Air Quality Prediction Using Machine Learning

[1]K.Satish kumar, [2]A.Sathvika, [3]P.Sejal, [4]R.Shailaj, [5]A.Shanthi Priya, [6]R.Sharanya Reddy, [7]Thayyaba Khatoon, [8]Manikkannan

[123456]Student, Department of Artificial Intelligence and Machine Learning (AI&ML) Malla Reddy University, Maisammaguda, Hyderabad

[78]Professor, Department of Artificial Intelligence and Machine Learning (AI&ML) Malla Reddy University, Maisammaguda, Hyderabad

Abstract- Machine Learning: To predict the Air Quality Index (AQI) for different locations in India. This involves using historical data on major pollutants like NO2, SO2, RSPM, and SPM, collected from the Indian government database. The goal is to develop a forecasting model that can estimate future AQI values for a given region. To analyse the data and extract valuable information. The model will consider factors like pollutant concentrations, meteorological data, industrial activities, and geographical features. By employing machine learning algorithms such as regression or time series analysis, you aim to create a reliable model that can predict air quality levels. This information is crucial for targeted efforts to improve air quality in different regions of India.

1. INTRODUCTION

1.1 Machine Learning

In the pursuit of a healthier and sustainable future, our project, focuses on the critical examination of air quality in diverse locations across India. The escalating impact of air pollution on public health and the environment necessitates a proactive understanding of pollutant dynamics and their implications. Leveraging data from the Indian government database, we aim to develop a robust forecasting model for the Air Quality Index (AQI), shedding light on future air quality conditions. Our approach involves a meticulous analysis of historical data, encompassing key pollutants such as NO2, SO2, RSPM, and SPM. By integrating meteorological variables, industrial activities, and geographical attributes, we seek to create a comprehensive model employing advanced machine learning techniques. This model not only predicts AQI values but also gives us information of air whether it is good, bad, average or hazardous.

2. LITERATURE REVIEW

2.1 Machine Learning

Reviewed various research papers, articles, or studies that have employed KNN for predicting AQI or analyzing air quality. Summarized the methodologies, datasets used, features selected, and the performance of KNN in these studies.

3. PROBLEM STATEMENT

3.1 Machine Learning

In the quest for a cleaner and healthier environment, our project, embarks on a comprehensive exploration of air quality dynamics in diverse Indian locations. The escalating impact of air pollution on public health demands a proactive understanding of pollutant intricacies and their future trends. Harnessing data from the Indian government, we strive to construct a robust predictive model for the Air Quality Index (AQI), including the complexities of major pollutants like NO2, SO2, RSPM, and SPM.

4. DATASET DESCRIPTION

4.1 Machine Learning

The dataset contains information related to air quality and environmental parameters, which are crucial for predicting the Air Quality Index (AQI). It includes data for various states, locations, and types of air quality monitoring stations.
Key columns in the dataset include:
state: The state in India where the monitoring station is located.
location: The specific location or city within the state.

type: The type of air quality monitoring station, such as residential, industrial, or rural areas.
so2: Sulfur Dioxide (SO2) levels in the air.
no2: Nitrogen Dioxide (NO2) levels in the air.
rspm: Respirable Suspended Particulate Matter (RSPM) levels.
spm: Suspended Particulate Matter (SPM) levels.
pm2_5: Particulate Matter with a diameter of 2.5 micrometers or less.

5. METHEDOLOGY

5.1 Machine Learning

1. Data Collection:
Gather comprehensive datasets from reliable sources, such as government environmental agencies, containing information on major pollutants (e.g., NO2, SO2, RSPM, SPM), meteorological variables, and geographical features across diverse locations in India.
2. Data Preprocessing:
Feature Engineering for Air Quality Indices
Calculation of Air Quality Index (AQI)
Categorizing AQI into Ranges
3. Feature Selection:
Identify and select relevant features based on domain knowledge and exploratory data analysis. Evaluate the importance of features using techniques like correlation analysis, feature ranking, and feature importance scores from machine learning models.
4. Model Selection:
Choose appropriate machine learning algorithms for air quality prediction. Common choices include:
Regression models, Time series models for capturing temporal patterns. Ensemble methods for robust predictions.
5. Model Training:
Split the dataset into training and validation sets to train and assess model performance. Fine-tune hyperparameters using techniques like cross-validation to prevent overfitting and enhance generalization.
6.Model Evaluation:
Assess the model's performance using relevant metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared for regression models.
Utilize validation datasets to ensure the model's robustness and generalization.
7. Interpretability and Visualization:
Enhance model interpretability through techniques values or partial dependence plots. Visualize predictions against actual values to understand model behavior and identify potential areas for improvement.

6 EXPERMENTAL RESULTS

6.1 Machine Learning

K-Nearest Neighbors (KNN) Model Evaluation:
Root Mean Squared Error (RMSE): Measures the average magnitude of the errors between predicted and actual values.
R-squared (R^2): Assesses the proportion of variance in the dependent variable explained by the independent variable(s).
Classifier Metrics:
Model Accuracy: Measures the percentage of correctly predicted AQI categories.
Kappa Score: Assesses the agreement between predicted and actual AQI categories while considering the possibility of agreement occurring by chance.

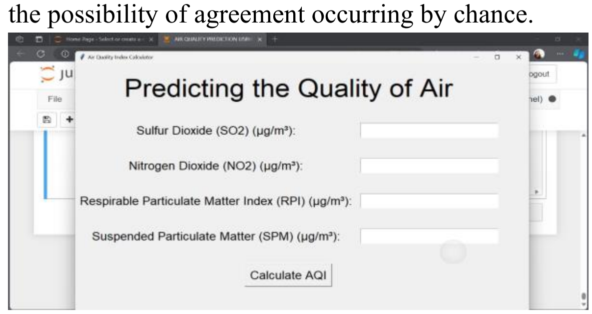

Fig 6.1 Output screen

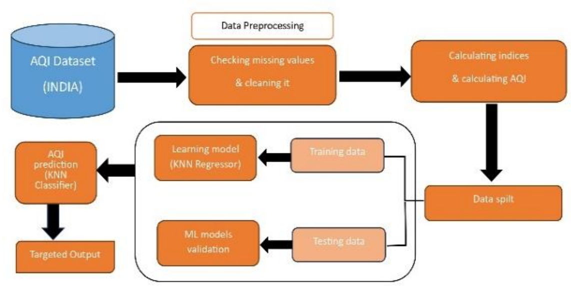

Fig 6.2 Architecture Diagram

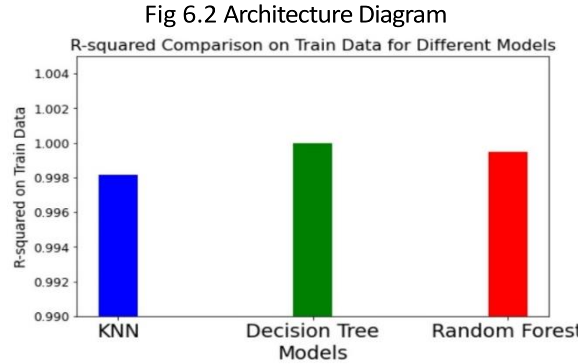

Fig 6.3 R-Squared Comparison on Train Data for Different Models

7. CONCLUSION

7.1 Machine Learning

In conclusion, the application of machine learning in air quality analysis represents a significant stride toward enhancing our understanding, prediction, and management of atmospheric conditions. The integration of advanced computational techniques enables a more nuanced and accurate interpretation of complex relationships within air quality data. Through an extensive review of literature and a methodical exploration of methodologies, several key observations and implications emerge.

8. FUTURE ENHANCEMENT

8.1 Machine Learning

1. Integration of Satellite Data: Explore the integration of satellite data for a more comprehensive understanding of atmospheric conditions. Remote sensing technologies can provide valuable insights into air quality parameters, contributing to more accurate and expansive models.
2. Real-time Adaptive Models: Develop real-time adaptive models that can continuously update and adjust predictions based on evolving conditions. This would enable more responsive and dynamic air quality management strategies.
3. Urban Planning and Policy Integration: Collaborate with urban planners and policymakers to integrate air quality models into city planning and policy decisions. Machine learning insights can inform sustainable urban development and pollution control measures.

9. REFERENCES

9.1 Machine Learning

[1] Dragomir, Elia Georgiana. "Air quality index prediction using K-nearest neighbour technique no.1 (2010): 103-108.

[2] Sivacoumar R, et al, “Air pollution modelling for an industrial complex and model performance evaluation “, Environmental Pollution 111.3 (2001): 471-477

[3] S. Halsana, “Air quality prediction model using supervised machine learning algorithms,” International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 8, pp. 190–201, 2020.

[4] C. R. Aditya, C. R. Deshmukh, N. D K, P. Gandhi, and V. astu,“Detection and prediction of air pollution using machine learning models,” International Journal of Engineering Trends and Technology, vol. 59, no. 4, pp. 204–207, 2018.

[5] M. Londhe, “Data mining and machine learning approach for air quality index prediction,” International Journal of Engineering and Applied Physics, vol. 1, no. 2, pp. 136–153, May 2021.