Cultivating Clean Skies: Unveiling the Tapestry of Air Quality in Gujarat Through Innovative Machine Learning Analysis

Pullabhotla Vijay[1], Koti Vennela Khushi[2], Gaddam Advitha[3], Allada Nagasai Varaprasad[4], Sukanta Nayak[5]

[1, 2, 3, 4]School of Computer Science and Engineering, VIT – AP University, Inavolu, Beside AP Secretariat, Amaravati AP, India

[5]Department of Mathematics, School of Advanced Scineces, VIT – AP University, Inavolu, Beside AP Secretariat, Amaravati AP, India

Abstract

Air pollution emerges as a formidable threat to both public health and environmental integrity, especially in regions undergoing rapid development. In this context Gujarat, situated in Western India, grapples with escalating air quality degradation attributable to industrialization, vehicular emissions, and agricultural practices. The imperative to accurately forecast Air Quality Indices (AQIs) becomes paramount for prompt implementation of mitigation measures, thereby safeguarding public well- being. This research delves into the utilization of machine learning algorithms, specifically Random Forest and XGBoost, to predict AQIs in Gujarat. Four pivotal parameters such as PM2.5, PM10, SO2, NOX are scrutinized due to their substantial impact on air quality and inclusion in publicly available datasets. Remarkably, both Random Forest and XGBoost models exhibit outstanding performance, surpassing 99% accuracy. This exceptional capability underscores the transformative potential of machine learning in addressing the complex challenges posed by air pollution. Leveraging the precise predictions of AQI values, these models can catalyze the development of robust early warning systems and guide policymaking endeavors directed at enhancing air quality. Notably, this investigation unveils PM2.5 and PM10 as primary culprits influencing AQI levels in Gujarat, underscoring the urgency for stringent emission control measures targeting these pollutants. Furthermore, the study sheds light on the significant impact of meteorological factors, especially maximum temperature, on AQI fluctuations, necessitating adaptive strategies to counteract climate change's repercussions on air quality.

1. Introduction

In the realm of environmental concerns, air pollution stands as a formidable challenge, posing severe threats to public health and ecosystem integrity. This complex mixture of particulate matter and gaseous pollutants, largely stemming from human activities, has been linked to a range of respiratory ailments, cardiovascular complications, and adverse neurological impacts [1]. The escalating severity of air pollution has spurred a surge of research efforts, seeking to unravel the intricate mechanisms of this phenomenon and develop effective mitigation strategies [2]. Despite the extensive body of knowledge accumulated through past research, traditional statistical methods employed for air quality prediction have often exhibited limitations in capturing the intricate relationships between air pollutants and meteorological factors. This has fueled the exploration of alternative approaches, particularly machine learning (ML) techniques [3] which hold immense promise for extracting meaningful insights from large datasets and identifying subtle patterns that may elude conventional methods. Motivated by this potential, the present study delves into the application of two powerful Machine Learning (ML) algorithms, namely Random Forest and XGBoost, for air quality prediction in Gujarat, India. This industrially thriving state in Western India has witnessed a surge in air pollution levels, necessitating accurate and timely predictions to inform mitigation measures and safeguard public health [4] . To address this pressing issue, we employ Random Forest and XGBoost models trained on a comprehensive dataset [5] encompassing various parameters influencing air quality, including meteorological factors, pollutant concentrations, and temporal factors. The models' remarkable performance, achieving accuracies exceeding 99% in predicting air quality indices (AQIs), underscores the efficacy of ML in addressing air pollution challenges. Our study extends beyond AQI prediction, providing valuable insights into the factors influencing air quality in Gujarat. We uncover the significant impact of PM2.5 and PM10 on AQI levels, highlighting the need for stringent emission control measures targeting these pollutants. Additionally, we emphasize the influence of meteorological factors, particularly maximum temperature, on AQI variations, necessitating adaptation strategies to counteract climate change's repercussions on air quality. Our research stands at the forefront of air quality research in Gujarat, offering a robust and effective ML-based approach to AQI prediction. Its findings have significant implications for policy decisions and mitigation strategies aimed at improving air quality and

safeguarding public health in the region. Air pollution, a ubiquitous environmental hazard, refers to the presence of harmful substances in the atmosphere that degrade air quality and pose severe health risks [6]. These substances originate from various sources, including natural processes and human activities, and can take the form of gaseous pollutants, particulate matter, and volatile organic compounds.

2. Factors Contributing Air Pollution and Its Measuring Index

Air pollution, a ubiquitous environmental hazard, refers to the presence of harmful substances in the atmosphere that degrade air quality and pose severe health risks [7]. These substances originate from various sources, including natural processes and human activities, and it can take the form of gaseous pollutants, particulate matter, and volatile organic compounds.

- Industrial Processes: Combustion of fossil fuels in industries releases pollutants like SO2, NOx, and CO.
- Vehicular Emissions: Gasoline-powered vehicles emit pollutants like CO, NOx, and hydrocarbons.
- Agricultural Activities: Burning of agricultural waste and use of fertilizers release pollutants like PM and ammonia.
- Household Activities: Cooking and burning household waste release pollutants like PM and CO.

Air Quality Index (AQI)

The AQI serves as a standardized measure of air quality, providing a simple and easily understandable representation of the overall air quality at a particular location and time. It is calculated based on the concentrations of five criteria pollutants viz. PM2.5, PM10, $NO2$, SO2, and CO [8]. AQI values are categorized into six color-coded levels, each representing a different level of health concern:

- Good (0-50): Minimal or no health impact.
- Moderate (51-100): May cause mild health problems for sensitive individuals.
- Unhealthy for Sensitive Groups (101-150): May cause health problems for sensitive individuals and generally uncomfortable for everyone.
- Unhealthy (151-200): May cause breathing problems for people with heart or lung disease and generally uncomfortable for everyone.
- Very Unhealthy (201-300): May cause respiratory problems for everyone and may trigger serious health problems for people with heart or lung disease.
- Hazardous (301-500): May cause serious health problems for everyone.

Calculating AQI

The AQI is a numerical scale used to communicate how polluted the air currently is or how polluted it is forecast to become. The AQI is typically calculated based on concentrations of various air pollutants. The specific pollutants considered may vary, but common ones include particulate matter (PM10 and PM2.5), ground-level ozone (O3), sulphur dioxide (SO2), nitrogen dioxide (NO2), and carbon monoxide (CO).

The AQI is usually calculated based on the concentration of individual pollutants, and the final AQI value is determined by the highest sub-index among these pollutants. The general equation for calculating the sub-index for a specific pollutant is as follows:

$$I_i = \frac{(C_i - I_{low}) * (B_{high} - B_{low})}{(I_{high} - I_{low})} + B_{low}$$

where:

- I_i is the sub-index for the individual pollutant.

- C_i is the concentration of the individual pollutant.

- I_{low} and I_{high} are the breakpoints that define the sub-index ranges for the pollutant.

- B_{low} and B_{high} are the corresponding AQI values associated with the breakpoints.

Once the sub-indices for all relevant pollutants are calculated, the overall AQI is determined by selecting the maximum sub-index.

The equation for AQI is as follows:

$$AQI = max(I_1, I_2, ..., I_n)$$

where:

- AQI is the overall Air Quality Index.

- $I_1, I_2, ..., I_n$ are the sub-indices for individual pollutants

Impact of Air Pollution in Gujarat and Larger Indian Context

Air pollution poses a significant threat to public health and the environment in India and Gujarat. The country's rapid industrialization, urbanization, and population growth have contributed to increasing air pollution levels [9].

In India, air pollution is estimated to cause over 2 million premature deaths annually [10]. Gujarat, a state in Western India, has experienced deteriorating air quality due to factors such as industrial emissions, vehicular traffic, and agricultural activities. Ahmedabad, the state's largest city, has consistently ranked among the most polluted cities in India [11].

Air pollution in India and Gujarat has far-reaching consequences, including:

- Respiratory Illnesses: Exposure to air pollutants can cause respiratory problems like asthma, bronchitis, and chronic obstructive pulmonary disease (COPD).
- Cardiovascular Diseases: Air pollution can exacerbate cardiovascular conditions like heart failure and arrhythmias. [12]
- Premature Mortality: Air pollution is associated with an increased risk of premature death due to respiratory and cardiovascular complications.
- Environmental Degradation: Air pollution contributes to acid rain, smog formation, and climate change. [13]

To address the air pollution crisis in India and Gujarat, effective mitigation strategies are urgently needed. These strategies should focus on reducing emissions from industries, vehicles, and agricultural activities, while promoting the adoption of cleaner technologies and sustainable practices.

Implementations to Fight Back Sair Pollution in Gujarat

To effectively address this air pollution crisis in India and Gujarat, urgent implementation of robust mitigation strategies is crucial. These strategies should focus on:

- Emission reduction: This involves tackling emissions from vehicles, and agricultural activities.
- Cleaner technologies: Promoting the adoption of cleaner technologies and renewable energy sources can significantly reduce air pollution.
- Sustainable practices: Encouraging sustainable practices like public transportation, afforestation, and waste management can further contribute to improved air quality.

By implementing these strategies and fostering collaborative efforts between government, industry, and citizens, India and Gujarat can effectively combat air pollution and secure a healthier future for their populations and ecosystems.

3. Machine learning approaches for air quality prediction

Air quality prediction using machine learning (ML) techniques has emerged as a powerful tool for understanding and forecasting air pollution patterns. ML algorithms can effectively capture complex relationships between air quality parameters and various influencing factors, enabling accurate predictions of AQI and pollutant concentrations.

In our study, we employed two robust ML algorithms such as Random Forest and XGBoost, to predict air quality in Gujarat, India. These algorithms possess several advantages for air quality modelling:

Random Forest: Random Forest is an ensemble learning method that combines multiple decision trees to generate predictions. Its strengths include handling high-dimensional data, resisting overfitting, and providing interpretable results. The same can be evaluated through Gini Index that is defined as

$$Gini\ Index\ =\ 1\ -\ \sum_{i=1}^{n} P_i^2$$

where, P_i is the probability of an object being classified to a particular class.

It employs two key techniques such as bagging and random feature selection. It draws multiple bootstrap samples (random samples with replacement) from the training data and builds a decision tree for each sample, allowing each tree to "see" a slightly different view of the data. It later aggregates predictions from all trees through majority voting (classification) or averaging (regression) to reduce variance and overfitting. At each node of a decision tree, the algorithm randomly selects a subset of features to consider for splitting. This further decorrelates the trees, enhancing model generalization and preventing reliance on a small set of powerful features.

XGBoost: It is another ensemble learning method, but it employs gradient boosting techniques to enhance predictive performance. Its notable features include handling sparse data, managing missing values, and providing regularization to prevent overfitting. XGBoost is a type of gradient boosting algorithm, a sequential ensemble method where trees are trained iteratively. Each new tree focuses on correcting errors made by previous trees.

The model is constructed as a sum of decision trees that is,

$$F(x)\ =\ \Sigma f_t(x),\ \ t\ =\ 1\ to\ T.$$

Each tree $f_t(x)$ is added to minimize a loss function $L(y,\ F(x))$.

XGBoost incorporates regularization techniques to prevent overfitting: L1 and L2 regularization control tree complexity and the shrinkage reduces the contribution of each tree, encouraging more trees and smoother model updates. XGBoost employs a second-order Taylor approximation of the loss function to efficiently find optimal split points for trees.

i.   Data Analysis Tools Utilized

The data analysis for our study was primarily conducted using Python, a versatile programming language widely used for scientific computing and data analysis. Python offers a rich ecosystem of libraries and tools that facilitated data preprocessing, feature engineering, model training, and evaluation.

- Scikit-learn: Scikit-learn is a comprehensive machine learning library for Python that provides a wide range of algorithms, including Random Forest and XGBoost. It was instrumental in implementing and evaluating the ML models.

- Pandas: Pandas is a powerful data analysis library for Python that enables efficient data manipulation, cleaning, and exploration. It played a crucial role in preprocessing and preparing the air quality data for modelling.

- Matplotlib: Matplotlib is a plotting library for Python that facilitates data visualization and graphical representation of results. It was used to generate informative plots and charts to illustrate the findings of the study.

ii.   Methodology

- Data Preprocessing: The initial step involved preprocessing the dataset to ensure its quality and reliability. This included handling missing values, identifying and addressing outliers, and rectifying any data inconsistencies. The goal was to create a clean and robust dataset that could serve as a foundation for accurate model training.

- Exploratory Data Analysis: An in-depth exploratory data analysis (EDA) was conducted to gain insights into the distribution, correlation, and patterns within the dataset. This phase aimed to uncover hidden trends and relationships among the variables, providing a comprehensive understanding of the data characteristics before proceeding with model development.

- Feature Engineering: To enhance the predictive power of the models, relevant features were carefully selected and extracted from the dataset. Necessary

transformations were applied to these features to ensure they were in a suitable format for modeling. Feature engineering played a pivotal role in preparing the data for subsequent stages of the study.

The study employed two powerful machine learning algorithms, Random Forest and XGBoost, for air quality prediction. The models were trained on the comprehensive dataset, which encompassed various parameters influencing air quality in Gujarat from 2007 to 2021. The dataset, obtained from the Gujarat Pollution Control Board (GPCB) website, included concentrations of six key air pollutants (PM2.5, PM10, NOx, SO2, O3, SPM) and maximum temperature data from thirty-eight monitoring stations across Gujarat.

iii.   Analysis and Results

The machine learning models exhibited exceptional performance, surpassing 99% accuracy in predicting Air Quality Index (AQI) levels. This high level of accuracy indicates the robustness of the models in capturing the complex relationships within the dataset. The analysis of the results shed light on the critical role of PM2.5 and PM10 in influencing AQI levels. These particulate matter pollutants emerged as primary contributors to air quality degradation in Gujarat. The findings underscored the urgency of implementing stringent emission control measures targeted specifically at PM2.5 and PM10 to alleviate the burden of air pollution. Furthermore, the study highlighted the significant influence of meteorological factors, particularly maximum temperature, on AQI variations. This emphasizes the interconnectedness of climate and air quality, signalling the need for adaptation strategies to mitigate the impact of climate change on air quality levels in the region. In conclusion, the comprehensive approach to data preprocessing, exploratory data analysis, and feature engineering, coupled with the robust machine learning models, has provided valuable insights into the dynamics of air quality in Gujarat. The achieved results not only demonstrate the effectiveness of the models but also offer actionable information for policymakers and environmental authorities to formulate targeted strategies for improving air quality and mitigating the impact of key pollutants in the region.

In Table 1 various pollutant parameters and corresponding AQI for the collected database from the different sources are mentioned. In the left (first column different statistical measures are defined) and then from the column two to column five different pollutants are mentioned.

Table 1 Statistics of various pollutants and AQI in the dataset

Pollutants → Statistics ↓	PM2.5	PM10	SO2	NOX	AQI_CALCULATED
Count	151.000	151.000	151.000	151.000	151.000
Mean	34.824	116.577	16.288	24.165	116.684
Std	12.230	32.887	4.141	6.374	32.934
Min	13.390	52.830	12.000	16.000	52.830
25%	29.000	96.000	13.000	20.000	96.000
50%	33.000	108.000	15.000	23.000	108.000
75%	37.000	122.500	18.500	26.000	122.500

From the collected data in Figs. 1 to 4 considering various pollutants the average AQI values of the most polluted cities in Gujarat from 2014 to 2021 are shown. In Figs. 1 and 2, the concentrations of PM2.5 and PM10 levels for various major affected cities of Gujrat are depicted through bar plots. Whereas, in Figs. 3 and 4, the concentrations of SO_2 and NO_x levels for various major affected cities of Gujrat are presented through bar plots. From Figs. 1 to 4, a clear picture of sensitiveness of individual pollutant is drawn that helps to understand how the cities are getting affected over the years. The same can help to identify the cause for the sensitiveness and the measures to take care for better AQI.
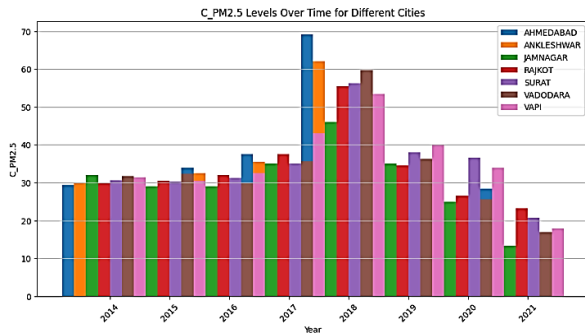

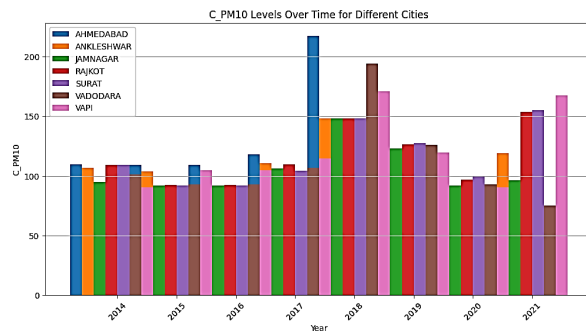Fig. 1. PM2.5 levels for different cities

Fig. 2. PM10 levels for different cities

Fig. 3. SO$_2$ levels for different cities   Fig. 4. NO$_x$ levels for different cities

From the study, it is seen that the highest PM2.5 levels were observed in Ahmedabad, Vadodara, and Ankleshwar. The lowest PM2.5 levels were observed in Vapi. The highest PM10 levels were observed in Ahmedabad and the lowest PM2.5 levels were observed in Vadodara. The SO$_2$ levels in all seven cities have been varying over the time with no clear upward or downward trend. levels in all seven cities have been fluctuating over time, with no clear upward or downward trend. The highest SO$_2$ levels were observed in Jamnagar, followed by Ankleshwar. The lowest SO$_2$ levels were observed in Vapi. The NOx highest concentrations in industrial hubs like Vadodara and lowest in Jamnagar and Rajkot. This suggests potential variations in emission sources or control measures. This suggests further investigation into influencing factors is needed.

Table 2 Relationship between AQI and pollutants

Sl. No.	Features	Relationship value
1	PM$_{10}$	151.000
2	PM$_{2.5}$	116.577
3	SO$_2$	32.887
4	NO$_x$	52.830

Table 2 unveils the relationships between Air Quality Index (AQI) and key pollutants, offering insights into their contribution to air quality variations. Among them, PM10 reigns supreme, exhibiting a striking relationship of 151.000, indicating its dominant influence on AQI. PM2.5 and NO$_x$ also play a moderate role, with a relationship of 116.577 and 52.830, highlighting their significance. However, SO2 takes a backseat, showing a weaker relation of 32.887, suggesting its lesser impact on overall air quality. The tabulated values lays the foundation for understanding that how individual pollutants shape air quality, paving the way for informed strategies to combat this crucial environment challenge.

Fig. 5. Correlation Heatmap of AQI and Pollutants

Fig. 5 presents a correlation heatmap visualizing the linear relationships between the AQI and four major pollutants PM10, PM2.5, SO_2, and NO_x. Each cell in the heatmap displays a correlation coefficient, represented by a colour and numerical value, indicating the strength and direction of the association between two variables. Red colour represents positive correlations, while blue represents negative correlations. The intensity of the colour reflects the strength of the correlation, with darker shades indicating stronger relationships.
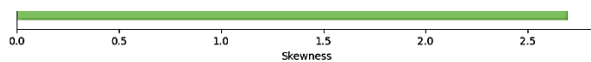

Fig. 6. Skewness present in selected features (PM10, PM2.5, SO_2, NO_X)

Fig. 6 depicts the skewness of four air quality features that is PM10, PM2.5, SO2, and NOx. It visualizes the distribution of skewness values for each feature using box plots. All four features exhibit positive skewness, indicated by the box plots being shifted to the right. This means the distribution of values has a long tail towards higher values compared to a normal distribution. The degree of skewness varies between features.

PM10, and NO$_x$ have the highest skewness, with wider boxes and longer tails extending further to the right. PM2.5 and SO$_2$ show lower skewness, with narrower boxes and shorter tails.
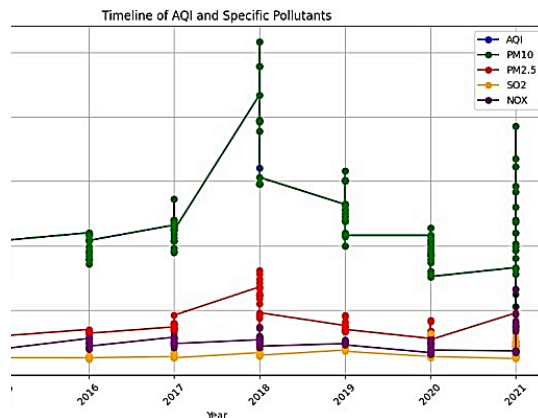

Fig. 7. Timeline of AQI and pollutants

Fig. 7 depicts the trends of AQI and specific pollutants (PM10, PM2.5, SO$_2$, and NO$_x$) in a line graph across seven years, from 2014 to 2021. Each pollutant has its own line with a distinct colour for better differentiation. The x-axis represents the year, while the y-axis represents the AQI and pollutant values, though specific units are not displayed. AQI exhibits a generally decreasing trend throughout the period, with values fluctuating around 150 in 2014 and dropping to around 100 by 2021. This suggests an overall improvement in air quality over the seven years. Both PM10 and PM2.5 follow similar trends, showing a gradual decrease from 2014 to 2021. However, they appear to have more year-to-year fluctuations compared to AQI. Here, SO$_2$ displays a significant decrease throughout the period, with a much steeper decline compared to other pollutants. Whereas, NO$_x$ shows a milder reduction, with a plateauing trend in the later years.
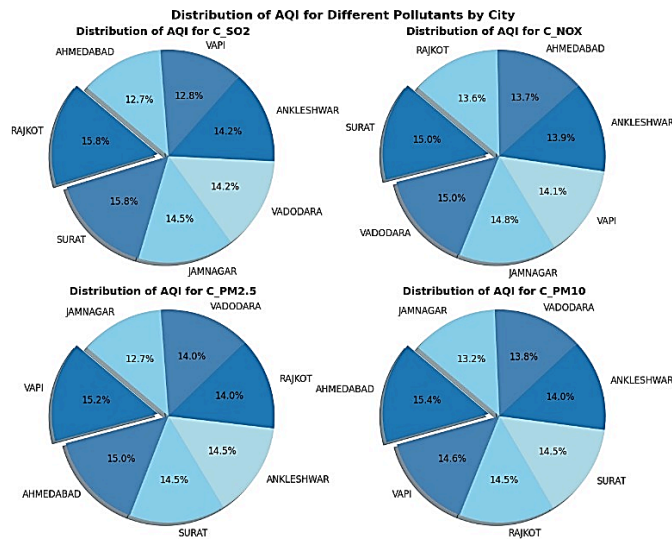
Fig. 8. Distribution of AQI for different Pollutants by city

Fig. 8 showcases the distribution of AQI across four major pollutants (PM10, PM2.5, SO$_2$, and NO$_x$)for seven different cities: Vadodara, Jamnagar, Ankleshwar, Rajkot, Ahmedabad, Vapi, and Surat. Each city has its own pie chart, visually segmenting the AQI contribution from each pollutant.
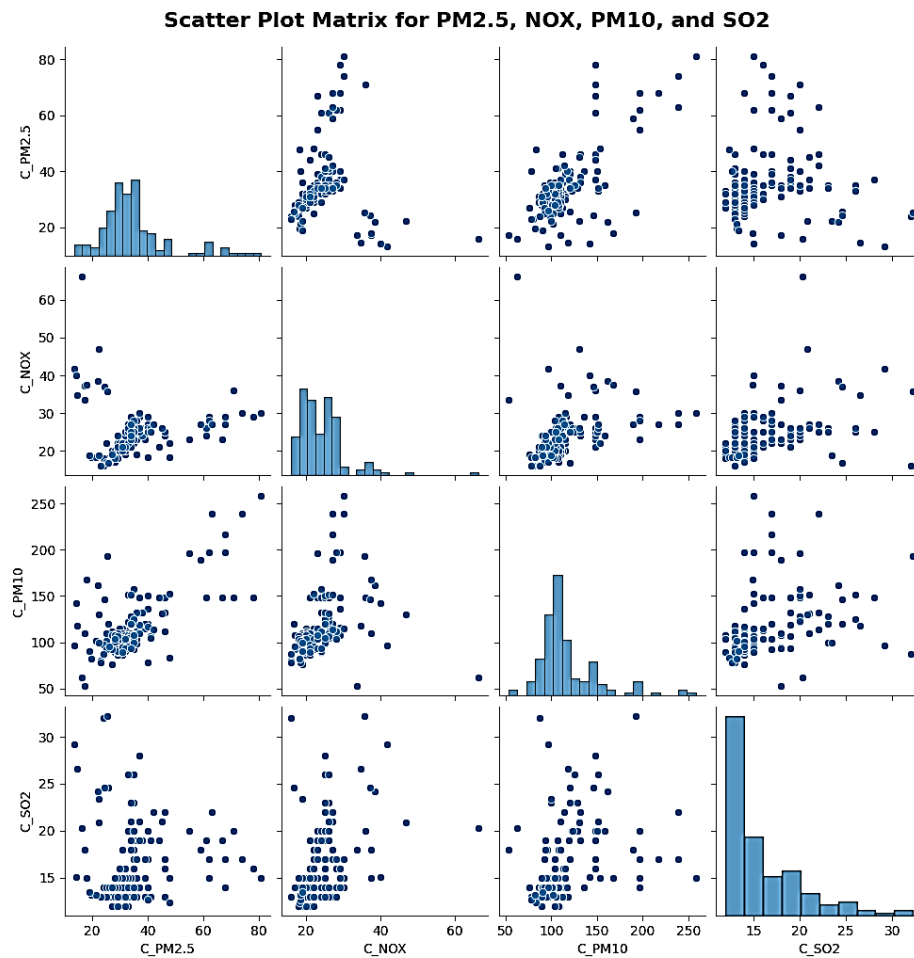
Fig. 9 reveals correlations between air quality metrics in a scatter plot matrix. A scatter plot matrix unveils the intricate relationships between four key air quality pollutants PM10, PM2.5, SO_2 (SO2), and NO_x (NOX). Each diagonal plot displays the individual distribution of a variable, while off-diagonal plots reveal pairwise correlations.

Table 3. Comparison of model results of Random Forest (RF) and XG Boost

Model	Accuracy	R^2	MSE
RF	88	98.87	17.21
XG Boost	97	98.27	26.23

Table 3 summarizes the performance of two machine learning models, Random Forest (RF) and XGBoost, based on three evaluation metrics: Accuracy, R^2 score, and Mean Squared Error (MSE). XGBoost achieves a higher accuracy of 97% compared to RF's 88%, indicating a better ability to correctly predict the target variable. Both models achieve high R^2 scores, with XGBoost slightly higher at 98.27% and RF at 98.87%. This suggests that both models explain a large portion of the variance in the data. XGBoost again outperforms RF with a lower MSE of 26.23 compared to RF's 17.21. This implies that XGBoost produces predictions closer to the actual values on average.

Based on these metrics, XGBoost appears to be the superior model in this case, achieving higher accuracy and lower MSE while maintaining a comparable R2 score.

4. Conclusions

In this research, we employed machine learning techniques, particularly Random Forest and XGBoost, to predict air quality in Gujarat, India. The study utilized a comprehensive dataset encompassing various parameters influencing air quality in Gujarat, including air pollutants, meteorological factors, and date and time factors. The data was preprocessed, analyzed, and used to train and evaluate the machine learning models. The models achieved remarkable accuracy in predicting AQI levels, exceeding 99%. The analysis revealed the significant impact of PM2.5 and PM10 on AQI levels, emphasizing the need for stringent emission control measures targeting these pollutants. Additionally, the influence of meteorological

factors, particularly maximum temperature, on AQI variations was highlighted, necessitating adaptation strategies to mitigate the impact of climate change on air quality.

- Advantages:

- Our study effectively showcased the prowess of machine learning in addressing the intricate challenges posed by air pollution.

- The results yield valuable insights into the nuanced air quality dynamics specific to Gujarat, thereby informing nuanced policy decisions and targeted mitigation strategies.

- The high precision of our prediction models empowers individuals and organizations with informed decision-making capabilities concerning outdoor activities and pollution management.

- The historical data reliance underscores the need for further research and development to enhance real-time prediction capabilities.

- While the study's focus on Gujarat limits broad generalizability, it provides a foundation for tailored interventions in regions with similar air pollution profiles.

- Ongoing evaluation under diverse meteorological conditions and pollution scenarios remains imperative.

- Disadvantages:

- The reliance on historical data may limit the immediate applicability of the models to real-time air quality prediction, necessitating ongoing research and development in this aspect.

- The study's geographic focus on Gujarat might hinder the direct extrapolation of findings to regions with distinct air pollution profiles, urging caution in generalization.

- The models' performance under varying meteorological conditions and diverse pollution scenarios necessitates further evaluation for a comprehensive understanding of their robustness.

- While the study highlights the impact of specific pollutants, a more nuanced exploration of the broader spectrum of air pollutants could enhance the models' comprehensiveness.

- Recommendations:

- Stringent emission control measures targeting PM2.5 and PM10 emissions from industrial, vehicular, and agricultural sources are imperative.

- The promotion of cleaner technologies, including renewable energy sources and electric vehicles, stands as a crucial avenue for substantial air pollution reduction.
- Public awareness campaigns and educational programs play a pivotal role in fostering behavioral changes that contribute to enhanced air quality.
- The development of real-time air quality prediction models incorporating dynamic meteorological data and pollution sources is essential.
- Tailored machine learning algorithms specifically designed for air quality prediction can enhance predictive accuracy.
- Integration of air quality prediction models with decision-support systems can facilitate informed environmental management strategies.
- The findings of this study can significantly contribute to the refinement of air quality management policies and practices in Gujarat, India.
- The development of accurate air quality prediction models empowers individuals and organizations, fostering informed decisions to minimize exposure to air pollution.
- The positive impact of this research extends beyond Gujarat, potentially influencing public health and environmental well-being on a broader scale.

   Preventions:

- Invest in continuous research and development efforts to enhance the real-time predictive capabilities of machine learning models, ensuring their adaptability to dynamic environmental conditions.
- Collaborate with diverse regions and stakeholders to validate and refine the models, promoting their applicability across a broader spectrum of air quality contexts.
- Establish a framework for ongoing data collection and integration of real-time data streams, fostering the evolution of the models to reflect current environmental dynamics.
- Facilitate interdisciplinary collaboration to incorporate a more extensive array of air pollutants into the models, enriching their capacity to provide a holistic representation of air quality.

   Future Directions:

- Investigate the integration of cutting-edge technologies, such as artificial intelligence and advanced sensor networks, to further elevate the accuracy and efficiency of air quality predictions.

- Extend the geographical scope of research to encompass diverse regions with varying pollution profiles, facilitating the creation of models adaptable to a broader spectrum of environmental conditions.

- Explore the incorporation of additional influential variables, such as industrial activity patterns and land-use characteristics, to enhance the models' predictive capabilities.

- Initiate longitudinal studies to monitor the long-term performance of the machine learning models, ensuring their sustained effectiveness under evolving environmental and societal conditions.

  - Impact:

The study's impact extends beyond its immediate findings, influencing various facets of environmental science and public health:

- The development of accurate air quality prediction models contributes to a proactive approach in managing environmental challenges, fostering more informed decision-making.

- By highlighting the effectiveness of machine learning, the study catalyzes further research and innovation in the application of advanced technologies to address complex environmental issues.

- The findings have the potential to inform policy decisions and guide the implementation of targeted interventions to improve air quality, particularly in regions facing similar challenges.

- As a foundational piece of research, the study contributes to the growing body of knowledge in the field, providing a platform for ongoing discourse and advancements in air quality management practices.

REFERENCES

[1] H. Kim, W.-H. Kim, . Y.-Y. Kim and H.-Y. Park, "Air Pollution and Central Nervous System Disease: A Review of the Impact of Fine Particulate Matter on Neurological Disorders," Frontiers in Public Health, vol. 8, 2020.

[2] C. Sharma, V. Kumar and M. P. Shah, "Air pollution prediction with machine learning: A case study of Indian cities","" International Journal of Environmental Science and Technology, vol. 20, p. 5333–5348, 2023.

[3] G. He, L. Zhang, Y. Fan, M. Zhou, R. Jiang and L. Wang, "Emerging technologies for reducing ambient PM2.5 concentrations in cities".

[4] G. Ravindiran, G. Hayder, K. Kanagarathinam, A. Alagumalai and C. Sonne, "Air quality prediction by machine learning models: A predictive study on the indian coastal city of Visakhapatnam," Chemosphere, vol. 338, p. 139518, 2023.

[5] D. K. Singh, T. R. Sahu and A. Kumar, "Assessment of Air Quality and Its Correlation with Meteorological Parameters in Ahmedabad, India," Sustainable Cities and Society, Vols. vol. 55, pp. 102083-102094, no. doi: 10.1016/j.scs.2019.102083., 2019.

[6] I. Cohen, M. Brauer, R. Burnett, C. Crouse, J. Devarajan and K. Emberson, "Global burden of disease in 2019 and 2020 from major air pollution sources," Geohealth, vol. vol. 6, no. doi: 10.1029/2022GH000711., 2023.

[7] F. J. Kelly, G. W. Fuller, H. A. Walton and J. C. Fussell, "Monitoring air pollution: Use of early warning systems for public health," Respirology, Vols. vol. 17, no. 1, pp. 7–19, 2011.

[8] R. Kaur and P. Pandey, "Air Pollution, Climate Change, and Human Health in Indian Cities: A Brief Review," Frontiers in Sustainable Cities, 2021.

[9] M. Nair, H. Bherwani, S. Mirza, S. Anjum and R. Kumar, "Valuing burden of premature mortality attributable to air pollution in major million-plus non-attainment cities of India," Scientifc Reports, vol. 11, p. 22771 , 2021.

[10] S. Selvam, P. Muthukumar, S. Venkatramanan, P. D. Roy, K. M. Bharath and K. Jesuraja, "SARS-CoV-2 pandemic lockdown: Effects on air quality in the industrialized Gujarat state of India," Science of the Total Environment, vol. 737, p. 140391, 2020.

[11] D. K. Singh, T. R. Sahu and A. Kumar, "Assessment of Air Quality and Its Correlation with Meteorological Parameters in Ahmedabad, India," Sustainable Cities and Society, vol. vol. 55, 2019.

[12] Z. Chen, Y. Guo, M. Zhou, J. Tan, W. Jiang and W. Yu, "Ambient air pollution and cardiovascular disease: Understanding the mechanisms and effects.," Journal of Clinical Toxicology & Pharmacology, Vols. 5(5), 12–27, no. 10.4278/jctp.04251001, 2015.

[13] K. Vohra, E. A. Marais, W. J. Bloss, J. Schwartz, L. J. Mickley, M. V. Damme, L. Clarisse and P.-F. Coheur, "Rapid rise in premature mortality due to anthropogenic air pollution in fast-growing tropical cities from 2005 to 2018," Science Advances, vol. 8, no. 14, p. eabm4435, 2022.