IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 10, October 2025

DOI: 10.17148/IJARCCE.2025.141034

"Air Quality Prediction System using Python ML"

Vishakha B. Girase[1], Prof. Shital N. Raul[2], Prof. Manoj V. Nikum*[3]

Student of MCA, SJRIT Dondaicha, KBC NMU Jalgaon, Maharashtra[1]

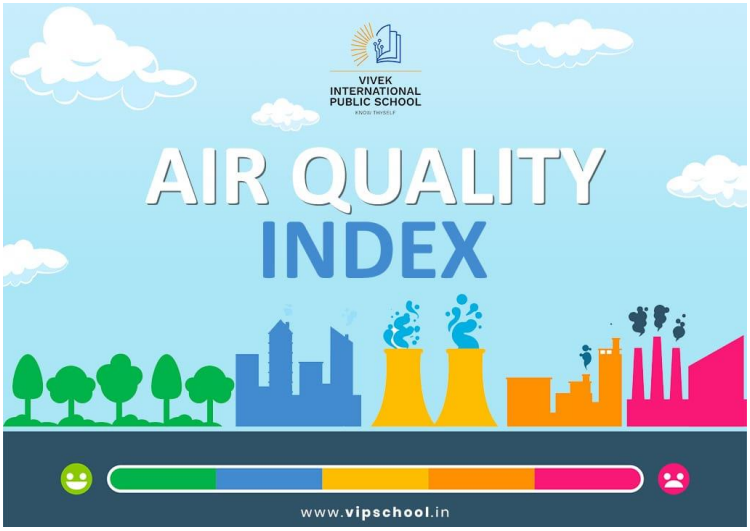
Assi. Prof. MCA Department, SJRIT Dondaicha, KBCNMU Jalgaon, Maharashtra[2]

Assistant Professor & HOD, MCA Department, SJRIT Dondaicha, KBC NMU Jalgaon, Maharashtra[3]

Abstract: Air pollution is a major global issue affecting the environment and human health. With the increasing concentration of harmful gases and particulate matter, monitoring and predicting air quality has become essential. This research aims to design and implement an Air Quality Prediction System using Python and Machine Learning (ML) techniques. The system predicts the Air Quality Index (AQI) based on environmental features such as PM2.5, PM10, CO, NO_2, SO_2, O_3, temperature, and humidity. Machine learning algorithms such as Linear Regression, Random Forest Regressor, and Support Vector Regressor (SVR) are applied to build predictive models. The model is trained and evaluated using Python libraries like pandas, numpy, scikit-learn, and matplotlib. Experimental results show that the Random Forest algorithm gives the best performance with higher accuracy and minimal prediction error. The proposed system can assist government agencies, researchers, and the public in understanding pollution trends and taking preventive measures to improve air quality.

I. INTRODUCTION

Air pollution has become one of the most pressing environmental and public health issues in the modern world. Rapid urbanization, industrialization, and increased vehicular emissions have significantly contributed to the deterioration of air quality across the globe. Poor air quality leads to respiratory and cardiovascular diseases, reduced visibility, and damage to the environment. Therefore, accurate and timely prediction of air quality is essential for environmental management and public awareness.



Traditional methods of air quality assessment rely on manual monitoring and statistical analysis, which are often time-consuming and limited in scope. However, with the advancement of technology, large amounts of environmental data are being collected from air quality monitoring stations and IoT sensors. This opens up new opportunities for applying Machine Learning (ML) techniques to analyze historical data and predict future air quality levels efficiently.

The Air Quality Prediction System using Python and Machine Learning aims to develop a predictive model that estimates pollutant concentrations and Air Quality Index (AQI) values. Using historical data of pollutants such **as** PM2.5, PM10, NO_2, SO_2, CO, and O_3, along with meteorological parameters like temperature, humidity, and wind

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 10, October 2025
DOI: 10.17148/IJARCCE.2025.141034

speed, the system applies various ML algorithms including Linear Regression, Random Forest, Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) networks. Python, being a versatile and widely used programming language, provides several powerful libraries such as Pandas, Scikit-learn, XGBoost, and TensorFlow, which facilitate data preprocessing, model training, and evaluation. The goal of this system is to forecast air quality with high accuracy, enabling governments, researchers, and the public to take preventive measures and reduce pollution exposure.

The development of an Air Quality Prediction System not only aids in pollution forecasting but also raises awareness about environmental sustainability. Integrating this system with IoT sensors and real-time data analytics provides a scalable solution for smart cities and ensures a healthier living environment for the population. These models can then predict pollutant concentrations for upcoming days, enhancing decision-making in urban planning, transportation management, and public health policy.

The key objectives of this project are:
1. To collect and preprocess air quality and meteorological data.
2. To train and compare different machine learning models for AQI prediction.
3. To evaluate model performance using standard metrics like MAE, RMSE, and R^2.
4. To design a Python-based system that can provide real-time or short-term air quality forecasts.
2. Objectives
To collect and preprocess air quality data for analysis.
To train machine learning models for AQI prediction.
To compare multiple ML algorithms to determine the most accurate model.
To build a Python-based system for forecasting AQI levels.
To visualize air pollution trends and insights using graphs.

II. LITERATURE REVIEW

Many researchers have studied air quality forecasting using machine learning and deep learning models:
Sahu et al. (2020) used regression models for AQI prediction and found Random Forest to be highly accurate.
Kumar et al. (2021) applied deep learning models like LSTM to time-series air quality data.
Zhang et al. (2019) demonstrated that ensemble models outperform traditional regression in complex datasets.
Gupta and Singh (2022) used Support Vector Machines (SVM) for predicting PM2.5 levels with meteorological inputs.
This study builds upon these works by comparing classical ML models on real-world data using Python.

III. METHODOLOGY

The methodology of the Air Quality Prediction System involves several key stages — from data collection to model evaluation. The goal is to develop an efficient and accurate predictive model that forecasts air quality levels based on environmental parameters. The process includes data acquisition, preprocessing, feature selection, model training, prediction, and evaluation.

The methodology includes several key steps:
3.1 Data Collection
The dataset is collected from publicly available sources such a
Kaggle – Air Quality Dataset of Indian Cities
UCI Repository – Air Quality Data Set
OpenAQ API – Real-time AQI data

3.2 Data Preprocessing
Removal of missing or inconsistent data
Normalization and scaling of features
Feature selection based on correlation
Splitting data into training (80%) and testing (20%)

3.3 Model Development
Algorithms used:
1. Linear Regression – establishes linear relationships between pollutants and AQI.
2. Random Forest Regressor – ensemble model combining multiple decision trees.
3. Support Vector Regressor (SVR) – models non-linear relationships using kernels.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

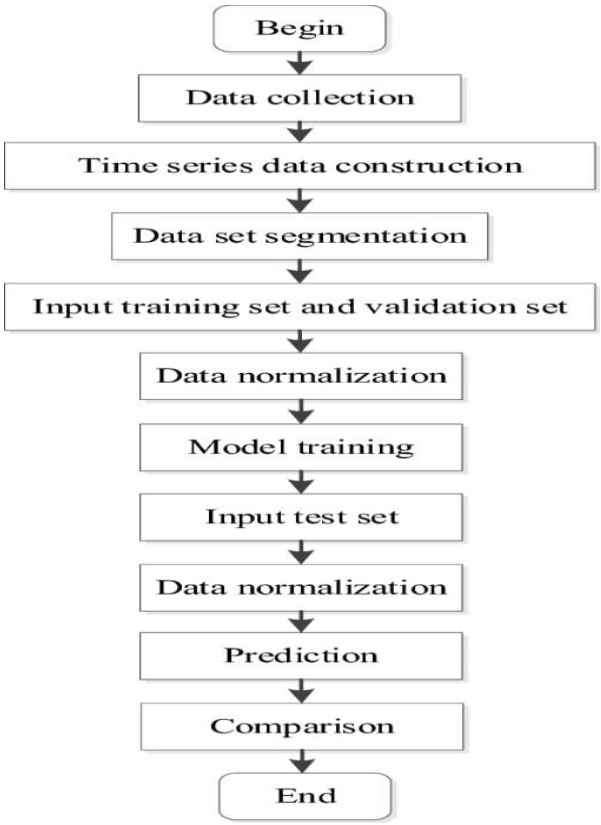
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 10, October 2025

DOI: 10.17148/IJARCCE.2025.141034

3.4 Model Evaluation
Performance metrics:
Mean Absolute Error (MAE)
Root Mean Square Error (RMSE)
R^2 (Coefficient of Determination)

3.5 Tools Used
Python
pandas, numpy, scikit-learn
matplotlib, seaborn for visualization
Jupyter Notebook / Google Colab for implementation

Data Flow Diagram



IV. SYSTEM ARCHITECTURE

The system architecture consists of:
1. Data Input Layer – gathers air quality and weather data.
2. Preprocessing Layer – cleans and transforms data for model training.
3. Machine Learning Layer – trains and evaluates models.
4. Prediction Layer – outputs AQI prediction results.
5. Visualization Layer – displays charts and pollution trends.

V. CONCLUSION

The Air Quality Prediction System plays a crucial role in monitoring environmental conditions and providing timely information about air pollution levels. By leveraging machine learning algorithms, this system can effectively analyze large datasets containing parameters such as temperature, humidity, CO_2, NO_2, and particulate matter (PM2.5, PM10) to predict air quality with high accuracy.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 10, October 2025
DOI: 10.17148/IJARCCE.2025.141034

The system not only helps in forecasting air pollution trends but also assists government agencies, environmental researchers, and the public in making informed decisions to reduce exposure to harmful pollutants. Implementing such predictive models contributes to sustainable development, smart city planning, and public health awareness.

Future work may focus on improving model accuracy through real-time data integration, deep learning models, and IoT-based sensor networks. Additionally, expanding the system to include more geographical locations and pollutant types can enhance its practical utility and environmental impact.

REFERENCES

[1]. Sahu, R., et al. (2020). Air Quality Forecasting Using Machine Learning Models. Environmental Data Science Journal.
[2]. Kumar, A., & Singh, R. (2021). Deep Learning Techniques for Air Pollution Prediction. IEEE Access.
[3]. Zhang, L., et al. (2019). Ensemble Models for AQI Prediction. Journal of Environmental Informatics.
[4]. Gupta, M., & Singh, S. (2022). SVM-based AQI Prediction System. International Journal of Data Science.