



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Machine Learning Optimization for Air Quality Forecasting

S.R.Sridhar, Rajula Prathap Reddy, Somapalle Prasad, Sriram Venkata Mahesh Raja

Assistant Professor, Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

ABSTRACT By using machine learning, we anticipate the air quality index for a certain place in India. The Indian air quality index is a commonly used indicator of pollutant (so₂, no₂, etc.) Levels across time. Based on historical data from previous years and forecasting over a certain forthcoming year as a Gradient Decent Boosted Multivariable Regression Problem, we constructed a model to predict the air quality index. For our predictive problem, we use cost estimation to increase the model's efficacy. When given historical data on pollutant concentration, our model will be able to accurately estimate the air quality index for an entire county, any state, or any contiguous region. By including the suggested parameter-reducing formulations into our model, we outperformed the traditional regression models in terms of performance. Our model predicts the air quality index for the entirety of India with 96% accuracy, and we also use the XG Boost Algorithm technique to determine the order of preference based on how closely it approaches the optimal answer. By utilising machine learning to estimate the air quality index of a certain place, we forecast India's air quality. The Indian air quality index is a commonly used indicator of pollutant (so₂, no₂, etc.) Levels across time. Based on historical data from previous years and forecasting over a certain forthcoming year as a Gradient Decent Boosted Multivariable Regression Problem, we constructed a model to predict the air quality index. For our predictive problem, we use cost estimation to increase the model's efficacy. When given historical data on pollutant concentration, our model will be able to accurately estimate the air quality index for an entire county, any state, or any bounded region. By including the suggested parameter-reducing formulations into our model, we outperformed the traditional regression models in terms of performance. Using our model, we apply the XG Boost Algorithm technique to discover the order of preference by similarity to the ideal answer, and we predict the air quality index of the entire country of India with 96% accuracy using the currently available dataset.

KEYWORDS: Accuracy, Air pollution, Detection, Machine learning, Prediction, Recommendation

I.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.

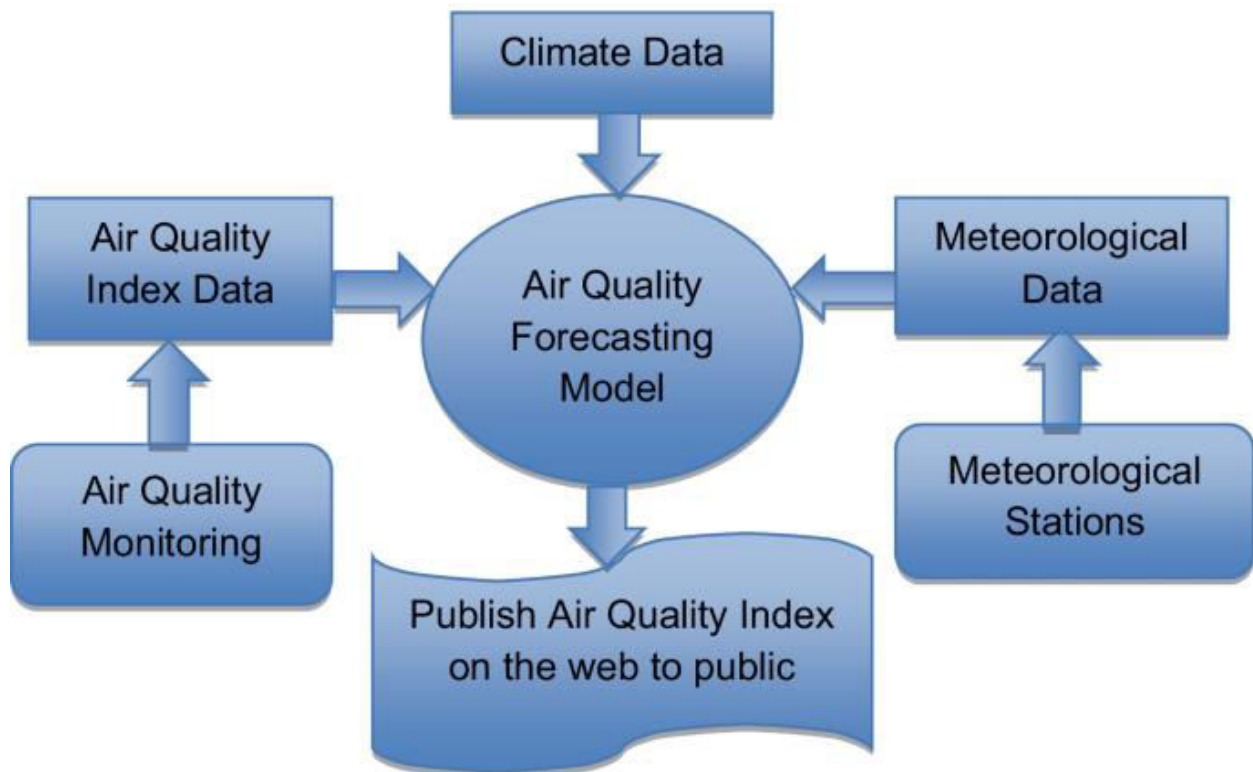


Fig 1: Analysis of deep learning approaches for air pollution

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 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.

II. LITERATURE REVIEW

Li Hong Xingfeng, 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 A, 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, , 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].

III.METHODS

The Environment is nothing but everything that encircles us. The environment is getting polluted due to human activities and natural disaster, very severe among them is air pollution. The concentration of air pollutants in ambient air is governed by the meteorological parameters such as atmospheric wind speed, wind direction, relative humidity, and temperature. If the humidity is more, we feel much hotter because sweat will not evaporate into the atmosphere. Urbanization is one of the main reasons for air pollution because, increase in the transportation facilities emits more pollutants into the atmosphere and another main reason for air pollution is Industrialization. The major pollutants are Nitrogen Dioxide (NO₂), Carbon Monoxide (CO), Particulate matter (PM), Sulphur Dioxide (SO₂), Carbon Dioxide (CO₂) etc. Carbon Monoxide is produced due to the deficient Oxidization of propellant such as petroleum, gas, etc. Nitrogen Oxide is produced due to the ignition of thermal fuel; Carbon monoxide causes headaches, vomiting; Benzene is produced due to smoking, it causes respiratory problems; Nitrogen oxides causes dizziness, nausea; Particulate matter with a diameter 2.5 micrometre or less than that affects more to human health. Measures must be taken to minimize air pollution in the environment.

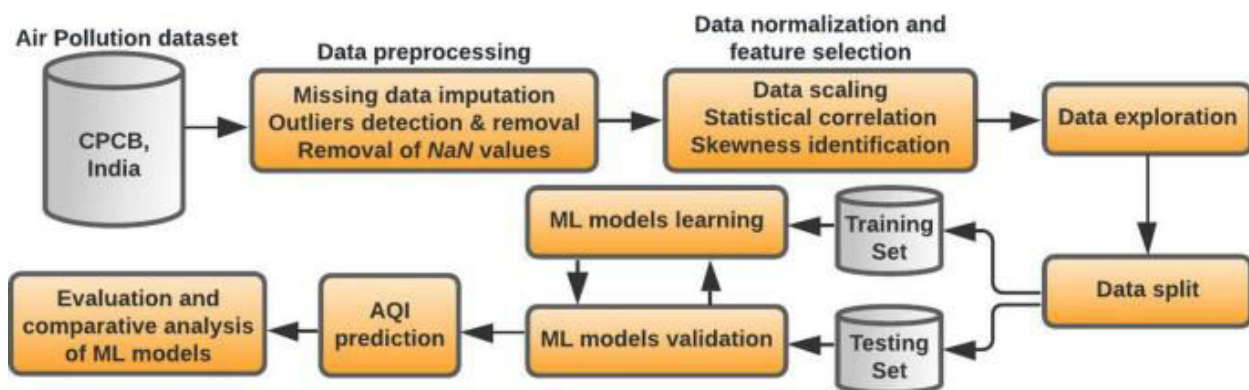


Fig 2: Work Flow

Air Quality Index (AQI), is used to measure the quality of air. Earlier classical methods such as probability, statistics were used to predict the quality of air, but those methods are very complex to predict the quality of air. Due to advancement of technology, now it is very easy to fetch the data about the pollutants of air using sensors. Assessment of raw data to detect the 3 pollutants needs vigorous analysis. Convolution Neural networks, Recursive Neural networks, Deep Learning, Machine learning algorithms assures in accomplishing the prediction of future AI so that measures can be taken appropriately. Machine learning which comes under artificial intelligence has three kinds of learning algorithms, they are the Supervised Learning, Unsupervised learning, Reinforcement learning. In the proposed work we have used supervised learning approach. There are many algorithms under supervised learning algorithms such as Linear Regression, Nearest Neighbour, XG Boost Algorithm. Compared to all other algorithms Random Forest gives better results, so our approach selects Random Forest to predict the accurate air pollution.

IV.RESULT ANALYSIS

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. 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.

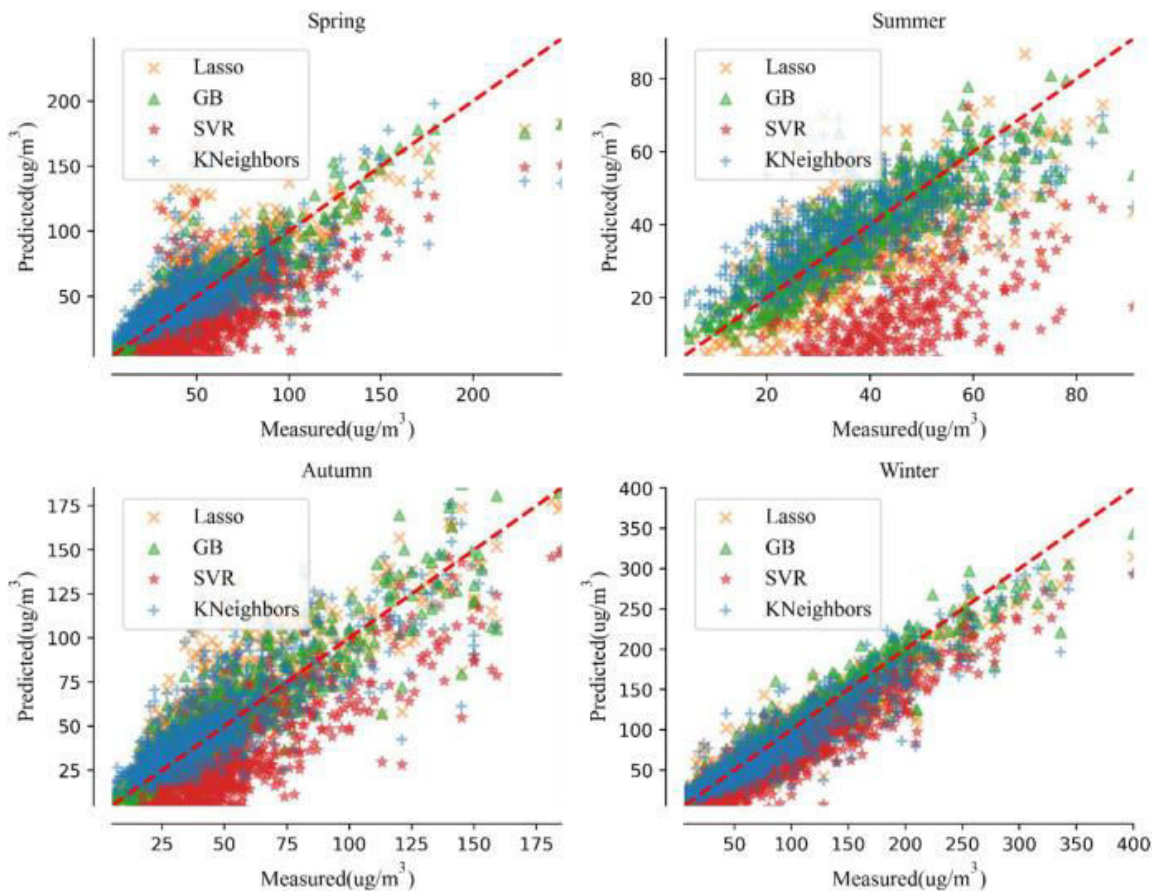


Fig 3: Air prediction Result analysis

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 . Advancing Air Quality Prediction Accuracy Analysis Advancing Air Quality Prediction Precision Analysis.

V.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] P. Rafaj, G. Kiesewetter, T. Gül, W. Schöpp, J. Cofala, Z. Klimont, and P. Purohit, "Outlook for clean air in the context of sustainable development goals," *Global Environ. Change*, vol. 53, pp. 111, Nov. 2018.
- [2] P. J. Landrigan, R. Fuller, N. J. R. Acosta, O. Adeyi, R. Arnold, A. B. Baldé, and R. Bertollini, "The Lancet commission on pollution and health," *Lancet*, vol. 391, no. 10119, pp. 462512, 2018.
- [3] R. A. Rohde and R. A. Müller, "Air pollution in China: Mapping of concentrations and sources," *PLoS ONE*, vol. 10, no. 8, Aug. 2015, Art. no. e0135749.
- [4] State of Global Air 2019, Health Effects Institute, Boston, MA, USA, 2019.
- [5] K. Al Hwaitat, S. Manaseer, R. M. Al-Sayed, M. A. Almaiah, and O. Almomani, "An investigator digital forensics frequencies particle swarm optimization for detection and classification of APT attack in FOG computing environment (IDF-FPSO)," *J. Theor. Appl. Inf. Technol.*, vol. 98, no. 7, pp. 937_952, 2020.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details