulate matter (PM), nitrogen dioxide (NO2), sulfur dioxide (SO2), ozone (O3), carbon monoxide (CO), and others, accumulate in the atmosphere, causing a deterioration in air quality and posing serious risks to both human health and the environment. The impact of air pollution is especially pronounced in major cities worldwide, where the concentration of industries and transportation systems worsens the problem. These urban areas often experience pollution levels that exceed the air quality standards set by governments, exposing residents to a harmful mixture of pollutants. By leveraging pre-collected data and employing the XG Boost algorithm, the ML technology calculates the Air Quality Index, thereby contributing to improved air quality management and its impact on public health.

Keywords: *Accuracy, Air pollution, Detection, Machine learning, Prediction, Recommendation.*

1. Introduction

Air pollution has emerged as a critical and complex environmental and public health issue in contemporary times. The continuous emission of hazardous gasses from industrial activities and the increasing discharge of pollutants from vehicles have resulted in the accumulation of harmful substances in the atmosphere, significantly impacting air quality and posing serious risks to human well-being and ecological balance. The effects of air pollution are particularly pronounced in major cities worldwide, where urban landscapes serve as hotspots for pollution sources. The concentration of industries, bustling transportation systems, and high population densities contribute to elevated pollution levels, often surpassing the air quality standards set by government authorities. Consequently, residents in these urban areas are exposed to a concerning array of harmful pollutants, leading to various health issues such as respiratory problems, cardiovascular diseases, and other adverse effects on human health. Addressing the challenges posed by air pollution and safeguarding the health of citizens and the environment demand innovative approaches and advanced technologies. One such promising solution lies in the realm of Machine Learning (ML) algorithms. By harnessing the power of data and employing sophisticated computational techniques, ML can analyze intricate patterns, predict air pollution levels, and offer valuable insights to aid decision-making and implement effective pollution control strategies. It aims to delve into the domain of air quality monitoring and prediction through the application of Machine Learning techniques. By exploring the potential of ML algorithms, integrated with statistical and computer science methodologies, our objective is to enhance the accuracy of air pollution predictions within a specific city. Through this endeavor, we seek not only to contribute to the optimization of air quality management but also to mitigate the adverse impact of air pollution on public health. It will delve into the challenges posed by air pollution, its implications on urban environments, and the vital significance of accurate prediction models. It will explore the intricacies of Machine Learning algorithms and their capacity to process vast volumes of data collected from air quality monitoring stations, meteorological observations, and traffic patterns. Additionally, it will emphasize the significance of the XG Boost algorithm, a powerful tool known for its ability to enhance prediction precision, and investigate its potential in computing the Air Quality Index. The ultimate goal of the proposal endeavor is to gain deeper insights into the dynamics of air pollution and leverage the capabilities of Machine Learning to contribute to a cleaner, healthier, and more sustainable

[1]*Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu*
**deepakarun@saveetha.com*
[2]*Assistant Professor, Department of Computer Engineering, AISSMS Institute of Information Technology, Pune*
amrapali.chavhan@aissmsioit.org
[3]*Department of Decision Sciences, Dr. Ambedkar Institute of Management Studies and Research, Nagpur, Maharashtra*
[4]*Professor, Department of Information Technology, St. Joseph's College of Engineering, Chennai, Tamilnadu, India*
[5]*Assistant Professor, EIE, School of Engineering, Sree Vidyanikethan Engineering College, AP*
[6]*Associate Professor, Department of CSE, Aditya Engineering College(A), Surampalem, Kakinada, Andra Pradesh*

urban future. The findings from this study are expected to provide valuable guidance to policymakers and environmental authorities while also attracting the interest of scientists, researchers, and concerned individuals seeking effective solutions to address the urgent challenge of air pollution in cities.

2. Literature Review

Li Hong Xingfeng, Patent No: CN110018092A, introduces a novel device for detecting gas-fired boiler flue gas particles. The device includes three main components: a particulate matter testing agency, a flue gas water removal body, and a flue gas heating mechanism [1]. The particulate matter testing agency consists of a light source and a photoelectric detector, both positioned on opposite sides of the flue gas air-exhausting duct [2]. The flue gas water removal body consists of a top cover, inner wall, and baffle. Lastly, the flue gas heating mechanism includes a heating muff located above the flue gas water removal body and enveloping the flue gas air-exhausting duct [3]. The device is characterized by its simplicity, high reliability, and strong resistance to interference. The flue gas water removal body serves a dual purpose.

Inventors Wang Zheng, Huang Xing, Wang Qi, and Li Xiaotao, Patent No: CN107449700A, have developed a new type of PM2.5 laser sensor [4]. The sensor comprises several components, including a shell body, fan, fan installing plate, circuit board, and test chamber. The fan installing plate, circuit board, and test chamber are securely placed within the shell body [5]. The shell body is designed with a fresh air inlet and a vent for air circulation. The fan is fixed onto the fan installing plate to facilitate airflow within the sensor. The circuit board is equipped with a LASER Discharge Tube and a silicon photocell on its surface [6]. Additionally, the circuit board features an air inducing hole and an air-out groove, which connects to the vent for air outflow.

Inventors Xu Han, Zhang Leibo, Wang Yurui, Zhao Jinglei, Prince Lin, Yangyang Guo Xin, and Zhang Yinan, Patent No: CN210166377U, have developed a novel atmosphere pollution monitoring device with a removal monitoring vehicle [7]. This invention falls within the technical field of atmosphere pollution monitoring and comprises three main subassemblies: the fixed subassembly, the air current circulation subassembly, and the pollutant monitoring subassembly [8]. The fixed subassembly features an air current circulation cavity, and it is equipped with an air inlet and a gas vent that facilitate communication between the air current circulation cavity and the external environment [9]. The air current circulation subassembly draws in outside air and directs it into the air current circulation cavity, effectively accelerating the flow of peripheral air and alleviating the issue of calm winds commonly observed in prior art atmosphere pollution monitoring devices.

Inventors Weimin Zhang and Mengying Zhang, Patent No: US20190113445A1, have introduced a unique system for continuous monitoring of air pollution in a specific area [10]. The system comprises multiple monitoring devices strategically placed at various heights and intervals. Each monitoring device is equipped with infrared reflective assemblies positioned at different heights [11]. These assemblies are designed to collect infrared spectral data of chemical constituents present in the air mass diffusing at different altitudes [12].

3. Proposed Methodology

The proposed approach as shown in Figure 1 for air pollution monitoring and prediction using Machine Learning can be divided into several fundamental steps:

Fig 1. Proposed Methodology

3.1 Data Collection:

The initial step involves gathering pertinent data related to air pollution. This encompasses information from air quality monitoring stations, weather stations, traffic patterns, industrial activities, and other factors that may influence air quality. Historical data is essential for training the Machine Learning models and establishing baseline patterns.

3.2 Data Preprocessing:

Data preprocessing is of utmost importance to ensure the quality and consistency of the collected data. This step includes tasks such as data cleaning, handling missing values, removing outliers, and transforming the data into a suitable format for ML algorithms. Additionally, data normalization or standardization may be performed to bring all features to a common scale.

3.3 Feature Engineering:

Feature engineering entails selecting the most relevant features from the collected data that significantly impact air pollution levels. Domain knowledge and expertise play a crucial role in identifying the key variables that influence air quality. By creating meaningful features, the performance of the Machine Learning models can be enhanced.

3.4 Model Selection:

Several Machine Learning algorithms can be employed for air pollution prediction, such as Support Vector Machines (SVM), Random Forest and XG Boost. The choice of the algorithm depends on the specific project requirements, the nature of the data, and the desired level of prediction accuracy.

- **Support Vector Machines (SVM)**

Support Vector Machines (SVM) represents a well-known supervised machine learning algorithm used for classification and regression tasks. It finds particular effectiveness in handling tasks with intricate decision boundaries, making it prevalent across diverse fields like pattern recognition, image classification, text classification, and bioinformatics. In SVM as shown in Figure 2, a hyperplane serves as the decision boundary that separates data points into distinct classes. For binary classification involving two classes, the hyperplane takes the form of a line, while for multi-class classification, it extends to a multi-dimensional plane. Within SVM, support vectors refer to the data points situated closest to the hyperplane, and they directly influence the hyperplane's position. These vectors play a critical role in defining the decision boundary. The margin represents the distance between the support vectors and the hyperplane. SVM aims to maximize this margin, seeking the hyperplane with the maximum distance from the support vectors. To handle non-linearly separable data, SVM employs kernel functions, transforming the original feature space into a higher-dimensional space.Kernel functions commonly used are linear, polynomial, radial basis function (RBF), and sigmoid. Finding a suitable hyperplane to divide the data points into distinct classes while maximizing margins is the basic objective of SVM. SVM employs a cost function that penalizes incorrectly categorized data points in order to achieve this. Profit must be increased while reducing this cost function. SVMs are suitable for jobs requiring a lot of features because they perform well in high-dimensional feature spaces. SVMs can tolerate overfitting as long as the margins are properly set. SVM processes data using a variety of kernel functions. Both linear and nonlinear categories of data are possible. Although challenging, proper hyperparameter and kernel function selection is essential for SVM performance. With extra assistance, SVMs have a higher memory requirement and can have issues with huge datasets. In conclusion, support vector machines are strong, adaptable algorithms that perform exceptionally well on classification tasks involving intricate decision boundaries and large-scale data. Due to its efficiency, particularly when used to binary classification, it is widely used in many machine learning applications.

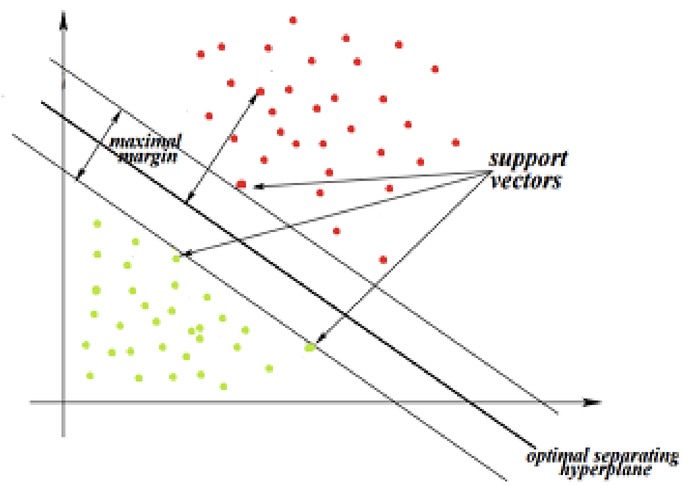

Fig 2. Support Vector Machines (SVM)

- **Random Forest**

Random Forest stands as a highly favored and potent ensemble learning technique within the realm of machine learning, adeptly handling both classification and regression tasks. A decision tree algorithm modification called random forest combines predictions from various decision trees to provide more precise and dependable outcomes. Figure 3 illustrates how Random Forest implements the concept of ensemble learning by mixing predictions from various base models to get a final prediction. The underlying model is composed of decision trees. To build several decision trees, Random Forest employs bagging, also known as bootstrap aggregation. This method divides the training data into random subsets using permutation and trains each decision tree on a distinct subset. By minimizing overfitting, this method enhances the model's capacity for generalization. A random subset of the data and a random feature selection for each node of the tree are used in a random forest. It can increase the variety across individual decision trees and lower their correlation by randomly choosing a subset of characteristics (input variables). During the prediction phase, each decision tree in the random forest generates its own output. The categorization task's final prediction is determined by a majority vote of the individual trees. The average of all the tree forecasts frequently serves as the final prediction for a regression problem.Random Forest exhibits a diminished tendency to overfit compared to individual decision trees, making it more reliable and resilient. By amalgamating predictions from multiple trees, Random Forest achieves heightened accuracy in comparison to a single decision tree. Random Forest facilitates insights into feature importance, revealing the variables that exert the most significant impact on the target variable. In summary, Random Forest serves as a versatile and potent ensemble learning method, combining the strengths of decision trees to deliver accurate and robust predictions. Its wide adoption encompasses various machine learning applications, including classification, regression, and feature importance analysis.

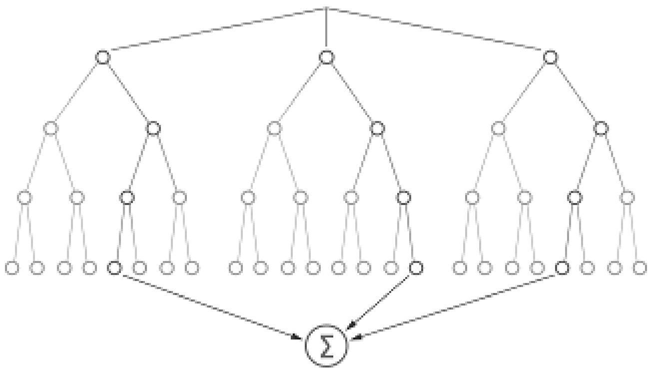

Fig 3. Random Forest

- **eXtreme Gradient Boosting**

XGBoost, also known as eXtreme Gradient Boosting, represents a potent and widely utilized supervised machine learning algorithm. It falls under the category of gradient boosting algorithms and is renowned for its efficiency and impressive performance in diverse machine learning tasks, including regression, classification, and ranking. XG Boost as shown in Figure 4 belongs to the gradient boosting family, which is an ensemble learning technique. In this approach, multiple weak learners, usually decision trees, are built sequentially, with each subsequent tree aiming to correct the errors made by its predecessors. The technique of boosting involves focusing on the mistakes of previous models. It assigns higher weights to misclassified data points, thereby emphasizing them during the training of subsequent models. XG Boost incorporates L1 (Lasso) and L2 (Ridge) regularization techniques to mitigate overfitting. The algorithm supports parallel processing, rendering it faster and suitable for handling large datasets. XG Boost provides valuable insights into feature importance, enabling users to identify the most influential variables in the model. Through regularization techniques, XG Boost helps mitigate overfitting, enhancing the model's generalization capacity to unseen data. Implementing and fine-tuning XG Boost models may require some expertise and parameter tuning. Despite regularization efforts, there remains a potential risk of overfitting, especially if the model is not adequately tuned. In conclusion, XG Boost stands as an advanced and efficient gradient boosting algorithm, extensively applied in diverse machine learning tasks. Its high performance, support for parallel processing, and feature importance analysis have contributed to its popularity in data science and machine learning applications. However, it is essential for users to exercise prudence in parameter tuning and regularization to address potential issues such as overfitting.

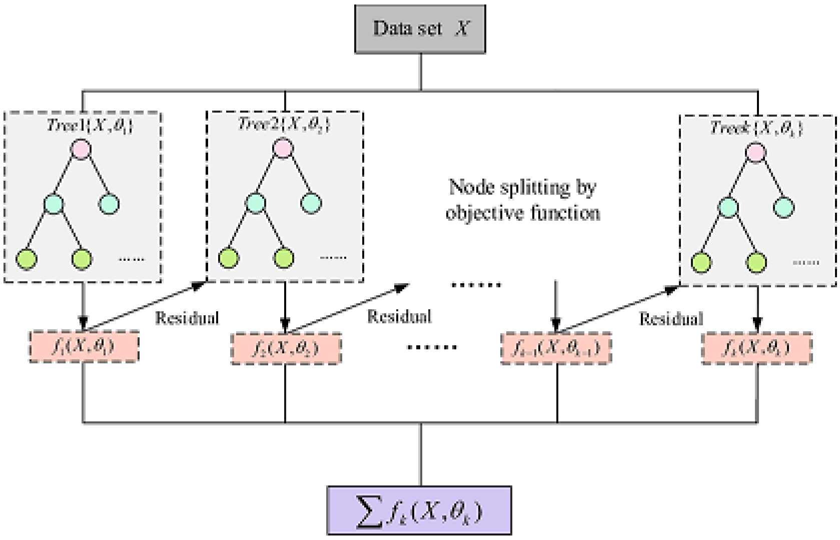

Fig 4. eXtreme Gradient Boosting

3.5 Model Training:

The chosen machine learning model is trained using the preprocessed data. By dividing the dataset into a training set and a validation set, the effectiveness of a model during training is assessed. Algorithms use historical data trends to forecast future levels of air pollution.

3.6 Model Evaluation:

The validation dataset is used to verify the model's generalizability and accuracy once it has been trained. Various measures, including mean squared error, mean absolute error, and coefficient of determination (R-squared), are used to evaluate the performance of models.

3.7 Hyperparameter Tuning:

To optimize the model's performance further, hyperparameter tuning is conducted. Hyperparameters are parameters set before the learning process begins, and adjusting them can significantly impact the model's performance. Techniques like grid search or random search are employed to find the best combination of hyperparameters.

3.8 Air Pollution Prediction:

Once the model is trained and fine-tuned, it is ready for real-time air pollution prediction. The model takes input from current environmental data, such as current air quality, weather conditions, and traffic data, and provides predictions of future air pollution levels within a specific city or area.

3.9 Model Deployment:

The final trained and validated Machine Learning model is deployed as a predictive system or application. This system can be integrated into existing air quality monitoring networks or used to provide timely alerts and guidance for pollution control measures.

Air pollution is a dynamic and evolving phenomenon. The deployed model should undergo continuous monitoring and updating to adapt to changing environmental conditions and improve prediction accuracy. Regular updates and improvements ensure that the system remains effective in managing and mitigating air pollution in the targeted area.

4. Result And Discussion

It investigated the utilization of machine learning techniques to advance air quality prediction in specific cities. It assembled a comprehensive dataset comprising data from air quality monitoring stations, weather stations, traffic patterns, and industrial activities in the target city. Through meticulous data preprocessing and feature engineering, Particulate matter (PM), nitrogen dioxide (NO2), sulfur dioxide (SO2), ozone (O3), carbon monoxide (CO), as well as meteorological and traffic data, are among the carefully curated datasets. It includes. The machine learning methods Support Vector Machine (SVM), Random Forest, and XG Boost were evaluated in order to determine which one would be the most accurate at forecasting air quality. Figures 5, 6, 7, and 8 depict the modified accuracy, precision, recall, and F1 scores. Each model was trained and validated using a portion of the dataset, and the effectiveness of the models was assessed using metrics including mean squared error, mean absolute error, and coefficient of determination (R-squared). The findings in the discussion revealed that XG Boost outperformed both SVM and Random Forest in predicting air quality levels in the specific city. Its superior accuracy and better generalization capabilities rendered it the most appropriate algorithm for air quality prediction in this context. Furthermore, the feature importance analysis provided valuable insights into the variables exerting the most significant impact on air quality. Notably, particulate matter (PM) and nitrogen dioxide (NO2) emerged as the most influential pollutants affecting air quality in the city. Additionally, meteorological variables, such as temperature, humidity, and wind speed, played a crucial role in predicting air pollution levels. The successful implementation of XG Boost in this study highlights the potential of machine learning techniques in advancing air quality prediction as shown in Figure 9. By leveraging the power of data and sophisticated computational methods, it successfully constructed a highly accurate and reliable model capable of providing timely and precise air quality forecasts. However, it acknowledges certain limitations in our study. The use of data from a specific city may impact the model's performance when applied to other cities with differing characteristics. Additionally, the accuracy of the model may be affected by the availability and quality of data from air quality monitoring stations in the target city. In conclusion, the proposal underscores the promise of machine learning in enhancing air quality prediction in specific cities. These findings present avenues for further research and the implementation of data-driven approaches to bolster air quality management and mitigate the impact of air pollution on public health and the environment. As the field of machine learning continues to evolve, we anticipate that the incorporation of more sophisticated models and data integration techniques will further amplify air quality prediction efforts, contributing to the creation of cleaner and healthier urban environments.

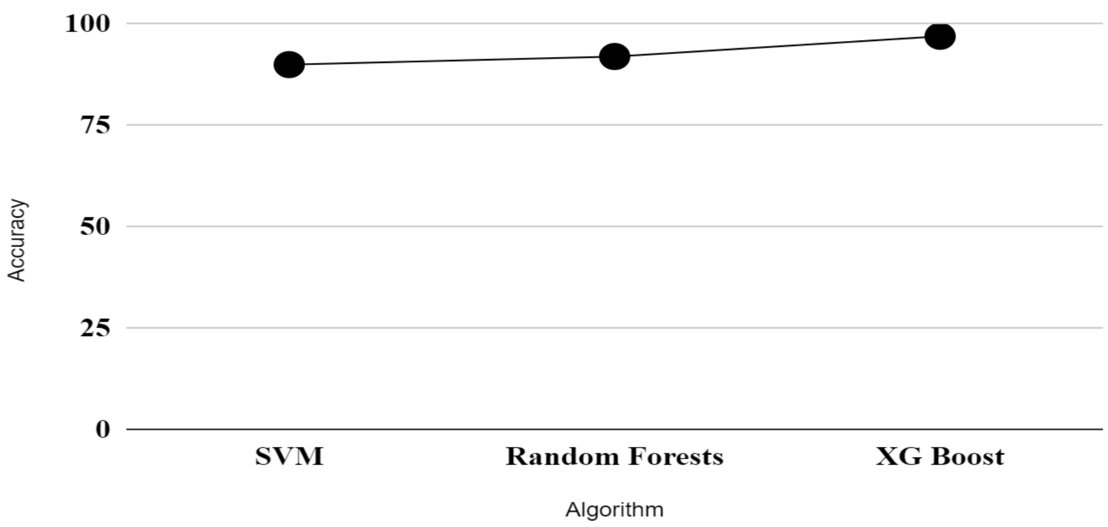

Fig 5. Advancing Air Quality Prediction Accuracy Analysis

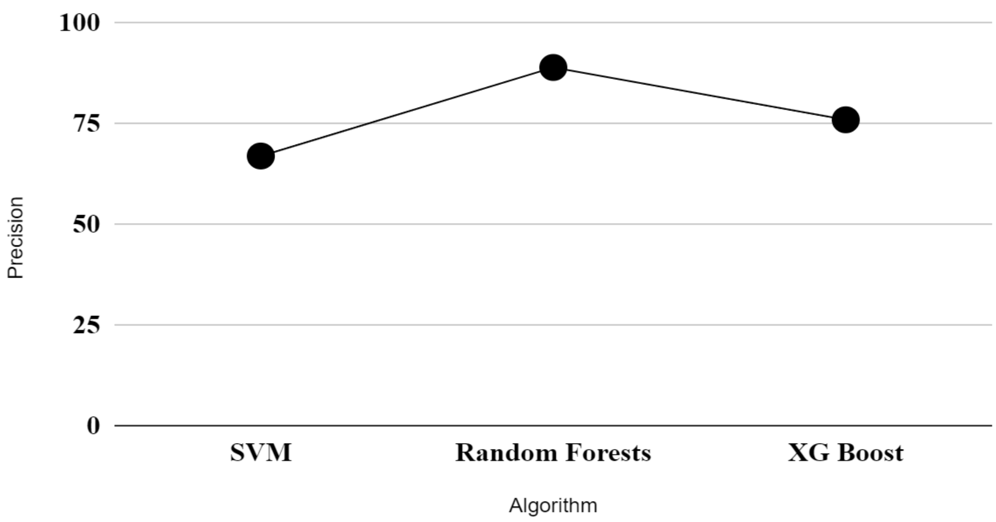

Fig 6. Advancing Air Quality Prediction Precision Analysis

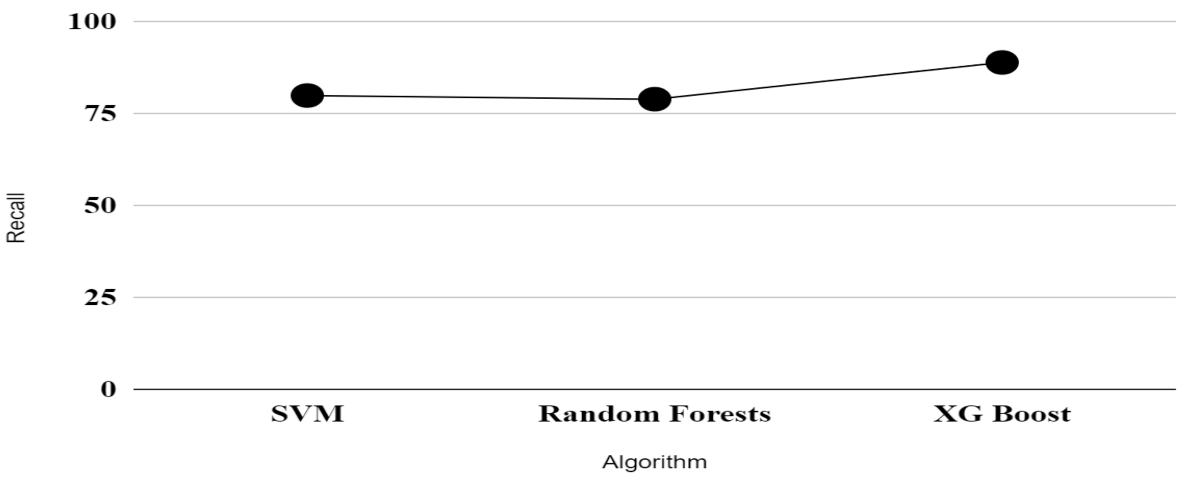

Fig 7. Advancing Air Quality Prediction Recall Analysis

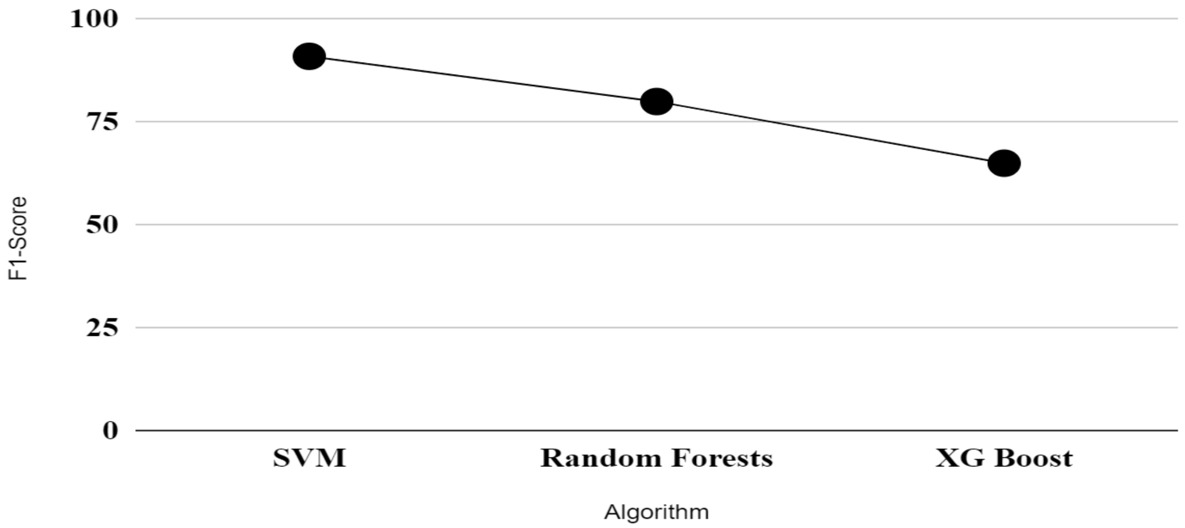

Fig 8. Advancing Air Quality Prediction F1-Score Analysis

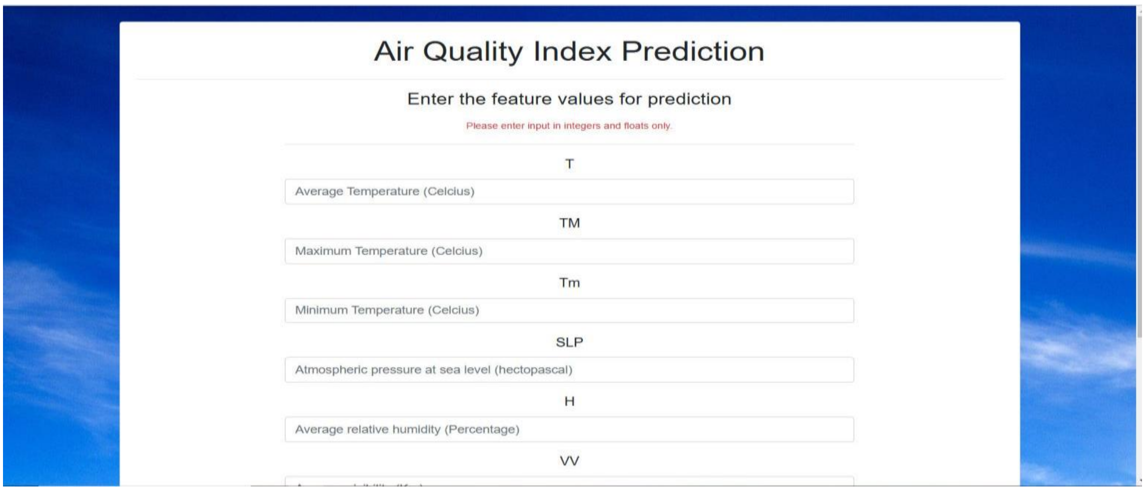

Fig 9. Output

5. Conclusion

The application of machine learning techniques to improve air quality prediction in specific cities. It compiled an extensive dataset incorporating data from air quality monitoring stations, weather stations, traffic patterns, and industrial activities in the targeted city. Through meticulous data preprocessing and feature engineering, we curated a dataset with meaningful and relevant features, including particulate matter (PM), nitrogen dioxide (NO2), sulfur dioxide (SO2), ozone (O3), carbon monoxide (CO), as well as meteorological variables and traffic data. Each model was subjected to training and validation using a subset of the dataset, and their performances were compared using evaluation metrics like mean squared error, mean absolute error, and coefficient of determination (R-squared). The successful utilization of XG Boost in this study underscores the immense potential of machine learning techniques in advancing air quality prediction. Through harnessing the power of data and sophisticated computational methods, we achieved the construction of a remarkably accurate and dependable model capable of delivering timely and precise air quality forecasts. The results presented in the discussion unveiled that XG Boost surpassed both SVM and Random Forest in accurately predicting air quality levels in the specific city. Its exceptional accuracy and improved generalization capabilities rendered it the most appropriate algorithm for air quality prediction in this particular context.

References

[1] Mrs. A. GnanaSoundariMtech, (Phd) ,Mrs. J. GnanaJeslin M.E, (Phd), Akshaya A.C. "Indian Air Quality Prediction And Analysis Using Machine Learning". International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 11, 2019 (Special Issue)

[2] Suhasini V. Kottur , Dr. S. S. Mantha. "An Integrated Model Using Artificial Neural Network(Ann) And Kriging For Forecasting Air Pollutants Using Meteorological Data". International Journal of Advanced Research in Computer and Communication Engineering ISSN (Online) : 2278-1021 ISSN (Print) : 2319-5940 Vol. 4, Issue 1, January 2015

[3] RuchiRaturi, Dr. J.R. Prasad ."Recognition Of Future Air Quality Index Using Artificial Neural Network".International Research Journal of Engineering and Technology (IRJET) .e-ISSN: 2395-0056 p-ISSN: 2395-0072 Volume: 05 Issue: 03 Mar-2018

[4] Aditya C R, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu ." Detection and Prediction of Air Pollution using Machine Learning Models". International Journal of Engineering Trends and Technology (IJETT) – volume 59 Issue 4 – May 2018

[5] Gaganjot Kaur Kang, Jerry ZeyuGao, Sen Chiao, Shengqiang Lu, and Gang Xie." Air Quality Prediction: Big Data and Machine Learning Approaches". International Journal of Environmental Science and Development, Vol. 9, No. 1, January 2018

[6] Alade IO, Rahman MAA, Saleh TA (2019a) Predicting the specific heat capacity of alumina/ethylene glycol nanofluids using support vector regression model optimized with Bayesian algorithm. Sol Energy 183:74–82.

[7] Alade IO, Rahman MAA, Saleh TA (2019b) Modeling and prediction of the specific heat capacity of Al2 O3/water nanofluids using hybrid genetic algorithm/support vector regression model. Nano-Struct Nano-Objects 17:103–111.

[8] Al-Jamimi HA, Saleh TA (2019) Transparent predictive modelling of catalytic hydrodesulfurization using an interval type-2 fuzzy logic. J Clean Prod 231:1079–1088.

[9] Al-Jamimi HA, Al-Azani S, Saleh TA (2018) Supervised machine learning techniques in the desulfurization of oil products for environmental protection: a review. Process Saf Environ Prot 120:57–71.

[10] Al-Jamimi HA, Bagudu A, Saleh TA (2019) An intelligent approach for the modeling and experimental optimization of molecular hydrodesulfurization over AlMoCoBi catalyst. J Mol Liq 278:376–384.

[11] Ayturan YA, Ayturan ZC, Altun HO, Kongoli C, Tuncez FD, Dursun S, Ozturk A (2020) Short-term prediction of PM2.5 pollution with deep learning methods. Global NEST J 22(1):126–131

[12] Bellinger C, Jabbar MSM, Zaïane O, Osornio-Vargas A (2017) A systematic review of data mining and machine learning for air pollution epidemiology. BMC Public Health.

[13] Bhalgat P, Bhoite S, Pitare S (2019) Air Quality Prediction using Machine Learning Algorithms. Int J Comput Appl Technol Res 8(9):367–370.

[14] K Murali Krishna, Amit Jain, Hardeep Singh Kang, Mithra Venkatesan, Anurag Shrivastava, Sitesh Kumar Singh, Muhammad Arif, Development of the Broadband Multilayer Absorption Materials with Genetic Algorithm up to 8 GHz Frequency, Security and Communication Networks

[15] P Gite, A Shrivastava, KM Krishna, GH Kusumadev, Under water motion tracking and monitoring using wireless sensor network and Machine learning, Materials Today: Proceedings, Volume 80, Part 3, 2023, Pages 3511-3516

[16] Anurag Shrivastava, Midhun Chakkaravathy, Mohd Asif Shah, A Novel Approach Using Learning

Algorithm for Parkinson’s Disease Detection with Handwritten Sketches’, Cybernetics and Systems, Taylor & Francis

[17] Mukesh Patidar, Anurag Shrivastava, Shahajan Mia h, Yogendra Kumar, Arun Kumar Sivaraman, An energy efficient high-speed quantum-dot based full adder design and parity gate for nano application, Materials Today: Proceedings, Volume 62, Part 7, 2022, Pages 4880-4890

[18] Castelli M, Clemente FM, Popoviˇc A, Silva S, Vanneschi L (2020) A machine learning approach to predict air quality in California. Complexity 2020(8049504):1–23.

[19] Doreswamy HKS, Yogesh KM, Gad I (2020) Forecasting Air pollution particulate matter (PM2.5) using machine learning regression models. Procedia Comput Sci 171:2057–2066.

[20] Fahad S, Sönmez O, Saud S, Wang D, Wu C, Adnan M, Turan, V (2021a) Plant growth regulators for climate-smart agriculture (1st ed.). CRC Press.

[21] Fahad, S, Sönmez O, Saud S, Wang D, Wu C, Adnan M, Turan V (2021b) Sustainable soil and land management and climate change (1st ed.). CRC Press.

[22] Liang Y, Maimury Y, Chen AH, Josue RCJ (2020) Machine learning-based prediction of air quality. Appl Sci 10(9151):1–17.

[23] Madan T, Sagar S, Virmani D (2020) Air quality prediction using machine learning algorithms–a review. In: 2nd international conference on advances in computing, communication control and networking (ICACCCN) pp 140–145.

[24] Anurag Shrivastava, S. J. Suji Prasad, Ajay Reddy Yeruva, P. Mani, Pooja Nagpal & Abhay Chaturvedi (2023): IoT Based RFID Attendance Monitoring System of Students using Arduino ESP8266 & Adafruit.io on Defined Area, Cybernetics and Systems, DOI: 10.1080/01969722.2023.2166243

[25] P. William, A. Shrivastava, H. Chauhan, P. Nagpal, V. K. T. N and P. Singh, "Framework for Intelligent Smart City Deployment via Artificial Intelligence Software Networking," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), 2022, pp. 455-460, doi: 10.1109/ICIEM54221.2022.9853119.

[26] Madhuri VM, Samyama GGH, Kamalapurkar S (2020) Air pollution prediction using machine learning supervised learning approach. Int J Sci Technol Res 9(4):118–123

[27] Velarde-Molina, J. F. ., Ancco, J. A. ., Rosado, M. B. ., Huamaní, E. L. ., Paico Campos, M. M. ., Meneses-Claudio, B. ., & Benancio, L. O. . (2023). Development of Mobile Application to Report Criminal Acts in Young Areas of Peru. International Journal on Recent and Innovation Trends in Computing and Communication, 11(4), 215–218. https://doi.org/10.17762/ijritcc.v11i4.6403

[28] Jacobs, M., Georgiev, I., Đorđević, S., Oliveira, F., & Sánchez, F. Efficient Clustering Algorithms for Big Data Analytics. Kuwait Journal of Machine Learning, 1(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/138

[29] Anupong, W., Azhagumurugan, R., Sahay, K. B., Dhabliya, D., Kumar, R., & Vijendra Babu, D. (2022). Towards a high precision in AMI-based smart meters and new technologies in the smart grid. Sustainable Computing: Informatics and Systems, 35 doi:10.1016/j.suscom.2022.100690