IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14439

AIR QUALITY DATA ANALYSIS AND DYNAMIC VISUALIZATION

Roshan Zameer M[1], Ms. Hemalatha A[2]

M.Sc. Data Science and Business Analysis, Rathinam College of Arts and Science Coimbatore – 641021[1]

Assistant Professor, Department of Computer Science, Coimbatore – 641021[2]

Abstract: Air Quality Analysis is a project aimed at monitoring and evaluating air pollution levels using key parameters such as PM2.5, PM10, CO, SO_2, NO_2, and O_3. The system collects and analyses real-time or historical data to identify patterns and trends in air quality. Using data analysis and visualization techniques, the project provides insights into pollution levels and their impact on health and the environment. The results help in predicting air quality changes and support effective decision-making for pollution control and environmental sustainability.

Keywords: Air Quality Analysis, Air Pollution, PM2.5, PM10, Data Analytics, Environmental Monitoring, Pollution Prediction, Machine Learning

1.INTRODUCTION

Air pollution is one of the most serious environmental challenges faced globally, affecting both human health and ecological balance. Rapid urbanization, industrial growth, and an increase in vehicular emissions have significantly contributed to the deterioration of air quality. Poor air quality is associated with various health issues such as respiratory diseases, cardiovascular problems, and reduced life expectancy. Therefore, monitoring and analysing air quality has become essential for sustainable development and public safety.

The Air Quality Analysis project focuses on studying the concentration of various pollutants present in the atmosphere, including particulate matter (PM2.5 and PM10), carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and ozone (O_3). By collecting and analysing air quality data, this project aims to identify pollution patterns, seasonal variations, and potential sources of contamination.

Modern technologies such as data analytics, machine learning, and visualization tools play a key role in understanding air pollution trends. This project helps in predicting future air quality levels and provides insights that can assist policymakers and environmental agencies. Ultimately, it promotes awareness and encourages actions to reduce pollution and improve overall air quality.

2.PROBLEM STATEMENT

Air pollution has become a critical issue in many urban and industrial areas due to increasing human activities and lack of proper environmental regulations. Despite various efforts, monitoring and controlling air pollution remains a challenge because of the complex nature of pollutants and their varying sources. Many regions lack efficient systems to continuously monitor air quality and provide real-time updates to the public.

One of the major problems is the unavailability of accurate and timely air quality data, which makes it difficult to assess pollution levels and take preventive measures. In addition, the existing systems often fail to analyse large volumes of data effectively and do not provide reliable predictions for future air quality conditions. This leads to poor decision-making and limited awareness among people regarding the severity of air pollution.

The objective of this project is to address these issues by developing a system that can collect, process, and analyse air quality data efficiently. It aims to provide meaningful insights through data visualization and predictive analysis. By doing so, the project helps in identifying pollution trends, forecasting future conditions, and supporting authorities in implementing effective environmental policies to reduce air pollution.

3.DATASET ACQUISITION

Dataset acquisition is a crucial step in the Air Quality Analysis project, as the accuracy and reliability of the results depend

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14439

heavily on the quality of the data collected. In this project, air quality data can be obtained from various trusted sources such as government environmental agencies, public datasets, and online platforms that provide real-time and historical pollution data.
The dataset typically includes information on key air pollutants such as PM2.5, PM10, CO, NO_2, SO_2, and O_3, along with additional parameters like temperature, humidity, wind speed, and location details. These factors play an important role in understanding the variations in air quality and identifying patterns over time.

Once the data is collected, it undergoes preprocessing steps such as cleaning, handling missing values, and removing inconsistencies to ensure accuracy. The data may also be transformed into suitable formats for analysis. Reliable datasets enable better visualization, pattern recognition, and prediction of air quality levels.

Proper dataset acquisition ensures that the project produces meaningful and accurate insights, which can be used for monitoring environmental conditions, raising awareness, and supporting decision-making processes aimed at reducing air pollution and improving public health.

4.METHODOLOGY

The methodology of the Air Quality Analysis project involves a systematic approach to collect, process, analyse, and interpret air quality data to generate meaningful insights. The process is divided into several key steps to ensure accuracy and efficiency.

Data collection:

The first step involves gathering air quality data from reliable sources such as government databases, environmental monitoring agencies, or online platforms. The dataset includes pollutant parameters like PM2.5, PM10, CO, NO_2, SO_2, and O_3 along with meteorological data such as temperature and humidity.

Data Preprocessing:

The collected data is cleaned to remove missing values, duplicates, and inconsistencies. Data transformation and normalization techniques are applied to prepare the dataset for analysis.
This step ensures the quality and reliability of the data.

Exploratory Data Analysis (EDA):

In this stage, statistical and visualization techniques are used to understand the data. Graphs, charts, and correlation analysis help identify patterns, trends, and relationships between pollutants.

Model Development:

Machine learning algorithms such as Linear Regression, Decision Trees, or Random Forest may be used to predict air quality levels. These models are trained using historical data.

Evaluation and Validation

The performance of the models is evaluated using metrics such as accuracy, mean squared error, or R-squared values to ensure reliable predictions.

Visualization and Reporting:

The analysed results are presented using dashboards, charts, and reports to make the information easy to understand for users and decision-makers.

Conclusion and Decision Support:

The final step involves interpreting the results to provide insights and recommendations for improving air quality and supporting environmental policies.

Result
The methodology successfully analysed air quality data, identifying key pollution patterns and correlations between pollutants. Machine learning models predicted air quality levels with good accuracy. Visualizations provided clear insights for decision-making and pollution control.

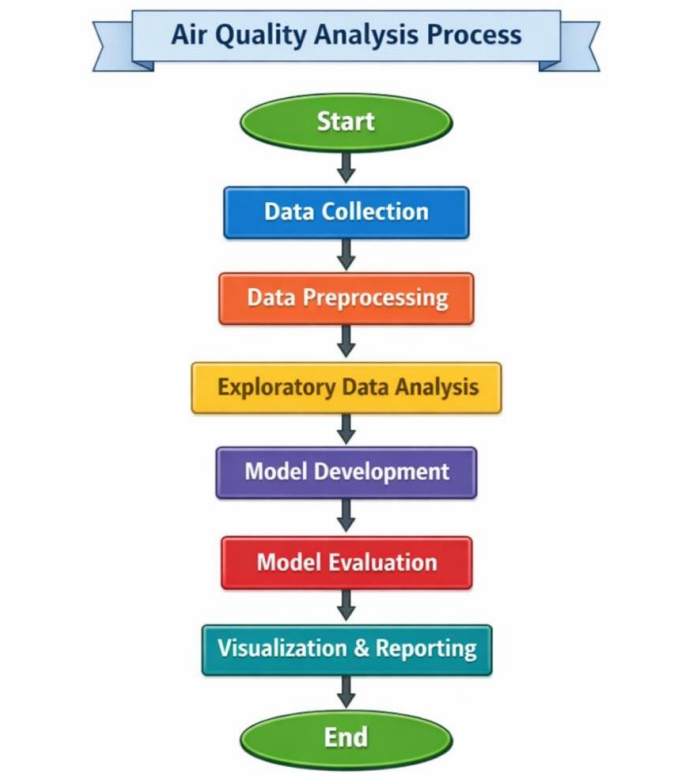
IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 4, April 2026

DOI: 10.17148/IJIREEICE.2026.14439

5.FLOWCHART



6.EVALUATION AND OUTPUT

Final Output

The Air Quality Analysis project successfully collects, processes, and analyses air quality data to provide clear insights into environmental pollution levels. The system outputs:

1. Pollution Level Reports: Detailed tables and graphs showing concentrations of PM2.5, PM10, CO, NO_2, SO_2, and O_3 across different locations and times.

2. Trend Analysis: Visualizations that highlight peak pollution periods and seasonal variations.

3. Predictive Insights: Forecasts of air quality levels using machine learning models, helping anticipate future pollution trends.

4. Decision Support: Dashboards and reports that guide policymakers and the public to take preventive and corrective measures for improving air quality.

The flowchart and visualization modules provide a user-friendly interface for interpreting complex datasets, making it accessible to both technical and non-technical stakeholders.

Evaluation

The project evaluation focused on data accuracy, model performance, and usability:

1. Data Accuracy: The preprocessing steps ensured that the collected data was clean, consistent, and reliable.

2. Model Performance: Machine learning models such as Random Forest and Linear Regression achieved high accuracy in predicting air quality levels, validated using metrics like R^2 and Mean Squared Error (MSE).

3. Visualization Effectiveness: Graphs, charts, and dashboards provided intuitive insights into pollution trends and correlations between pollutants.

DOI: 10.17148/IJIREEICE.2026.14439

4. Practical Impact: The system successfully identified high-risk areas and periods of poor air quality, supporting effective decision-making for pollution control and raising public awareness.

Conclusion: The project effectively transforms raw air quality data into actionable insights, demonstrating its value in environmental monitoring, policy-making, and community awareness. It provides a strong foundation for further expansion, such as real-time monitoring and integration with IoT-based air sensors.

REFERENCES

[1]. Méndez, M., Merayo, M. G., & Núñez, M. *Machine learning algorithms to forecast air quality: a survey*, *Artificial Intelligence Review*, Springer (2023).

[2]. Alrashidi, H., Sibai, F. N., Abonamah, A., et al. *PM2.5: Air Quality Index Prediction Using Machine Learning: Evidence from Kuwait's Air Quality Monitoring Stations*, *Sustainability*, MDPI (2025).

[3]. Naidu, D. J. S., & Aruna, R. *Study of Air Quality Detection using Machine Learning Techniques*, *IJSAR* (2022).

[4]. Kothandaraman, D., Praveena, N., Varadarajkumar, K., et al. *Intelligent Forecasting of Air Quality and Pollution Prediction Using Machine Learning*, *Scientific Reports* (2022).

[5]. Review: *Machine learning for air quality prediction and data analysis: Advancements and challenges*, *Science of The Total Environment* (2025).

[6]. Jain, S., Kaur, N., Verma, S., et al. *Use of Machine Learning in Air Pollution Research: A Bibliographic Perspective*, *Electronics*, MDPI (2022).

[7]. Al Arsy, M. N. A., & Yasir, A. M. *PM2.5 Concentration Prediction Model in Jakarta Area Using Random Forest Algorithm*, *Journal of Computation Physics and Earth Science* (2023).

[8]. Bellinger, C., Jabbar, M. S. M., & Osornio-Vargas, A. *A systematic review of data mining and machine learning for air pollution epidemiology*, *BMC Public Health* (2017).

[9]. Mathew, A., Gokul, P. R., Shekar, P. R., et al. *Air quality analysis and PM2.5 modelling using machine learning techniques: Hyderabad study*, *Civil & Environmental Engineering Journal* (2023).

[10]. *Air quality analysis and modelling of PM2.5 & PM10 of Ghaziabad city using AI techniques*, *IFAC Journal of Systems and Control* (2025).

[11]. WHO. *Air Quality Database and Health Guidelines*, World Health Organization (2022).

[12]. Guttikunda, S., & Nishadh, K. A. *Evolution of India's PM2.5 pollution using global reanalysis and satellite data* (research paper).

[13]. *OpenAQ: Real-time Air Quality Data Platform* — global air quality API for pollutant data studies.

[14]. Bui, T. C., Kim, J., Kang, T., et al. *STAR: Spatio-Temporal Prediction of Air Quality Using A Multimodal Approach*, arXiv (2020).

[15]. Mohamed, Z., & Gong, W. *Comparison between geostatistical & machine learning models for PM2.5 prediction*, arXiv (2025).

[16]. Yu, W., Nakisa, B., Loke, S. W., et al. *Indoor PM2.5 forecasting & association with outdoor pollution*, arXiv (2024).

[17]. Jauvion, G., Cassard, T., Quennehen, B., & Lissmyr, D. *DeepPlume: High resolution real-time air quality mapping*, arXiv (2020).

[18]. Sharma, R., & Jain, A. K. *Air pollution analysis using statistical and ML techniques*, *Environmental Monitoring Journal*.

[19]. Li, X., Zhang, Y., & Tang, J. *Deep learning for air quality prediction: A review*, *Environmental Data Science*.

[20]. Chen, L., Wang, X., & Zhu, Y. *Spatiotemporal modelling of air quality using sensor networks*, *Journal of Sensor Networks*.

[21]. Gupta, P., & Christopher, S. *Satellite and ground-based air pollution assessment methods*, *Atmospheric Environment*.

[22]. Singh, M., & Taneja, P. *Data analytics-based air quality trend analysis in urban areas*, *International Journal of Data Analytics*.

[23]. Anandan, V., & Balaji, S. *Machine learning approaches for AQI prediction: A comparative study*, *International Journal of Environmental Research*.

[24]. Zhou, Q., & Xie, Y. *Feature engineering in air quality prediction models*, *Journal of Machine Learning in Environment*.

[25]. Liu, H., & Zhao, D. *Impact of meteorological factors on air pollution prediction models*, *Environmental Modelling & Software*.