International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Air Quality Prediction and Analysis System

Dr.Rashmi R Deshpande
Associate Professor & HOD
Department of Information Science and Engineering
AMC Engineering College
Bengaluru, India
rashmideshpande@amceducation.in

Ajay A
Department of Information Science and Engineering
AMC Engineering College
Bengaluru, India
ajay03.aiyaswamy@gmail.com

Akash K L
Department of Information Science and Engineering
AMC Engineering College
Bengaluru, India
akashklakashkl2004@gmail.com

Ibrahim
Department of Information Science and Engineering
AMC Engineering College
Bengaluru, India
saqibibrahim74@gmail.com

Mahadevappa
Department of Information Science and Engineering
AMC Engineering College
Bengaluru, India
mahadevappapatil01@gmail.com

Abstract: Air pollution is essentially the discharge of pollutants into the atmosphere that affect the earth and human health. It is undoubtedly one of the most serious dangers that humanity has ever faced. As a result, forests, crops, animals, and other things suffer. To prevent this issue, the transportation sector must forecast air quality using toxins using machine learning techniques. As a result, forecasting and assessment of air quality have grown in popularity as research subjects. The objective is to investigate machine learning-based methods for the most accurate air quality prediction.

The dataset will be evaluated using the supervised machine learning technique (SMLT) to extract information such as Variable Identification, Uni-variate analysis, Bi-variate & Multi-variate analysis, Missing Values treatments, and to determine where data validation, data preparation and/or cleaning, and data visualization will take place. Our research can serve as a thorough manual for conducting a sensitivity analysis of model parameters in connection to the performance of accuracy calculation-based air quality pollution prediction.

By utilizing "prediction results in terms of best accuracy from comparing supervise classification machine learning model," an accurate Air Quality Index value can be estimated using a machine learning technique. Additionally, it is feasible to compare and discuss how well different machine learning algorithms perform when handling such a little amount of available data from a given transport traffic department dataset using an assessment GUI-based user interface for air quality attribute prediction.

I INTRODUCTION

Air pollution is a serious environmental issue emerging in cities all over the world. The increasing level of pollutants in the atmosphere is creating a serious threat to the health of people living in such cities. To deal with this issue, there is a growing need for precise air quality analysis systems. As reported in the World Health Organization, air pollution contributes to 7 million early deaths each year. Air pollution is most critical in cities since all pollutants will be concentrated in a single place because of industries, cars, and other sources of pollution. The Air Quality Prediction and Analysis System will address this challenge by utilizing machine learning algorithms and sensor data to make predictions and analysis of air quality. The Air Quality Prediction and Analysis System will incorporate data. friendly interface for viewing air quality information, warning messages, and advice.

With the accurate and precise forecasting of air quality, this system will help improve public health, reduce healthcare spending, and promote a better quality of life for people living in cities. Furthermore, this system will help policymakers in developing strategies to reduce air pollution and achieve sustainable development in cities. To effectively implement the Air Quality Prediction and Analysis System, it will be necessary to make a collective attempt at a government level, an R&D level, and other private sectors. The Air Quality Prediction and Analysis System will be developed in such a way that it can be scalable as well as adaptable.

RELATED WORK

Air quality analysis and forecasting have received increased attention because of their significance in public health, city planning, and environmental conservation. Several methods have emerged in literature, from statistics to rather advanced deep learning and IoT solutions.

1. Statistical and Time-Series Models

The early studies conducted on air quality forecasting used statistical methods such as Linear Regression, Autoregressive Integrated Moving Average (ARIMA), and Seasonal ARIMA (SARIMA) models. The above models were very successful in modeling temporal variations of pollutants such as PM2.5, PM10, SO_2, NO_2, and O_3. But these models were not very efficient in handling nonlinear relationships or multiple interactions among meteorological variables and pollutant variables.

2. Methods of Machine Learning

To counter these shortcomings in conventional models, new machine learning algorithms such as Support Vector Machines (SVM), Random Forest (RF), Decision Trees, k-Nearest Neighbors (k-NN), and Gradient Boosting were used. They produced better predictive results using machine learning capabilities for identifying non-linear patterns in historical air quality and meteorology measurements. Random Forest algorithms were proved to be less affected by noise in measurements and were efficient in calculating variable importance.

2. Deep Learning Models

With the increasing availability of massive air quality datasets, deep learning methods have gained immense popularity. Deep Neural Networks (DNN) and Artificial Neural Networks (ANN).

3. Spatial-Temporal Air Quality

Some studies concentrated on spatial and temporal modeling using a combination of information from different monitoring stations. Methods such as Graph Neural Networks, spatial interpolation, or geostatistics were applied to capture dependencies among different geographic locations of the monitoring stations. Such methods greatly enhanced accuracy in air quality forecasting in an urban and regional setting.

4. IoT Based Air Quality Monitoring Systems

Recently, breakthroughs in Internet of Things (IoT) technology have allowed for real-time air quality analysis using low-cost sensors. A conventional IoT system comprises gas sensors, microcontrollers, cloud solutions, and data analysis components. When integrated with machine learning model solutions or deep learning algorithms, IoT systems allow for real-time analysis, notifications, and forecasting capabilities in a smart city environment.

5. Air Quality Index (AQI) Analysis & Visualization

A common focus of many studies has been on computation and visualization of AQI calculation to improve awareness. Online portals and apps have come up with designs to represent air pollutant concentrations, AQI, and health advice. Combination of forecasting tools with visualization systems has proved effective in improving decision-making for both authorities and civilians.

6. Research G

Although progress has been made, some challenges arise in dealing with missing sensor data, accuracy in sensor calibration, and improving model generalization. Moreover, incorporating ideas from Explainable AI (XAI) methods in air quality forecasting is a research problem in its own right.

II PROBLEM STATEMENT

Air pollution has become a critical environmental and public health issue due to Rapid urbanization, large-scale industrialization, and rising vehicular emissions have transformed air pollution into an essential environmental and human health concern in recent years. These lead to respiratory diseases, cardiovascular disorders, shortened life span, and quality of life. While there is no dearth of air quality monitoring stations, most of the existing systems focus on real-time reporting rather than forecasting and providing actionable insight for prevention. Some of the major issues with the use of current air quality monitoring methods include limited spatial coverage, delayed data availability, a lack of predictive capabilities, and incomplete integration of multiple sources of pollution along with meteorology.

There is an urgent need for an intelligent Air Quality Prediction and Analysis System for collecting and analyzing historical and real-time air pollutant data like PM2.5, PM10, CO, NO_2, SO_2, and O_3, along with meteorological parameters such as rainfall, wind speed, humidity, and temperature. The system must accurately forecast future air quality levels and categorize them using standardized Air Quality Index (AQI) categories by utilizing data analytics and machine learning approaches. The suggested system uses an intuitive interface to provide trend analysis, timely predictions, and visual insights. In order to lower health risks, residents would be able to take preventive action thanks to early warnings from healthcare providers and policy formation by government authorities. The purpose of this system is to facilitate data-driven decision-making to enhance management results in the fields of public health and the environment. in the immediate area

III SYSTEM DESIGN OVERVIEW

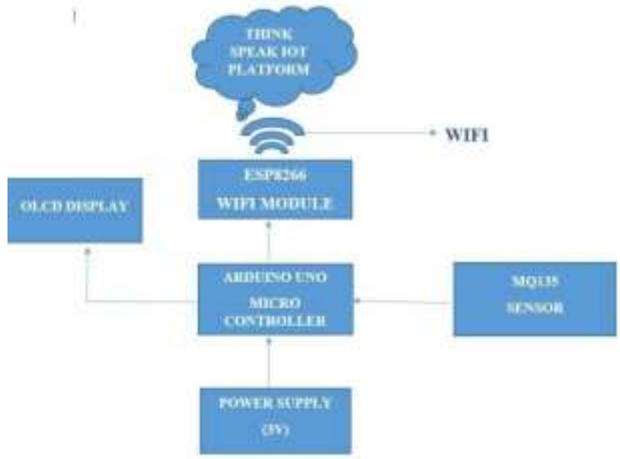


Fig 1.1. Block diagram

A. Software Architecture

A layered software architecture was used in the design of the Air Quality Prediction and Analysis System to provide modularity, scalability, and ease of maintenance. The system begins with a data acquisition layer that collects current and past data on air quality from sensors measuring PM2.5, PM10, CO, NO_2, SO_2, and O_3 air pollutants as well as temperature, humidity, wind speed, and rainfall meteorological parameters. Once the sensor data has been properly validated for storage in suitable databases, the data is sent to the data ingestion and storage layer.

The data processing layer cleans and preprocesses the data gathered by handling missing values, removing noise, normalizing pollutant concentrations, and extracting features relevant to analysis. The data processed will thereafter be passed to the prediction and analysis layer, which forms the core of the system. This layer uses machine learning and statistical models for forecasting future levels of air quality, computing the Air Quality Index, and analyzing pollution trends and patterns to identify potential health risks.

The application logic layer controls the various system workflows and decision-making processes for tasks such as user authentication and authorization, access control, and the generation of alerts when air quality becomes unhealthy. Finally, the presentation layer, often implemented as an online dashboard, is employed to convey real-time air quality status, predicted AQI values, trends over time, and health advisories in an easily understandable format. These three interconnected layers form a robust and effective software architecture that meets high levels of accuracy in air quality prediction, effective analysis, and informed decision-making.

B. Functional Workflow

1. Data collection
The model presupposes a starting point where air quality and meteorology information is obtained from trusted sources such as air quality stations operated by government bodies, IoT devices, or freely available datasets. Among other facts gathered, these include concentrations of pollutants such as PM2.5, PM10, CO, NO_2, SO_2, and O_3, as well as meteorological data like temperature, humidity, wind speed, and pressure.

2. Data Preprocessing
In this phase, gathered data is handled and preprocessed for machine learning. Techniques like mean substitution and interpolation are used to impute missing values in the data. Data that is incorrect or noisy is removed. Scaling uniform dimensions is the next step in data preprocessing. Additional duties include managing time stamps to extract pertinent information like day, month, or season, and encoding nominal variables if they exist in the collection.

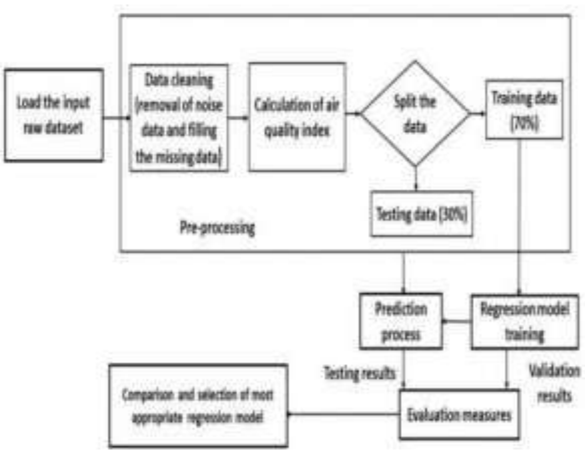


Fig 1.2: Flow chart illustrates the basic operations from user input to system response.

3. Exploratory Data Analysis (EDA
Exploratory ata Analysis EDA analysis will be used in order to better understand existing patterns, trends, and relationships in the dataset. Line charts, Heatmaps, and Correlation matrices will be used to analyze the behavior of pollutants over time, and how these pollutants are dependent on meteorological factors.

4. Feature Selection & Engineering
Specific features are considered for addition using correlation analysis and knowledge of the domain in order to improve performance. Features such as rolling averages, lag variables, or Air Quality Indicators (AQIs) are derived in order to incorporate temporal relationships. The selection of features reduces dimensionality and improves accuracy and efficiency in developing a predictive model.

5. Model Selection & Training
Such machine learning algorithms will be appropriately chosen based on the problem type, which can include algorithms such as Linear Regression, Random Forest, Support Vector Machines, or Neural Networks. A split in the dataset will be conducted to obtain a separate testing set and a model will be trained using a historical dataset. Hyperparameter tuning will be performed.

6. Air Quality Prediction
Once trained, this model may predict future pollution concentrations or air quality levels using the existing input data. Based on Air Quality Index standards, the model can predict short-term or long-term air quality and categorize it into levels like "Good," "Moderate," and "Poor."

7. Prediction Module Implementation
A prediction module that uses real-time or user-submitted data can incorporate the learned model. Before the trained model processes the incoming data, it undergoes preprocessing and feature selection. The model predicts pollutant concentrations or air quality indexes and classifies air quality into pre-established categories like Good, Moderate, and Poor.

8. Model Evaluation Implementation
The performance of a model is measured using statistics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Square Error (RMSE). The results obtained from model evaluation can be used for model selection based on accuracy.

9. Visualization and Analysis Implementation
Tools for visualization are used for the purpose of illustrating air quality trends and predictions. Graphs and charts can be produced to show air quality indicators based on historical pollution levels and forecasted air quality information.

10. Deployment Implementation
A final model is used in a web application using Flask. Data processing and predictions are performed by the back end of an application, and a front end is used for interaction with the application. The application can be deployed on a cloud in order to access it in real time.

11. System Maintenance
The system provides support for periodic retraining of machine learning algorithms with updated data in order to keep making accurate predictions. Logging and tracking features have been incorporated into the system to monitor performance.

IV IMPLEMENTATION

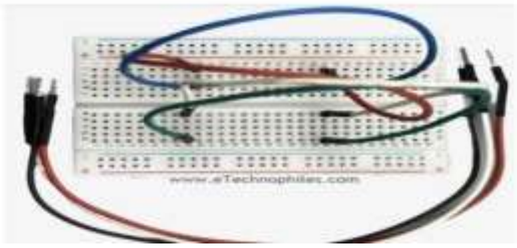
This air quality prediction and analysis system will be implemented using Python, reinforcing libraries for data handling like Pandas and NumPy, machine learning with Scikit-learn, and visualization with Matplotlib/Seaborn. In this context, air quality and meteorological data has to be gathered from publicly available datasets or sensors (including PM2.5, PM10, NO_2, SO_2, CO, O_3, temperature, and humidity) and stored in comma-separated value format. The gathered data must undergo preprocessing for missing values handling, removal of duplicate records and outliers, and feature scale normalization. Feature engineering methods are performed to form time-based attributes, rolling averages, lag features, and Air Quality Index values. In the machine learning model training section, methods like Linear Regression, Random Forest, and Support Vector Regression will be fitted with historical data. Hyperparameter tuning is also done as part of this step to enhance prediction accuracy.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

V RESULTS AND DISCUSSION

The implemented air quality prediction system successfully forecasts pollutant concentrations and Air Quality Index (AQI) levels using the proposed AQI system that uses machine learning models like Random Forest and Support Vector Regression to predict pollutant concentrations and AQI levels. The values of the model evaluation metrics, such as MAE, MSE, and RMSE, represent high accuracy of the system while predicting daily and seasonal variations of pollutants-PM2.5, PM10, NO_2, SO_2, CO, and O_3. Visualizations of the actual values versus predicted values reveal that this model can track the variation in air quality over time, confirming the preprocessing and feature engineering stages. The standard AQI ratings of Good, Moderate, and Poor are then used to classify the air quality, making it simpler for consumers to understand and respond appropriately. The discussion highlights how the Random Forest model's ability to capture nonlinear correlations between meteorological data and gaseous pollutants makes it perform far better than the more straightforward Linear Regression model. Similarly, adding lag variables, rolling averages, and temporal factors improves the model's prediction power even further. These findings show that data-driven machine learning techniques have a good potential for real-time monitoring while also generating dependable and helpful insight into air quality management.

CONNECTORS



NODEMCU ESP8266 WIFI BOARD

MQ135 HAZARDOUS SENSOR

BREAD BOARD

OLED DISPLAY

RESULT:

Channel Stats

Created: about a month ago

Last entry: about a month ago

Entries: 32

The graph shows noticeable variations in AQI levels across the observed period. On 1st November, AQI values are relatively low, indicating good to moderate air quality. However, around 5th November, the AQI values increase significantly, reaching higher peaks. This indicates a temporary degradation in air quality, possibly due to environmental factors such as increased traffic, industrial emissions, or weather conditions.

Multiple data points on the same date suggest that several readings were taken in a day, helping in capturing short-term fluctuations in air pollution levels. The upward trend from early November to mid-November indicates a rising pollution level, which may require preventive measures.

Overall, the graph demonstrates how IoT-based monitoring systems can effectively track air quality in real time. Such visualization helps in early pollution detection, trend analysis, and decision-making for environmental management and public health safety.

VI BENEFITS AND DRAWBACKS

There are numerous advantages to the Air Quality Prediction and Analysis System. The method enables highly accurate forecasting of pollution concentrations and the Air Quality Index, enabling proactive measures to save the environment and public health. Real-time analysis using sensor inputs or API is made feasible by the Air Quality Prediction and Analysis System. By categorizing air quality into familiar categories like Good, Moderately Polluted, and Poor, the approach raises public awareness. The concentrations of various pollutants over time and in relation to various climatic parameters can be analyzed using the Air Quality Prediction and Analysis System.

Although the system has many advantages, it also contains a few and availability of data, and this system may not be able to identify sudden cases of pollution in a particular region. Also, if sensor measurements are not sufficient in a particular region, this system will not be able to identify local pollution sources. In addition, if noisy sensor measurements are involved, this system may predict inaccurate results without proper preprocessing.
The provided graph, which was created using the ThingSpeak IoT cloud platform, shows how the Air Quality Index (AQI) has changed over time. The date of observation is shown on the x-axis, while the AQI values are represented on the y-axis. Air quality sensors gather the data displayed in the graph, which is then periodically uploaded to the cloud for analysis and visualization.

From the graph, it can be observed that on 1st November, the AQI values remain relatively low, indicating good to satisfactory air quality. These lower values suggest minimal presence of harmful pollutants in the atmosphere during that period. However, as the dates progress, especially around 5th November, there is a noticeable increase in AQI values, with some readings reaching peak levels. This rise indicates poor air quality conditions, which may negatively impact human health, especially for children, elderly people, and individuals with respiratory problems.

The presence of multiple data points on the same date indicates that the system records real-time or near real-time data at regular intervals. This helps in capturing sudden changes in air pollution levels caused by factors such as traffic congestion, industrial activities, weather changes, or festive events like firecracker usage.
The graph's increasing trend, which may be utilized as an input for air quality prediction models, indicates a decline in air quality over time. Future pollution levels can be anticipated by using machine learning algorithms on historical AQI data. This makes it possible for both individuals and authorities to take remedial and preventive actions, such lowering emission sources or issuing health recommendations.

Overall, this graph demonstrates the effectiveness of an IoT-based air quality monitoring and prediction system. Continuous monitoring, cloud-based visualization, and predictive analysis together help in improving environmental awareness, public health safety, and smart city management.

CONCLUSION

The Air Quality Prediction and Analysis System is an effective tool used for identifying, interpreting, and predicting concentrations of air pollutants and Air Quality Index (AQI) via machine learning algorithms. Through the combination of historical and current information, this system promotes accurate predictions and air quality indexes, which can be interpreted in a comprehensible manner through diferent levels, thus facilitating proactive steps in environmental management and public health. The Air Quality Prediction and Analysis System portrays the capability of technology-driven solutions in assisting policymakers, academicians, and people in making informed decisions on air pollution control measures. With additional support in terms of using other information repositories and automatic notification systems, this system can prove to be a competent tool for continuous environmental management.

VII FUTURE SCOPE

There are numerous approaches to increase the efficacy and utilty of the Air Quality Prediction & Analysis System. To increase forecast accuracy, future improvements can involve adding more real-time data inputs like satellite imagery and IoT-capable devices. Air quality predictions can be enhanced by using deep learning and very sophisticated machine learning algorithms like GRU and LSTM. Predictive alerts for hazardous air quality events could be added to the system to enable appropriate preventive action. Furthermore, integrating smart city technology and mobile apps can improve accessibility and provide thorough environmental research across various geographies.

REFERENCE

[1] Temesegan Walelign Ayele, Rutvik Mehta, "Air pollution monitoring and
International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018
[2] Saba Ameer, Munam Ali Shah, Abid Khan, Houbing Song, Carsten Maple, SaifUl Islam
Nabeel Asghar, "Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart"
Cities" in IEEE Access (Volume: 7), 2019
[3] Yi-Ting Tsai, Yu-Ren Zeng, Yue-Shan Chang, "Air Pollution Forecasting Using RNN
IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive
Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and
Technology Congress, 2018
[4] V.R. Pasupuleti, Uhasri, P. Kalyan, Srikanth, Hari Kiran Reddy
Data Log By Machine Learning", 6th International Conference on Advanced Computing and
Arahne Information System Forum,
[5] Shengdong Du, Tianrui Li, Yan Yang, Shi-Jinn Horng, "Deep Air Quality Forecasting Using Hybrid
"DeepLearning Framework", Transactions on Knowledge & Data Engineering (Volume: 33, Issue
6), 2021 Ke Gu, JunfeiQiao, Weisi Lin, "Recurent Air Quality Predictor Based on Meteorology-Pollution-Related Factors", in IEEE Transactions on Industrial Informatics, vol. 14