

# Air Quality Index detection using Machine Learning Approach

Raghav Mundra<sup>1</sup>, Rahul Singh<sup>2</sup>, Poojith R.Pejawar<sup>3</sup>, Mohammed Shanin<sup>4</sup>, Keerthi P<sup>5</sup>

<sup>12345</sup> Department of Computer Science, Atria Institute of Technology, Bangalore

**Abstract:** The emission of contaminants is referred to as air pollution. It harms forests, crops, animals, and other things. To avoid this issue, air quality must be predicted from pollutants and preventive measures should be taken. By foreseeing outcomes with the highest degree of accuracy, the goal is to investigate machine learning-based air quality forecasting solutions. The model offers a Model parameter sensitivity analysis manual with performance in terms of accuracy. The accuracy of the Support Vector Machine model is the lowest, while that of The maximum is the Gaussian Naive Bayes model. Using established performance characteristics, these models' performances are compared and evaluated. The XGBoost model surpassed the competition and achieved the strongest linearity between the predicted data and the actual data.

**Keywords:** Random Forest Classifier

## I. INTRODUCTION:

Air quality is affected by different forms of pollution caused by traffic, electricity, fuel, etc. It also contributes to ozone depletion. Smog, eutrophication, and global climate change are some other environmental repercussions of air pollution. In many industrial and urban locations today, preserving and monitoring the quality of the air has become one of the most important tasks. The model helps to in of identifying the pollutants to save humansto anticipate if the air quality is terrible or good based on given and attributes, saving a lot of meteorological time and resources. Since roughly 20 years ago, the quality of the air in New Delhi has been monitored. This has made it possible to comprehend how air pollution fluctuates in response to certain activities and governmental laws, yet there are still issues with air pollution in New Delhi computing air pollution prediction. Air quality is a major environmental concern worldwide, with increasing levels of air pollution leading to negative health impacts. The air quality index (AQI) is a measure of air quality that provides information on the health risks associated with exposure to different pollutants. In recent years, machine learning algorithms have been used to develop models for AQI detection. This paper presents a study on the development of an AQI detection system using machine learning algorithms.

## II. LITERATURE SURVEY

A literature review is a piece of writing that summarises the most important aspects of current understanding and/or methodological approaches to a specific issue. It is secondary sources and discusses material that has been published in a certain subject area, and occasionally information that has been published in a specific subject area within a specific time period. Its main purpose is to introduce the reader to the most recent literature on the topic and form the starting point for another goal, like future study. precedes a research proposal and may only be a brief review of the sources that are needed in the field. It often follows a structure and involves synthesis and summarization.

One of the most crucial jobs in combating urban air pollution is forecasting it. WRFchem, which is a numerical model, has the disadvantage that the source list cannot be updated in time. There is nature. As an input feature. Pollutant Percentage in the Atmosphere Creating an Extensive Evaluation Framework to Boost Predictive Performance. We conducted experiments with different feature sets and machine learning classification algorithms to find the best model. Different sets of feature selection and model selection from studies in various cities produce the best outcomes. Experimental results the prospect of improving accuracy increased with the introduction of more features. From a methodological point of view, the combined model results are superior to the proprietary model. The amount of contaminants in the air has substantially increased recently due to China's economy and society developing so quickly. Public attention was drawn to issues with environmental air quality assessment and pollution control. Governmental organisations utilise the Air Quality Index (AQI) as a measure of air pollution to let the public know how bad it is or might get. An increasing percentage of the populace is anticipated to endure more severe unfavourable health consequences as the AQI rises. The government can use air quality prediction to provide timely information on the changing trends in atmospheric environment quality. The Air A quantitative way to characterise the degree of air pollution is the Quality Index (AQI). An index for describing daily air quality is the daily AQI.

### iii.METHODOLOGY

The study involved the collection of air quality data from various locations using air quality sensors. The collected data was preprocessed to remove missing values and outliers. Feature selection was performed using correlation analysis to identify the most relevant features for AQI prediction.

### iv.IMPLEMENTATION

#### Data Collection:

This marks the beginning of the actual process of building a machine learning model and gathering data. This is a crucial phase since how well the model performs will be influenced by how much more and better data we can collect.

**Dataset:** There are 12456 distinct pieces of data in the collection. The dataset has 12 columns, each of which is detailed below.

Table 1. Schematic diagram of data source format2

	SO <sub>2</sub> ( $\mu\text{g}/\text{m}^3$ )	NO <sub>2</sub> ( $\mu\text{g}/\text{m}^3$ )	PM10 ( $\mu\text{g}/\text{m}^3$ )	PM2.5 ( $\mu\text{g}/\text{m}^3$ )	O <sub>3</sub> ( $\mu\text{g}/\text{m}^3$ )	CO ( $\text{mg}/\text{m}^3$ )
00:00	7	14	32	20	134	0.7
01:00	9	21	33	16	138	0.7
02:00	8	19	33	16	140	0.7
03:00	9	22	40	20	182	0.7
....	....	....	....	....	....	....

#### Model Selection:

We used Random Forest Classifier machine learning algorithm. We got Train Accuracy of 99% so we implemented this algorithm.



### v. CONCLUSION

Data preparation and processing were the first steps in the analytical process. Missing value analysis, exploratory analysis, and model construction and evaluation came next. This tool can help India's meteorological agency forecast the condition of the air and its future trends, enabling appropriate action to be done. In the future, we are applying some more algorithms to train and check the accuracy of the model. We will also implement a GUI-based application where users can provide input and the pollutant class will be displayed to the user.

The study demonstrated the potential of machine learning algorithms for AQI detection. The results showed that the random forest algorithm performed the best, with the decision tree algorithm also showing promising results. The developed AQI detection system can be used to provide accurate information on air quality to policymakers and the public, helping to address the issue of air pollution and its negative impacts on health. Further studies can be conducted to explore the use of deep learning algorithms for AQI detection and to improve the performance of the developed system

**vi. REFERENCES**

- [1] Dragomir, Elia Georgiana. "Air quality index prediction using K-nearest neighbour technique no. 1 (2010): 103-108
- [2] Kumar, Anikender and P. Goyal, " Forecasting of daily air quality index in Delhi", Science of the Total Environment 409, no. 24(2011): 5517-5523.
- [3] Challa Venkata Srinivas et al," Data Environment 60 (2012): 37-50
- [4] Sharma ME A McBean and U.Ghosh, "Prediction of atmospheric sulphate deposition at sensitive receptors in northern India", Atmospheric Environmet