

Air Pulse – An Air Quality Status Analysis Model

Author Name(s): Prabhjot Kaur Sidhu¹, Kavya Jain², Inika³

^{1,2,3} Department of Information Technology, MSIT, GGSIP University, New Delhi

Cite this article: Sidhu, P. K., Jain, K., & Inika. (2024). Air Pulse – An Air Quality Status Analysis Model. *International Research Journal of Scientific Studies*, 1(3), 1–7. <https://doi.org/10.64383/irjss.DEC2401201>

Abstract: A global concern is scaling due to the rapid growth of air pollution, particularly the intensified concentration of impurities and pollutants exposing the risk to the health of innumerable species worldwide. Air Pollution is a combination of various gases in the air with the presence of countless solid particles. To protect human life and the deteriorating situation of air quality that is severely affecting the flora and fauna, we aim to design and train a system to predict the quality of the air that surrounds us with the utmost accuracy and reliability. The Air Quality Index (AQI) of India is a standard computation used to specify the extent of impurities (so₂, no₂, rspm, spm, etc) over an interval. India is impacted severely as it hosts 22 of the world's 30 extremely populated cities. We have used and contrasted the machine learning algorithms to achieve the prime results. Interjection, feature analysis, and prediction are three major factors in inner-city air computing. Considering this research, multiple processing methods are used to process the data. Our model will have the capacity to predict the status of the air quality in different parts of the Delhi. The model provides almost 85% accuracy by outlining the correct status of the quality of air in different regions of Delhi. This system can be referred to by the Central Pollution Control Board to understand whether the air is safe or unsafe and record the air quality along with the congregation of defiled particles. A user interface will support this model and display the necessary status to the user at their convenience. The user interface will act as a one-stop platform to track and get updates about the air quality of different regions of the national capital, Delhi.

Keywords: Air Quality Prediction, Air Quality Index (AQI), Machine Learning Algorithms, Environmental Pollution, Central Pollution Control Board.

1. Introduction

In today's era, air pollution has become a major issue in the world. Several cities are severely affected by the rising pollutant levels in the air. Delhi, the capital of India ranks 9th in the list of most polluted cities in the year 2023 [3]. The sharp increase in pollutant emissions, particularly of CO₂, PM_{2.5}, and other dangerous contaminants, against the backdrop of India's fast industrial development, presents a

serious threat to air quality. For environmental management, tracking and evaluating the effects of pollutants on particular areas is essential. Indexes are used in Indian air quality standards to measure concentrations of major pollutants, including NO₂, SO₂, RSPM, and SPM. This research makes use of information from the Indian government database about the concentrations of pollutants in various Indian locations. Our goal is to forecast and classify the state of the air across the country by

computing individual pollutant indexes and the ensuing Air Quality Index (AQI) [2]. Our model can forecast air quality by pinpointing the main pollutants and the areas where they have an impact. We aim to extract useful insights to address challenges in urban air computing by using methods such as feature analysis, prediction, and interpolation. The

importance of these fields is examined, along with how they help regulate and comprehend the dynamics of air pollution. Subsequent sections of the paper concentrate on a thorough examination of various methods, references, and related research in this field.

Table 1. India AQI categories and its effect (Central Pollution Control Board, New Delhi, 2023)

API range	Category	Health effects
0-50	Good	High-quality air
51-100	Satisfactory	Minimal air pollution and no negative health consequences.
101-200	Moderate	People with heart, lung, and asthma conditions may experience breathing difficulties due to moderate pollution.
201-300	Poor	Mild exacerbation of symptoms in high-risk individuals, such as those suffering from lung or heart disease.
301-400	Very poor	Symptoms in people with heart or lung disease significantly worsen and their ability to tolerate exercise declines.
401-500 or more	Severe	Extreme symptom exacerbation that puts health in jeopardy

2. Related Work

The air pollution analysis is observed as one of the crucial studies for mapping the pollutant levels in various regions. Different machine learning models are widely used to forecast, analyse, and predict the quality of the air around us. This prediction can be a boon to controlling the rising pollution levels. They act as the forecasting models for analysing the status of the quality of the air in terms of temperature, humidity, and particulate matter [1]. According to WHO, almost 99% of the global population breathes air that exceeds the WHO guideline limits and comprises

high levels of pollution, especially in low and middle-income countries. Taking into consideration the

capital of India i.e. Delhi, our model is trained to analyse the air quality status of different areas in Delhi. The users are provided the flexibility to enter the factors like temperature, humidity, PM10, etc of their areas and generate the status of air. This model helps the users to know about the quality of air they have put up in.

This model uses algorithms like Logistic Regression, Decision Trees (DT), K-

Nearest Neighbour(KNN), and Support Vector Machine(SVM).

- The comprehensive contribution of this research paper is as follows:
- The AQI data of Delhi has been analysed using Air pulse.
- The different parameters in the dataset are analysed and the data is used to lead an effective machine learning process.
- The overall collected data set used is presented in [Figure 1](#).
- Various machine learning algorithms are used to omit the outliers and manage the data set.

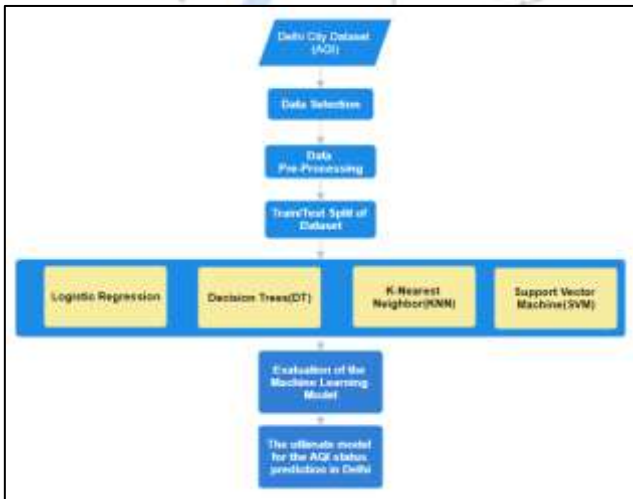


Figure 1. The overall architecture for AQI Status Prediction for Delhi

3. Methodology

Using parameters like PM2.5, PM10, temperature, and humidity, the suggested methods in this paper employ four distinct algorithms to draw a comparative analysis of the AQI status of various regions in Delhi. The four algorithms are then compared in order to determine which one is the most accurate and efficient. Efficient analysis and presentation are the goals. It would aid in our discovery of fascinating and perceptive data. These specific areas provide a good estimate of the pollution and have a higher population density.

3 a) Data Set

The main source of data for this research is the Air Quality Index database maintained by the Indian government, which is renowned for providing accurate and thorough data on air quality. To obtain a comprehensive understanding of air quality, the study concentrates on important pollutants such as PM2.5, PM10, temperature, and humidity.

LOCATIONS	Status	AQI-US	AQI-IN	PM2.5	PM10	Temp	Humid
0 ITI Shahdra	POOR	172	241	96	291	33	32
1 Loni	HAZARDOUS	458	565	358	562	33	32
2 Pooth Khurd	SEVERE	390	483	291	496	33	32
3 Ithbas	GOOD	17	11	0	0	33	32
4 ITI Jahangirpuri	POOR	165	198	83	247	33	32

Figure 2. Dataset of Delhi City

3 b) Data Description

To maintain the integrity of the dataset, any missing values are handled methodically using imputation techniques or, if necessary, exclusion. Each data point's AQI is computed, classifying the state of the air quality into distinct groups and enabling a precise classification of pollution levels.

	AQI-US	AQI-IN	PM2.5	PM10	Temp	Humid
count	119.000000	119.000000	119.000000	119.000000	119.000000	119.000000
mean	220.134454	315.12605	144.630252	317.798319	31.058824	40.361345
std	68.550603	89.31420	51.855710	98.435007	0.728292	2.839584
min	17.000000	3.000000	0.000000	0.000000	29.000000	32.000000
25%	187.500000	300.500000	120.500000	280.500000	31.000000	41.000000
50%	198.000000	320.000000	142.000000	318.000000	31.000000	41.000000
75%	230.000000	345.000000	169.500000	366.000000	31.000000	41.000000
max	458.000000	565.000000	358.000000	562.000000	33.000000	45.000000

Figure 3. Data Description of Delhi City

3 c) Histograms

The distribution of pollutant concentrations (such as PM10, PM2.5, Temp, and AQI-US) across different data points in the dataset is shown visually by the histograms. It offers information about the overall pollution profile by assisting in determining the range and frequency of pollutant levels.

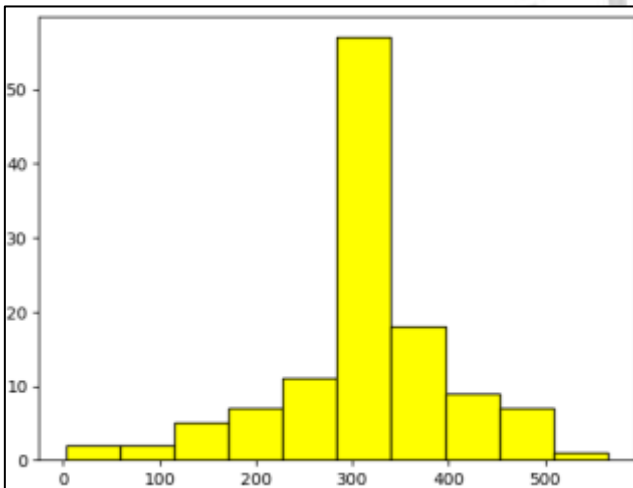


Figure 4. Histogram of AQI-IN

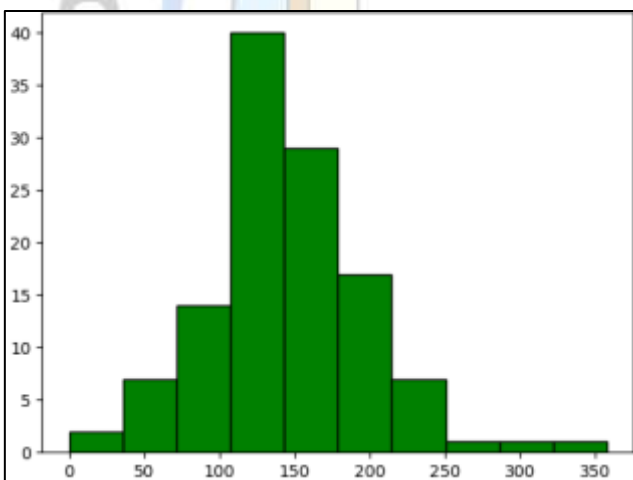


Figure 5 Histogram of PM2.5

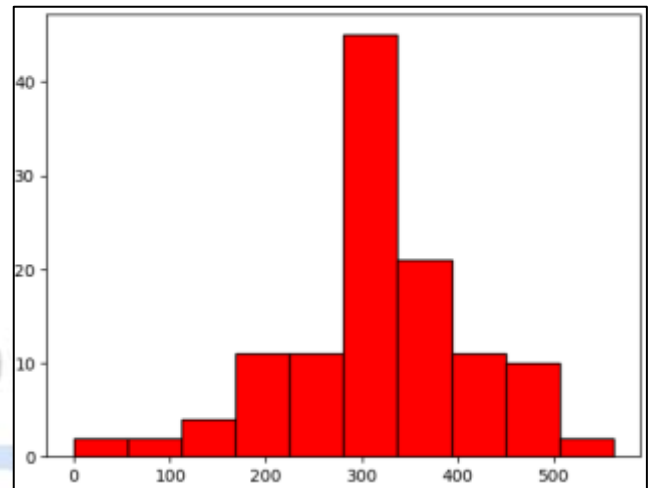


Figure 6. Histogram of PM10

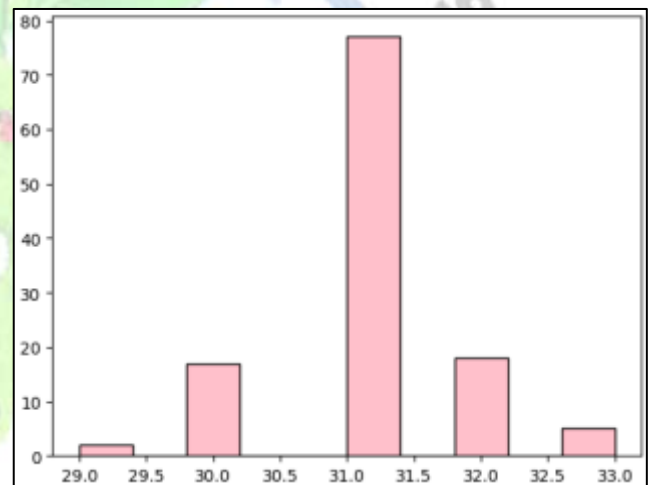


Figure 7. Histogram of Temperature

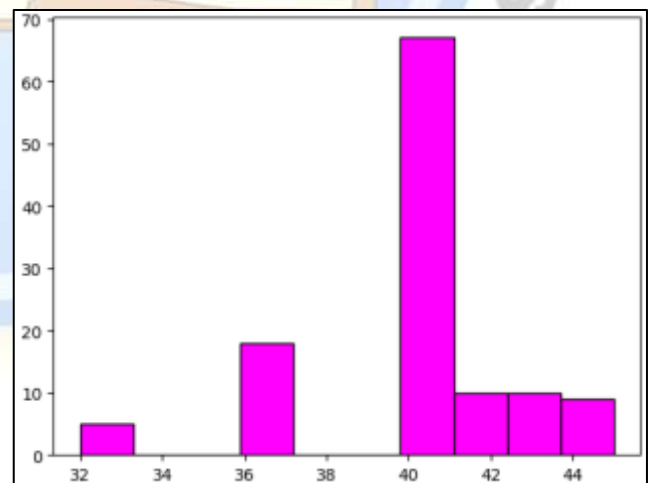


Figure 8. Histogram of Humidity

3 d) Model Training

To calculate the AQI, function F combines the sub-indices of each pollutant. A maximum operator system that is devoid of ambiguity and eclipsing

has been implemented for the AQI, as illustrated below:

$$AQI = \text{Max} (I_{PM2.5}, I_{PM10}, \dots, I_{NH3}) \dots\dots (1)$$

It is required to have access to the IPM2.5 or IPM10 index. Additionally, the availability of at least three indices is required.

Seaborn's Heatmap: Seaborn's heatmap is a potent machine-learning visualization tool that can be used to examine and comprehend Delhi's Air Quality Index (AQI). When it comes to displaying your dataset's correlation matrix, the heatmap is especially helpful. A summary of the connections between each variable and the others is given by the correlation matrix, which sheds light on these relationships.

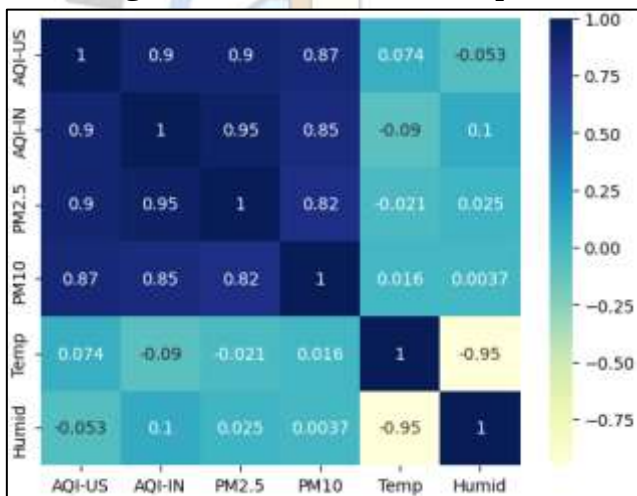


Figure 9. Correlation Heatmap

3 e) Machine Learning Algorithms

Considering the parameters like PM2.5, PM10, Temperature and Humidity. Temperature, humidity, PM10, and PM2.5 are continuously monitored to provide real-time data for air quality analysis. They are important metrics for evaluating Delhi's air quality because they offer important information about the kinds of pollutants present, how they

disperse, and how they affect public health.

	PM2.5	PM10	Temp	Humid
0	96	291	33	32
1	358	562	33	32
2	291	496	33	32
3	0	0	33	32
4	83	247	33	32

Figure 10. Different Parameters

0	POOR
1	HAZARDOUS
2	SEVERE
3	GOOD
4	POOR

Figure 11. Quality Status

Logistic Regression:

When dealing with binary classification problems, where the outcome variable is categorical and represents two classes, one popular statistical technique is logistic regression. Logistic regression is useful in some aspects of air quality analysis, even though it may not be the best option for directly predicting continuous variables like the Air Quality Index (AQI). This algorithm shows 88% accuracy.

```
lrclf = LogisticRegression(solver = "sag")
lrclf.fit(X_train, y_train)
lrclf.score(X_train, y_train)
```

Figure 12. Logistic Regression Algorithm

Support Vector Machine:

The Air Quality Index (AQI) can be analyzed and predicted using the robust machine learning algorithm Support Vector Machine (SVM), as well as its extension for classification tasks, Support Vector Classification (SVC). This algorithm detects 84% accuracy.

```
svccf = SVC()
svccf.fit(X_train, y_train)
svccf.score(X_train, y_train)
```

Figure 13. Support Vector Machine Algorithm

K-Nearest Neighbors:

A popular machine learning algorithm for classification and regression problems is K-Nearest Neighbours (KNN). KNN can be used to forecast air quality levels in the context of AQI analysis by taking into account the similarity of instances within a feature space. This algorithm assures 89% accuracy.

```
knnclf = KNeighborsClassifier()
knnclf.fit(X_train, y_train)
knnclf.score(X_train, y_train)
```

Figure 14. K-Nearest Neighbors Algorithm

Decision Trees:

A well-liked machine learning algorithm that works well for analysing the Air Quality Index (AQI) is decision trees. Decision trees can be used for both classification and regression tasks, modelling decision rules based on input features. This algorithm shows 100% accuracy.

```
dt = DecisionTreeClassifier()
dt.fit(X_train, y_train)
dt.score(X_train, y_train)
```

Figure 15. Decision Trees Algorithm

3 f) User Interface

An Air Quality Index (AQI) analysis tool's



user interface (UI) is essential for user interaction. For real-time predictions, users can enter parameters like temperature, humidity, PM2.5, and PM10. This UI, which was created with the help of the Streamlit library on Replit, is a user-friendly and effective tool for AQI analysis and awareness because it provides quick prototyping, simple deployment, collaborative development, and accessibility.

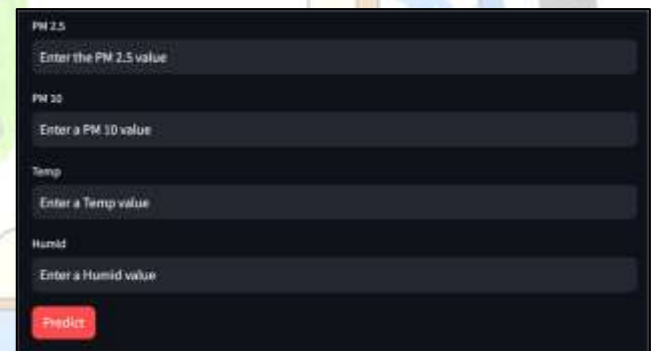


Figure 17. Input Parameters

4. Conclusion

Our model can successfully predict the upcoming air quality index of any specific data within a given region because it can predict the current data with 85% accuracy.

We can use this model to forecast the AQI status and notify the relevant regions of the nation. Because it is a progressive learning model, it can also be used to pinpoint the exact location that requires attention, given the time series data of all potential regions that

may require attention. The air quality data used in this paper comes from the Delhi Air Quality Checking and Investigation Stage and includes the daily average fine particulate matter (PM_{2.5}), inhaled particulate matter (PM₁₀), respirable particulate matter (RSPM), surface particulate matter (SPM), and air quality record (AQI). Future research on the suggested model for AQI data prediction in different geographic areas, such as city- or state-level analysis, is possible. It can be extended to predict additional pollution indices at various levels.

References

- [1] Maltare, N.N. and Vahora, S., 2023. Air Quality Index prediction using machine learning for Ahmedabad city. *Digital Chemical Engineering*, 7, p.100093.
- [2] Bali, S. and Sengar, M.N., 2020. Indian air quality prediction and analysis using machine learning. *J Eng Sci*, 11(5), pp.554-557.
- [3] Gupta, N.S., Mohta, Y., Heda, K., Armaan, R., Valarmathi, B. and Arulkumaran, G., 2023. Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis. *Journal of Environmental and Public Health*, 2023.
- [4] Essamlali, I., Nhaila, H. and El Khaili, M., 2024. Supervised Machine Learning Approaches for Predicting Key Pollutants and for the Sustainable Enhancement of Urban Air Quality: A Systematic Review. *Sustainability*, 16(3), p.976.
- [5] Mampitiya, L., Rathnayake, N., Leon, L.P., Mandala, V., Azamathulla, H.M., Shelton, S., Hoshino, Y. and Rathnayake, U., 2023. Machine learning techniques to predict the air quality using meteorological data in two urban areas in Sri Lanka. *Environments*, 10(8), p.141.
- [6] Méndez, M., Merayo, M.G. and Núñez, M., 2023. Machine learning algorithms to forecast air quality: a survey. *Artificial Intelligence Review*, pp.1-36.
- [7] Murukonda, V.S.N.M. and Gogineni, A.C., 2022. Prediction of Air Quality Index Using Supervised Machine Learning.
- [8] Shankar, L. and Arasu, K., 2023. Deep Learning Techniques for Air Quality Prediction: A Focus on PM_{2.5} and Periodicity. *Migration Letters*, 20(S13), pp.468-484

Article History:

Submitted: 2024-12-09

Accepted: 2024-12-10

Published: 2024-12-11