IJARST
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

AIR QUALITY INDEX PREDICTION USING DIFFERENT ML ALGORITHMS

M.Anitha[1],Y.Naga Malleswarao[2],G.Sri Tharun Naidu[3]

#1 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada.
#2 Assistant Professor in the Department of MCA , SRK Institute of Technology, Vijayawada.
#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada

ABSTRACT_ Air quality plays a pivotal role in human health and environmental sustainability, necessitating continuous monitoring and assessment to mitigate potential hazards. This project endeavors to leverage machine learning techniques to predict air quality levels, thereby enhancing public safety measures. Drawing upon the dataset sourced from the Air Quality Index Data Platform, the project aims to develop predictive models capable of generating timely alarms and informing stakeholders about hazardous gas emissions.

The proposed machine learning models encompass Linear Regression and Random Forest Regression. Linear Regression offers a foundational approach for establishing baseline predictions by analyzing linear relationships between input features and air quality indicators. Meanwhile, Random Forest Regression proves efficacious in discerning complex patterns and non-linear correlations within air quality data, thereby enhancing prediction accuracy.

By harnessing the power of machine learning, this project endeavors to fortify public safety measures by anticipating fluctuations in air quality and fostering proactive interventions. Through predictive modeling, alarm systems, regulatory frameworks, and informative outreach initiatives, stakeholders can be empowered to safeguard human well-being and environmental sustainability in the face of air quality challenges

1.INTRODUCTION

Air quality is a paramount concern for public health and environmental sustainability worldwide. The quality of the air we breathe is intricately linked to various factors, including industrial emissions, vehicular exhaust, agricultural activities, and natural phenomena such as wildfires and volcanic eruptions. Poor air quality poses significant risks to human health, contributing to respiratory ailments, cardiovascular diseases, and even premature mortality. Additionally, it can adversely affect ecosystems, leading to biodiversity loss and ecosystem degradation. In recent years, the proliferation of urbanization, industrialization, and transportation has exacerbated air quality issues in many regions. Rapid economic development, coupled with inadequate environmental regulations and enforcement, has led to heightened pollution levels in numerous urban centers globally. Furthermore, climate change-induced phenomena, such as extreme weather events and temperature fluctuations, can exacerbate air pollution levels, exacerbating the challenges faced by communities striving to maintain clean and breathable air.

In response to these challenges, governments, environmental organizations, and public health agencies have intensified efforts to monitor and mitigate air pollution. The establishment of air quality monitoring networks, comprising stationary and mobile monitoring stations equipped with sophisticated sensors, has facilitated real-time data collection and analysis. These networks generate vast amounts of data, encompassing various air pollutants such as particulate matter (PM), nitrogen dioxide (NO2), sulfur dioxide (SO2), carbon monoxide (CO), and ozone (O3).However, the sheer volume and complexity of air quality data present challenges in its interpretation and utilization. Traditional statistical methods may struggle to capture the intricate relationships and nonlinear dynamics inherent in air quality datasets. As such, there is a growing imperative to harness advanced computational techniques, particularly machine learning, to extract meaningful insights and facilitate predictive modeling.

Machine learning, a subfield of artificial intelligence (AI), offers powerful tools and algorithms capable of learning from data, identifying patterns, and making predictions with remarkable accuracy. By leveraging machine learning techniques, researchers and policymakers can develop predictive models to anticipate changes in air quality levels, thereby enabling proactive interventions and public safety measures.

The proposed machine learning approach for air quality prediction holds immense promise in enhancing public safety and environmental stewardship. By analysing historical air quality data alongside auxiliary variables such as meteorological conditions, land use patterns, and industrial activities, machine learning models can discern complex relationships and make informed predictions regarding future air quality levels. These predictions can be instrumental in guiding policymakers, urban planners, and public health officials in implementing targeted interventions to mitigate pollution sources and protect vulnerable populations. Moreover, the integration of machine learning-based alarm systems can facilitate timely notifications and alerts in the event of deteriorating air quality conditions. Such alarms can prompt individuals to take preventive measures, such as reducing outdoor activities or using protective masks, thereby minimizing exposure to harmful pollutants.

The application of machine learning techniques holds great potential in addressing the multifaceted challenges posed by air quality degradation. By leveraging advanced computational methods, stakeholders can enhance their capacity to monitor, predict, and mitigate air pollution, ultimately safeguarding public health and environmental well-being.

2.LITERATURE SURVEY

2.1 AIR QUALITY INDEX USING MACHINE LEARNING[2]

Publication:An international journal of advanced computer technology,2020

Methodology: The system suggested by this study uses machine learning techniques to calculate the air quality index. It takes into account a readily available Kaggle dataset and provides it as an input to the algorithms to determine the value of the Air Quality Index. compared the effectiveness of the algorithms under consideration. Decision Tree, Random Forest, and Support Vector Machine are three examples of machine learning techniques used in this.

Advantages:

Random forest and support vector machine (SVM), produce promising results for air quality index (AQI) level predictions.

These algorithms can also handle large datasets.

Disadvantages:

Random forest contain large number of trees which can make the algorithm too slow and ineffective for real-time predictions.

2.2 Federated Learning for Air Quality Index Prediction using UAV Swarm Networks[3]

Publication: IEEE Xplore,2021

Methodology: The paper proposes a distributed and decentralized Federated

Learning approach within a UAV swarm. Each UAV used its locally gathered data to train a model before transmitting the local model to the central base station. The central base station creates a master model by combining all the UAV&#x0027;s local model weights of the participating UAVs in the FL process and transmits it to all UAV s in the subsequent cycles.

Advantages:

With the help of Unmanned Aerial Vehicle&#x0027;s onboard sensors, we can collect air quality data easily.

It is more efficient than machine learning algorithms.

3.PROPOSED SYSTEM

The proposed system for air quality monitoring integrates advanced technologies, particularly machine learning, to overcome the limitations of the existing system and enhance air quality management efforts. The proposed system comprises several key components, including comprehensive sensor networks, predictive modeling algorithms, real-time data analytics, and user-friendly interfaces for data visualization and dissemination. By leveraging these components, the proposed system aims to provide accurate, timely, and actionable information about air quality levels to stakeholders, thereby facilitating more effective decision-making and intervention strategies..

3.1 IMPLEMENTATION

1. Importing Necessary Libraries:

- Import essential libraries such as NumPy, Pandas, and scikit-learn modules required for data manipulation, model training, and evaluation.

2. Loading the Dataset:

- Load the dataset containing AQI data. In your code, the dataset is loaded using Pandas' `read_csv()` function.

3. Preprocessing (if needed):

- Preprocess the dataset, if necessary, which may include handling missing values, encoding categorical variables, or scaling numerical features. In your code, preprocessing steps such as one-hot encoding are applied to categorical features.

4. Splitting the Dataset:

- Split the dataset into training and testing sets to train the models on one portion of the data and evaluate their performance on another. This is accomplished using scikit-learn's `train_test_split()` function.

5. Model Training and Evaluation:

- Train multiple machine learning models using different algorithms and evaluate their performance. The code provided implements the following algorithms:

- Multiple Linear Regression (MLR)

- Polynomial Regression (PR)

- Decision Tree Regression (DTR)

- Random Forest Regression (RFR)

- Support Vector Regression (SVR)

- For each model, the following steps are performed:

- Model initialization.

- Model training using the training dataset.

- Prediction on the testing dataset.

- Calculation of evaluation metrics such as RMSE, MAE, R-squared, and RMSLE to assess model performance.

- Printing or storing the evaluation metrics for each model.

6. Visualization (Optional):

- Optionally, visualize the results to gain insights into model predictions and compare the performance of different algorithms. This could include plotting actual vs predicted AQI values or comparing model performance metrics graphically.

7. Additional Functions:

- Define any additional functions needed for preprocessing or evaluating the models. In your code, a function `rmsle()` is defined to calculate the Root Mean Squared Logarithmic Error (RMSLE).

8. Final Output:

- Print or display the evaluation metrics for each model to compare their performance.

4.RESUTS AND DISCUSSION

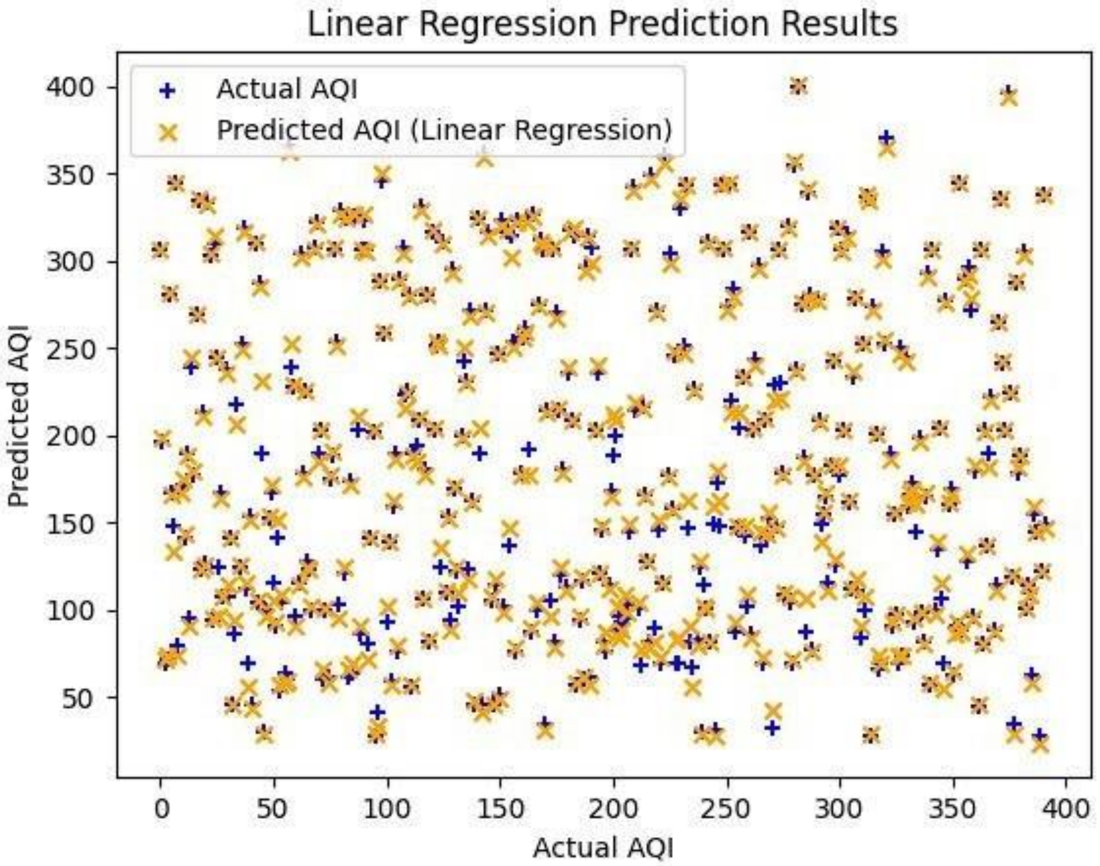

Figure 1: Linear Regression Prediction Results

Scatter Plot for Linear Regression Predictions:

- This plot compares the actual AQI values (blue plus signs) against the predicted AQI values generated by the Linear Regression model (orange 'x' markers).
- Each point represents a data instance.
- Ideally, the points should align closely along the diagonal line, indicating accurate predictions.

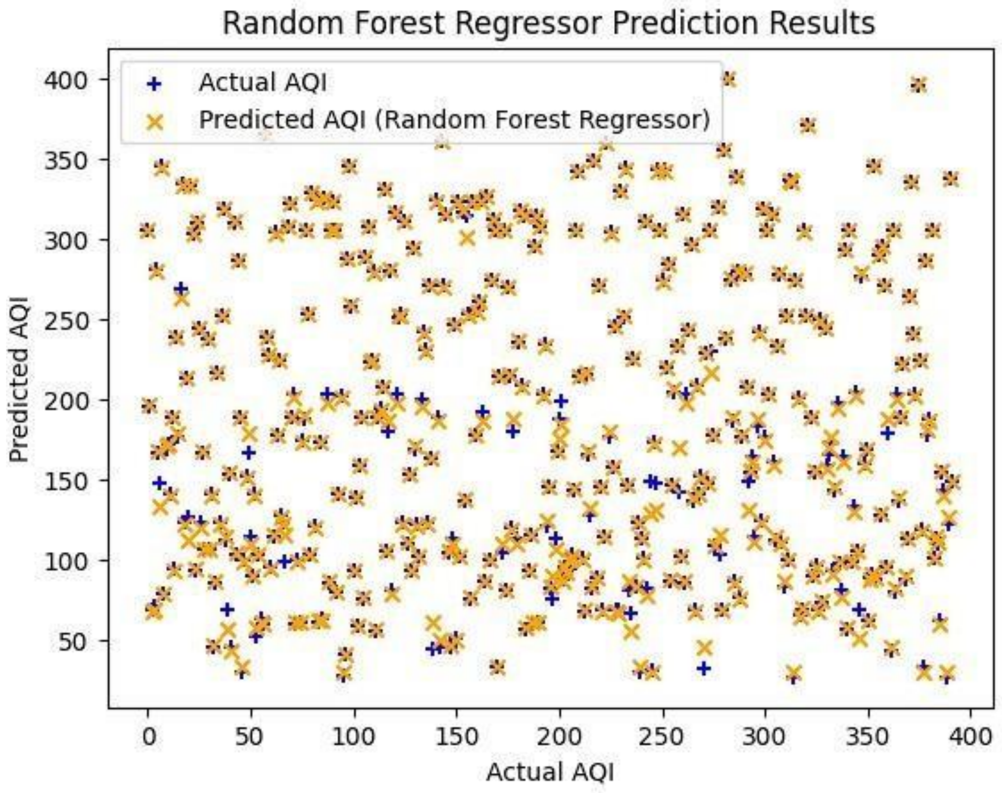

Figure 2: Random Forest Regression Results

Scatter Plot for Random Forest Regression Predictions:

- Similar to the first scatter plot, this one compares the actual AQI values (blue plus signs) with the predicted AQI values from the Random Forest Regression model (orange 'x' markers).
- The alignment of points indicates the accuracy of the predictions made by the Random Forest Regression model.

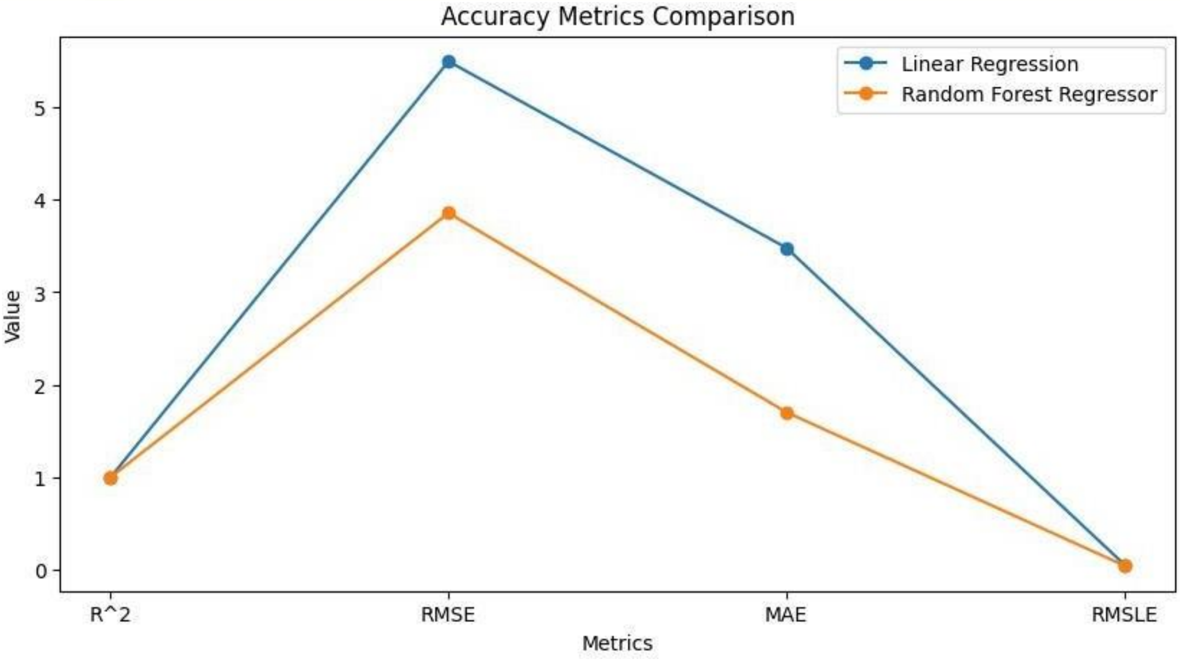

Figure 3: LR Vs RFR Accuracy Comparison

Comparison Plot of Error Metrics:

- This plot shows a comparison of various error metrics (RMSE, MAE, R², and RMSLE) between the Linear Regression and Random Forest Regression models.
- Each metric for both models is represented by a different line (Linear Regression in blue, Random Forest Regression in orange).
- The x-axis represents the different error metrics, and the y-axis represents the corresponding values of the error metrics.
- Lower values of RMSE, MAE, and RMSLE, and higher values of R² indicate better model performance.

5.CONCLUSION

The development and evaluation of machine learning models for predicting Air Quality Index (AQI) have been conducted extensively in this project. The project aimed to leverage various regression algorithms to forecast AQI levels accurately, which is crucial for public safety and environmental well- being. Through the implementation and analysis of different regression techniques, valuable insights have been gained regarding their performance and suitability for AQI prediction tasks.

The project began with a comprehensive understanding of the problem statement, emphasizing the importance of continuous monitoring and assessment of air quality to mitigate potential health risks and environmental impacts. Leveraging machine learning techniques for AQI prediction opens avenues for timely interventions and the establishment of regulatory measures to safeguard public health and the environment.

Various machine learning algorithms were explored and implemented, including Multiple Linear Regression (MLR), Polynomial Regression (PR), Decision Tree Regression (DTR), Random Forest Regression (RFR), and Support Vector Regression (SVR). Each algorithm was evaluated based onits ability to accurately predict AQI levels using appropriate error metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared (R^2), and Root Mean Squared Logarithmic Error (RMSLE).

The results demonstrated notable variations in the performance of different algorithms. MLR and RFR exhibited high levels of accuracy, with low RMSE, MAE, and RMSLE values, as well as high R^2 scores, indicating robust predictive capabilities. On the other hand, PR and SVR showed comparatively poorer performance, with higher error metrics and lower R^2 scores, suggesting limitations in capturing the underlying patterns in the data. Furthermore, visualizations such as scatter plots provided intuitive representations of the model predictions, allowing for a qualitative assessment of the models' performance in predicting AQI levels. The comparison plot of error metrics facilitated a comprehensive evaluation of the models' performance across multiple metrics, aiding in the selection of the most suitable algorithm for AQI prediction tasks.

Overall, the project underscores the significance of machine learning in addressing environmental challenges such as air quality monitoring. By harnessing the power of data-driven approaches, stakeholders can make informed decisions, implement timely interventions, and formulate effective policies to mitigate air pollution and safeguard public health. However, it isessential to acknowledge the limitations and challenges associated with machine learning models, including data quality issues, model interpretability, and the need for continuous refinement and validation. Moving forward, future research directions may include the integration of additional features and data sources, the exploration of advanced machine learning techniques, and the development of ensemble models to further enhance the accuracy and robustness of AQI prediction systems. Additionally, efforts should be made to promote interdisciplinary collaborations between data scientists, environmental experts, policymakers, and community stakeholders to foster innovation and drive meaningful impact in addressing air quality challenges on a global scale.

REFERENCES

1.J. Lee, ”Acoustical perceptions of building occupants on indoor environment quality in naturally-ventilated building facades, ”Journal of Acoustics, vol.4,no.3,2019.

2.K. Nahar, A. Jaradat, M. Atoum, and F. Ibrahim, “Sentiment analysis and classification of arab jordanian facebook comments for jordanian telecom companies using lexicon-based approach and machine learning,” Jordanian J. Comput. Inf. Technol., vol. 6, no. 03, pp. 247–263, 2020.

3.Chhikara P., Tekchandani R., Kumar N., Chamola V., and Guizani M., “Dcnn-ga: A deep neural net architecture for navigation of uav in indoor environment, ” IEEE Internet of Things Journal, vol. 8, no. 6, pp. 4448–4460, 2021.

4.X. Lin, H. Wang, J. Guo and G. Mei, "A Deep Learning Approach Using Graph Neural Networks for Anomaly Detection in Air Quality Data Considering Spatiotemporal Correlations," in IEEE Access, vol. 10, pp. 94074-94088, 2022, DOI: 10.1109/ACCESS.2022.3204284.

5.Fatima Ezzahra Mana, Blaise Kévin Guépié, Raphaèle Deprost, Eric Herber, Igor Nikiforov, "The air pollution monitoring by sequential detection of transient changes", IFAC-papers online, Volume 55, Issue 5,2022, Pages 60-65, ISSN 2405-8963,https://doi.org/10.1016/j.ifacol.2022.07.640.

6.M. Molinara, M. Ferdinandi, G. Cerro, L. Ferrigno and E. Massera, "An End to End Indoor Air Monitoring System Based on Machine Learning and SENSIPLUS Platform," in IEEE Access, vol. 8, pp. 72204-72215, 2020, doi: 10.1109/ACCESS.2020.2987756

7.Z. J. Andersen, L. C. Kristiansen, K. K. Andersen, T. S. Olsen, M. Hvid-berg, S. S. Jensen, et al., "Stroke and Long-Term Exposure to Outdoor Air Pollution From Nitrogen Dioxide", Stroke, vol. 43, no. 2, pp. 320-325, 2019.

AUTHOR PROFILES

Ms. M.Anitha Working as Assistant Professor & Head of Department of MCA ,in SRK Institute of technology in Vijayawada. She done with B.Tech, MCA ,M. Tech in Computer Science .She has 14 years of Teaching experience in SRK Institute of technology, Enikepadu, Vijayawada, NTR District. Her area of interest includes Machine Learning with Python and DBMS.

Mr.Y.Naga Malleswarao Completed his Masters of Technology from JNTUK,MSC(IS) from ANU,BCA from ANU. He has System Administrator ,Networking Administrator and Oracle Administrator. He also a web developer and python developer, Currently working has an Assistant Professor in the department of MCA at SRK Institute of Technology, Enikepadu, NTR District. His area of interest include Artificial Intelligence and Machine Learning.

Mr.G.Sri Tharun Naidu is an MCA Student in the Department of Computer Application at SRK Institute Of Technology, Enikepadu, Vijayawada, NTR District. He has Completed Degree in B.Sc.(computers) from VTJIM & IVTR Degree College in Mangalgiri. His area of interest are DBMS and Machine Learning with Python.