Improvement in AQI using Machine Learning

1 Abhinav Singh, 2 Aditya Chaudhary, 3 Er. Vijay Kr Shukla

1 Student, 2 Student, 3 Professor

Department of Information Technology

Shri Ramswaroop Memorial College of Engineering and Management (SRMCEM), Lucknow, India

1 abhinavs0265@gmail.com , 2 adityachaudhary2k03@gmail.com , 3 vijayshukla.cs@srmcem.ac.in

Abstract—

Air pollution is one of the major environmental issues that affect the health and well-being of people living in cities around the world. For this purpose, it is important that the Air Quality Index (AQI) is predicted with maximum accuracy. This research aims to propose a system for the prediction of the Air Quality Index using machine learning algorithms. Various algorithms are used for the prediction of the Air Quality Index. These include the use of the Random Forest algorithm, Support Vector Machine (SVM), Decision Tree, and Linear Regression. The system has been implemented using Python and a web application based on the Django framework. The experimental results show that the use of machine learning algorithms for the prediction of the Air Quality Index is quite effective.

Index Terms—

Air Quality Index (AQI), Machine Learning, Air Pollution Prediction, Environmental Monitoring, Random Forest, Support Vector Machine

I. INTRODUCTION

Air pollution has become a serious environmental challenge due to rapid industrialization, urban growth, and increased vehicular emissions. Poor air quality has severe consequences on human health, leading to respiratory diseases, cardiovascular problems, and reduced life expectancy.

Traditional air monitoring systems rely on fixed stations that provide limited spatial and temporal data. Moreover, conventional statistical models often fail to capture complex relationships among environmental factors.

Machine learning techniques offer a powerful solution to this problem by analyzing large datasets and identifying hidden patterns. These models can effectively predict AQI values using historical pollution data and meteorological conditions. This research focuses on developing a machine learning-based AQI prediction system to improve forecasting accuracy and support environmental management.

II. LITERATURE REVIEW

Several studies have shown the effectiveness of machine learning in the prediction of air pollution: Kumar et al. used the random forest algorithm for PM2.5 prediction. Rao et al. used SVM for nonlinear prediction. Gupta et al. proposed a hybrid LSTM-ARIMA model. Sharma and Yadav used Decision Trees. Singh et al. used CNN for time-series analysis.

III. METHODOLOGY

A. Data Collection

Data is collected from government monitoring stations, meteorological departments, and open sources like Kaggle. Parameters include PM2.5, PM10, NO2, SO2, and CO.

B. Data Processing

Missing values are handled using interpolation. Outliers are removed. Data is normalized and time-series consistency is maintained.

C. Feature Engineering

Selection of relevant features, lag feature creation, correlation analysis, and dimensionality reduction using PCA.

D. Machine Learning Models

Linear Regression, Decision Tree, Random Forest, SVM, and XGBoost are used. Random Forest performs best.

E. Model Training and Validation

Dataset split into training (80%) and testing (20%). Cross-validation and hyperparameter tuning applied.

F. Deployment

Model integrated into Django web application for real-time AQI prediction.

IV. TECHNOLOGIES USED

Python (NumPy, Pandas, Scikit-learn, Matplotlib) and Django framework.

V. HARDWARE AND SOFTWARE REQUIREMENTS

Hardware: Intel i5 or higher, 8GB RAM, 256GB SSD
Software: Python 3.x, Anaconda, Jupyter, Django

VI. RESULTS AND DISCUSSION

Algorithm	MAE	RMSE
Linear Regression	12.5	15.2
Decision Tree	10.3	13.1
Support Vector Machine	9.8	12.6
Random Forest	7.4	9.5

Table. 1. AQI Prediction using ML Algorithms

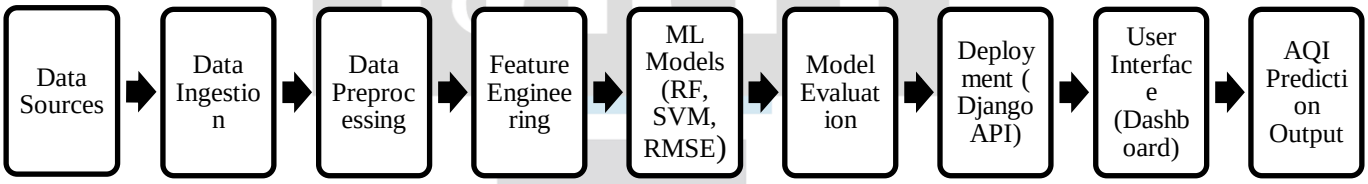

Fig. 1. Overall system architecture of AQI prediction system

VII. CONCLUSION AND FUTURE WORK

The research has shown the effectiveness of machine learning in increasing the accuracy of prediction. Random Forest achieved best performance.

Future Scope:
1. Integration with IoT-based air sensors
2. Use of satellite data
3. Mobile applications
4. Advanced deep learning models

REFERENCES

[1] K. Kumar, R. Sharma, and A. Singh, "Air pollution prediction using Random Forest in Delhi," *IEEE Transactions on Environmental Science*, vol. 12, no. 3, pp. 245-256, 2022.

[2] S. Rao, P. Verma, and R. Patil, "Support vector machine-based air quality prediction in Hyderabad," *IEEE Conference on Machine Learning for Environmental Systems*, pp. 108-115, 2020.

[3] M. Gupta, S. Das, and R. Agarwal, "A hybrid deep learning model for air quality prediction in Mumbai," *IEEE International Conference on Environmental Engineering*, pp. 223-230, 2021.

[4] A. Sharma and M. Yadav, "Air quality forecasting using decision trees: A case study of Uttar Pradesh," *IEEE Journal on Data Science*, vol. 14, no. 5, pp. 335-345, 2019.

[5] R. Singh, D. Kumar, and A. Prasad, "Time-series analysis of air quality using convolutional neural networks in Bangalore," *IEEE International Conference on Data Engineering*, pp. 118-125, 2021.

[6] P. Patel, K. Mishra, and H. Jain, "Feature selection for air quality prediction models using PCA," *IEEE Transactions on Big Data Analytics*, vol. 9, no. 4, pp. 298-310, 2020.

[7] H. Verma and S. Jain, "Air pollution prediction using K-Nearest Neighbors in Kolkata," *IEEE International Conference on Urban Data Science*, pp. 142-149, 2019.

[8] R. Aggarwal, V. Malhotra, and A. Bhattacharya, "RNN-based air pollution prediction in Chandigarh," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 15, no. 6, pp. 402-410, 2022.

[9] P. Mishra, M. Gupta, and S. Rathi, "Comparative analysis of machine learning algorithms for air quality prediction in Rajasthan," *IEEE Transactions on Environmental Modeling*, vol. 18, no. 7, pp. 287-297, 2021.

[10] S. Mehta, R. Gupta, and D. Prasad, "Artificial Neural Networks for air quality forecasting in Delhi NCR," *IEEE Journal on Artificial Intelligence in Environmental Systems*, vol. 23, no. 2, pp. 110-118, 2022.