International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 82-86, April-2024

https://doi.org/10.47001/IRJIET/2024.804011

Robust Predictive Model to Forecast Air Quality Index Level

[1]Aditya Arolkar, [2]Dhaval Smart, [3]Gaurav Waghmare, [4]Pratham Atale, [5]Prof. Sonali Despande

[1,2,3,4]Student, Smt. Indira Gandhi College of Engineering, Ghansoli, New Mumbai, Maharashtra, India

[5]Professor, Smt. Indira Gandhi College of Engineering, Ghansoli, New Mumbai, Maharashtra, India

Abstract **- The research on air quality index (AQI) prediction in India utilizing machine learning models, particularly the SARIMAX model, highlights the significance of advanced modeling techniques for accurate AQI forecasting. The study incorporates artificial intelligence in AQI prediction based on air pollution data from major Indian cities like Delhi. The dataset used includes attributes like PM 2.5, PM 10, NO, NO_2, CO, SO_2, O_3, and more, with AQI categorized into six levels from good to severe. The research emphasizes the need for comprehensive assessments in urban areas, addressing computational complexities, and integrating real-time data for enhanced forecasting. Various machine learning algorithms like RF, ANN, SVM, and NN have been employed by researchers to predict AQI, with the SARIMAX model being utilized for AQI prediction in cities like Ahmedabad. The study underscores the critical role of accurate AQI prediction in combating air pollution and its adverse effects on public health and the environment in India.**

Keywords: Air quality index (AQI) prediction, Machine Learning, Auto-Regression Module.

I. INTRODUCTION

Air pollution in India is a pressing concern, with the country ranking among the most polluted globally. The situation has worsened due to rapid population growth, unplanned urbanization, and industrialization. The World Health Organization reports that ambient air pollution in India contributes to approximately 17.8% of all fatalities, primarily due to ambient particulate matter (PM) and residential air pollution. Machine learning models have been employed to predict AQI (Air Quality Index) levels, aiding scientists in devising effective emission reduction strategies and identifying pollution trends. The SARIMAX model, a specific machine learning technique, has been utilized for AQI prediction in cities like Delhi. This model, along with other machine learning algorithms like RF, ANN, SVM, and NN, has proven effective in predicting AQI levels, contributing to more accurate and reliable environmental monitoring systems. In this study, we aim to analyze the AQI prediction data provided for Mumbai, Shillong, Lucknow, Delhi, Vishakhapatnam, Patna, and Bhopal using the SARIMAX model. By accurately forecasting AQI levels, authorities can implement effective strategies and measures to safeguard public health from air pollution. This analysis will provide insights into the performance of the SARIMAX model in predicting AQI levels for these cities, contributing to more accurate and reliable environmental monitoring systems.

1.1 Project Aims and Objectives

Project Aim:

The overarching aim of this project is to conduct an in-depth analysis of the AQI (Air Quality Index) prediction data provided for the major cities of Mumbai, Shillong, Lucknow, Delhi, Vishakhapatnam, Patna, and Bhopal using the advanced machine learning technique of the SARIMAX model.

Project Objectives:

1. To gain a comprehensive understanding of the SARIMAX model and its specific application in the context of AQI prediction, including its underlying principles, components, and strengths.
2. To thoroughly analyze the AQI prediction data provided for the aforementioned cities, with a focus on evaluating the accuracy and reliability of the SARIMAX model in predicting AQI levels for these urban areas.
3. To rigorously assess the performance of the SARIMAX model in predicting AQI levels for the selected cities, taking into account various factors such as data quality, model parameters, and computational complexities.
4. To provide detailed insights and recommendations based on the analysis of the AQI prediction data and the performance of the SARIMAX model, with the aim of contributing to more accurate and reliable environmental monitoring systems and informing authorities about effective strategies and measures to safeguard public health from air pollution.
5. To contribute to the broader discourse on air pollution and its adverse effects on public health and the environment in India, highlighting the potential of machine learning models in predicting AQI levels and

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 82-86, April-2024

https://doi.org/10.47001/IRJIET/2024.804011

providing a foundation for further research and development in this critical area.

1.2 System Objectives

- Data Preprocessing: The system should be able to preprocess the AQI data, including cleaning, normalization, and feature selection, to ensure the accuracy and reliability of the model.
- Model Training: The system should be able to train the selected SARIMAX model using the preprocessed data, including the estimation of the model parameters and the calculation of the forecast errors.
- Model Evaluation: The system should be able to evaluate the performance of the trained SARIMAX model, including the accuracy and reliability of the forecasts, using appropriate metrics such as RMSE, MAE, MSE, and MAPE.
- Model Deployment: The system should be able to deploy the trained and evaluated SARIMAX model in a real-world environment, including the integration with other systems and the provision of user-friendly interfaces for data input and forecast output.
- Model Monitoring: The system should be able to monitor the performance of the deployed SARIMAX model over time, including the detection of any changes in the data characteristics and the adjustment of the model parameters accordingly.
- Model Scalability: The system should be able to scale the SARIMAX model to handle larger datasets and more complex data characteristics, including the consideration of independent variables and the integration with other forecasting algorithms.
- Model Interpretability: The system should be able to provide clear and understandable explanations of the SARIMAX model, including the interpretation of the model parameters, the identification of the trends and patterns in the data, and the assessment of the model uncertainty.
- Model Integration: The system should be able to integrate the SARIMAX model with other systems and tools, including data sources, visualization tools, and decision-making systems, to provide a comprehensive solution for AQI prediction and management.

1.3 Project Background

The project background for the topic of predicting AQI (Air Quality Index) using the SARIMAX model with data from 2012 to 2020 involves a comprehensive overview of the issue of air pollution in India and the need for accurate and reliable AQI prediction to address the adverse effects of air pollution on public health and the environment. The historical context of air pollution in India and the availability of AQI data from 2012 to 2020 is presented, highlighting the growing concerns and the evolution of air pollution over time. Prior solutions or efforts related to air pollution and AQI prediction are discussed, including the use of various machine learning models and their effectiveness in predicting AQI levels. Relevant data and research related to air pollution and AQI prediction from 2012 to 2020 are incorporated, citing credible sources. External factors such as economic, social, political, and technological trends that could impact the project's feasibility and success are considered. Legal and regulatory requirements related to air pollution and AQI prediction are mentioned, along with potential risks and challenges and strategies for risk management. The project's specific objectives and goals are clearly defined, explaining how they will address the identified issue or opportunity using the SARIMAX model with data from 2012 to 2020.

II. SOFTWARE COMPONENT

The software components for the project on predicting AQI using the SARIMAX model in Python with the Anaconda environment include:

Python, Anaconda, Jupyter Notebook, Arima/Sarima Model.

These software components provide a powerful and flexible environment for data analysis, modeling, and visualization, and are widely used in various fields, including air quality monitoring and management.

III. METHODOLOGY

- Data Collection: Gather historical AQI data for the selected cities, including past AQI values for Mumbai, Shillong, Patna, Vishakhapatnam, Delhi, and Lucknow.
- Data Preprocessing: Clean the data, removing any missing or inconsistent values. Normalize the data if necessary to ensure that the values are on a similar scale.
- Auto Regression: Apply auto regression techniques to the preprocessed data to identify any underlying patterns or trends.
- Model Selection: Choose the SARIMAX model for AQI prediction. This model is a combination of the ARIMA model and exogenous variables, which can help account for external factors that may influence AQI levels.
- Model Training: Train the SARIMAX model on the preprocessed data. This involves selecting the appropriate parameters for the model, such as the order of the ARIMA component (p, d, q) and the number of exogenous variables.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 82-86, April-2024

https://doi.org/10.47001/IRJIET/2024.804011

- Model Evaluation: Evaluate the performance of the model using appropriate accuracy measures, such as MSE, RMSE, Med AE, Max error, and MAE.
- Model Optimization: Fine-tune the model's parameters using Bayesian optimization to improve its accuracy and efficiency.
- Model Comparison: Compare the performance of the SARIMAX model with other traditional and revolutionary models using IoT data from sensors.
- Model Implementation: Implement the SARIMAX model for real-time AQI prediction in the selected cities.
- Decision-Making: Use the predicted AQI values to inform decision-making and enable effective management of air pollution.

This can help policymakers and environmental organizations make informed decisions to manage air pollution and protect public health.

IV. MATHEMATICAL LOGIC

The Mathematical Logic used in the project is:

- RMSE (Root-mean-square deviation) = sqrt [(Σ(Pi – Oi)²) / n]
- MSE (Mean Squared Error) = Σ(yi − pi)2n
- MAE (Mean Absolute Error) = (1/n) Σ(i=1 to n) |y_i – ŷ_i|
- MAPE = (1/n) * Σ(|actual – forecast| / |actual|) * 100

Where:

Σ – a fancy symbol that means “sum”
n – sample size
actual – the actual data value
forecast – the forecasted data valueR^2 = 1 - RSS/TSS

- R^2 = 1 - RSS/TSS
  Where,
  R^2 = coefficient of determination
  RSS = sum of squares of residuals
  TSS = total sum of squares

- Mean Error = Sum of all error values/Number of records

IV. Results

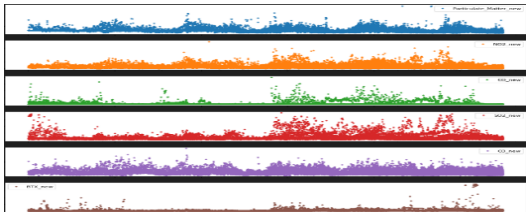

Figure 1: Visualization of Pollutants Year Wise

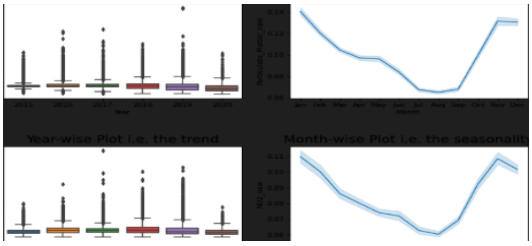

Figure 2: Year Wise i.e trend/ Month Wise i.e Plot

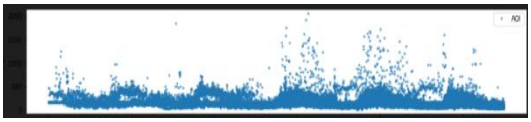

Figure 3: Visualization Of AQI

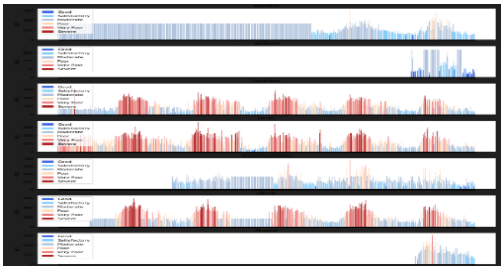

Figure 4: Illustration of AQI Visualizing City Wise

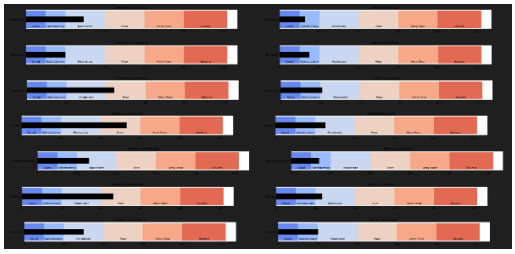

Figure 5: Visualization of AQI before Lockdown vs After Lockdown

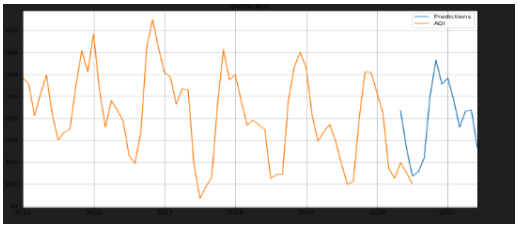

Figure 6: Visualization of AQI Prediction After Analyzing Previous Data

```
Root Mean Squared Error:  33.481045609363534
Mean AQI: 240.71425371223765
Bias: 5.343937
MSE: 1120.9804150962814
MAE: 26.155436064942666
RMSE: 33.481045609363534
MAPE: 0.11530528492842286
r2score: 0.8762677450758353
explainedVariance_score: 0.8794199004816929
me: 86.9858909017222
mpd: 4.1680077450658075
mgd: 0.018491245211032587
```

Figure 7: Error Values after Prediction

ISSN (online): 2581-3048
Volume 8, Issue 4, pp 82-86, April-2024
https://doi.org/10.47001/IRJIET/2024.804011

V. CONCLUSION

The strong rejection of the null hypothesis indicates the significance of the findings. The model shows good performance with a low RMSE and high R-squared score, indicating accurate predictions of AQI values. The analysis provides valuable insights into air quality prediction in Delhi supporting effective air quality management and public health protection.

VI. FUTURE SCOPE

The future scope for predicting AQI in India using the SARIMAX model with data from Mumbai, Shillong, Patna, Vishakhapatnam, Delhi, and Lucknow includes:

- Integration of Real-Time Data: Future research should focus on incorporating real-time data from IoT sensors to enhance the accuracy and reliability of AQI predictions.
- Addressing Computational Complexities: The optimization of machine learning models for AQI prediction should consider computational complexities to ensure efficient and scalable solutions.
- Comprehensive Assessments: The analysis of AQI levels should include risk assessments to identify major emission sources and analyze air pollution concentrations using multivariate models.
- Advanced Modeling Techniques: The use of advanced modeling techniques, such as SARIMAX, can improve the accuracy of AQI predictions and support more effective environmental monitoring systems.
- Machine Learning Algorithms: Machine learning algorithms, such as support vector machine, artificial neural networks, random forest, and stacking ensemble approaches, can be used to predict AQI levels with high accuracy.
- Auto-Regressive Integrated Moving Average (ARIMA) Models: ARIMA models can be used to determine daily mean ambient air pollutant levels and analyze air pollution problems in highly polluted cities like Delhi.
- Time-Series Studies: Time-series studies with spline variables, such as relative humidity, daily mean temperature, and visit date, can be conducted to analyze the impact of these variables on AQI levels.
- Data Analytics and Forecasting: Data analytics and forecasting techniques, such as time series regression forecasting, can be used to predict future values of air pollutants based on previous records.
- Health Impact Analysis: The analysis of the health impacts of air pollution in cities like Ahmedabad can inform decision-making and enable effective management of air pollution.
- Community Needs Assessment and Coalition Building: Community needs assessment and coalition building can support the development of air information and response plans to protect public health.

ACKNOWLEDGEMENT

The authors would like to extend their sincere gratitude to everyone who helped my study project be completed successfully. Before anything else, we want to express our deepest gratitude to the technical team for all of their help and support during the project. They overcame a number of technical obstacles and produced the expected results thanks to their knowledge and commitment. We are also grateful to our professors, research guide, supervisor, and mentor, who provided us with valuable direction and input during the research process. Their sage advice and insightful critiques helped us refine our ideas and raised the bar on our work. Furthermore, we would like to thank our institute and our beloved HOD madam for providing us with the resources and facilities we needed to complete this project. Their assistance and inspiration were crucial in the accomplishment of our research. Last but not least, we would like to express our gratitude to the project team for their cooperation, dedication, and hard work. Their assistance was essential in attaining the project's goals and finishing on schedule. Finally, we would want to express our sincere gratitude to everyone who has supported us on this trip. We could not have accomplished it without their help and contributions, which are priceless.

REFERENCES

[1] S. S. Ganesh, S. H. Modali, S. R. Palreddy, and P. Arulmozhivarman, "Forecasting air quality index using regression models: A case study on delhi and houston," in 2017 International Conference on Trends in Electronics and Informatics (ICEI), 2017, pp. 248–254.

[2] M. Castelli, F. M. Clemente, A. Popovič, S. Silva, and L. Vanneschi, "A machine learning approach to predict air quality in california," Complexity, vol. 2020, 2020.

[3] Kumar and P. Goyal, "Forecasting of air quality in delhi using principal component regression technique," Atmospheric Pollution Research, vol. 2, no. 4, pp. 436–444, 2011.

[4] T. Madan, S. Sagar, and D. Virmani, "Air quality prediction using machine learning algorithms –a review," in 2020 2nd International Conference on

ISSN (online): 2581-3048
Volume 8, Issue 4, pp 82-86, April-2024
https://doi.org/10.47001/IRJIET/2024.804011

Advances in Computing, Communication Control and Networking (ICACCCN), 2020, pp. 140–145.

[5] J. K. Sethi and M. Mittal, "An efficient correlation based adaptive LASSO regression method for air quality index prediction," Earth Science Informatics, vol. 14, no. 4, pp. 1777–1786, Dec. 2021. [Online]. Available: https://doi.org/10.1007/s12145-021-00618-1

[6] Dr, D. Patel, and A. Jain, "Air quality index forecasting using autoregression models," in 2020 IEEE International Students' Conference on Electrical Electronics and Computer Science (SCEECS), 2020, pp. 1–5

[7] Zhang, K., Zheng, Y., & Zhao, H. (2019). Air quality prediction with spatio-temporal data using deep learning models: A case study in Beijing, China. Information Fusion, 46, 268-278.

[8] Wei, J., & Wu, Q. (2020). A novel hybrid model for air quality prediction based on deep learning and random forest. Atmospheric Environment, 234, 117613.

[9] Liang, X., Zou, T., & Li, Z. (2020). Air quality prediction using machine learning and social media data. Environmental Pollution, 259, 113765.

[10] Li, C., Wang, Z., & Tang, J. (2018). Air quality prediction with LSTM based on PCA. In 2018 15th IEEE International Conference on Networking, Sensing and Control (ICNSC) (pp. 1-6). IEEE.

AUTHORS BIOGRAPHY

Aditya Arolkar,
Student, Smt. Indira Gandhi College of Engineering, Ghansoli, New Mumbai, Maharashtra, India.

Dhaval Smart,
Student, Smt. Indira Gandhi College of Engineering, Ghansoli, New Mumbai, Maharashtra, India.

Gaurav Waghmare,
Student, Smt. Indira Gandhi College of Engineering, Ghansoli, New Mumbai, Maharashtra, India.

Pratham Atale,
Student, Smt. Indira Gandhi College of Engineering, Ghansoli, New Mumbai, Maharashtra, India.

Prof. Sonali Despande,
Professor, Smt. Indira Gandhi College of Engineering, Ghansoli, New Mumbai, Maharashtra, India.

Citation of this Article:

Aditya Arolkar, Dhaval Smart, Gaurav Waghmare, Pratham Atale, Prof. Sonali Despande, "Robust Predictive Model to Forecast Air Quality Index Level", Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 8, Issue 4, pp 82-86, April 2024. Article DOI https://doi.org/10.47001/IRJIET/2024.804011

****

