

# Air Quality Prediction Using Machine Learning

Mr M.Dhanasekar<sup>1</sup> Ms S.Sowmiya<sup>2</sup> Ms M.Shahanas<sup>3</sup>

<sup>1</sup>*Assistant Professor, Department of Computer Science, Hindusthan College of Arts & Science, Coimbatore*

<sup>2</sup>*II PG Student, Department of Computer Science, Hindusthan College of Arts & Science, Coimbatore*

<sup>3</sup>*II PG Student, Department of Computer Science, Hindusthan College of Arts & Science, Coimbatore*

**Abstract:** Examining air quality has become one of the most essential activities for the peoples in many industrial and urban areas today. The quality of air is adversely affected due to various forms of pollution caused by electricity, transportation, fuel uses etc. The deposition of harmful gases is creating a serious problem for the quality of life in smart cities. With this increasing air pollution, we need to implement efficient air quality monitoring and prediction models which collect information about the concentration of air pollutants and provide assessment of air pollution in area. Particulate matter contains microscopic solids or liquid droplets that are so small that they can be inhaled and cause serious health problems. Of these, particles less than 2.5 micrometres in diameter, also known as fine particles or PM2.5, pose the greatest risk to health. Hence, air quality evaluation and prediction has become an important research area. The focus of this project is to study about the data mining and machine learning techniques used for prediction of air pollution and the mainly focus is on the prediction of PM2.5 on the basis of all the other air pollutants and temperature and humidity

**Keywords:** Machine learning, Linear aggression, prediction.

## I.INTRODUCTION

In the developing countries like India, the rapid increase in population and economic upswing in cities have lead to environmental problems such as air pollution, water pollution, noise pollution and many more. Air pollution has direct impact on humans health. There has been increased public awareness about the same in our country. Global warming, acid rains, increase in the number of asthma patients are some of the long-term consequences of air pollution. Précised air quality forecasting can reduce the effect of maximal pollution on the humans and biosphere as well.

Hence, enhancing air quality forecasting is one of the prime targets for the society. The main sources of air pollution are burning of fossil fuels, emission of noxious gases and solid substances from vehicles and industries. Such materials are Sulphur oxides, nitrogen dioxides, particulate matter, carbon monoxide. Currently supervising and scrutinizing air quality is a very crucial issue to have a healthful life, and it also very important. By applying data mining techniques air pollution can be analysed, so that apt actions can be taken for reduction of air pollution. Data Mining is simply an approach for extracting intention-based knowledge from the raw data set.

## II. LITRATURESURVEY

A literature review includes the current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic. Literature reviews do not report new or original experimental work. A literature survey, or literature review, means that anyone read and report on what the literature in the field has to say about any topic or subject. There may be a lot of literature on the topic or there may be a little. Either way, the goal is to show that he has read and understand the positions of other academics who have studied the problem issue that he is studying. It allows reader to establish his theoretical framework and methodological focus. Even if he/she is proposing a new theory or a new method, he/she is doing so in relation to what has been done. 1. Ranjana Waman Gore et.al.in have proposed an approach in which Naïve Bayes and J48 classification algorithms are used for analyzing the air quality levels. The accuracy of dataset by using Naïve Bayes was 86.66% and the accuracy with J48 decision tree

algorithm was 91.99%. And author also justify that J48 algorithm gives more accurate results than Naïve Bayes algorithm. 2. Sandhya P in have proposed a method in which author aim is to predict the PM2.5 by using random forest, Naïve Bayes, and decision tree algorithm. 3. Bonny Paulose et.al. in proposed mainly focused on analysis of air quality of Delhi and also find the reason behind the pollutants that cause air pollution by using K-means clustering algorithm. And the author showed that Anand Vihar, R k Puramand, Punjabi Bagh are one of the mostly polluted regions.

### III. MACHINE LEARNING

The Machine Learning Tutorial covers both the fundamentals and more complex ideas of machine learning. Students and professionals in the workforce can benefit from our machine learning tutorial.

A rapidly developing field of technology, machine learning allows computers to automatically learn from previous data. For building mathematical models and making predictions based on historical data or information, machine learning employs a variety of algorithms. It is currently being used for a variety of tasks, including speech recognition, email filtering, auto- tagging on Facebook, a recommender system, and image recognition.

Features of Machine Learning:

- Machine learning uses data to detect various patterns in a given dataset.
- It can learn from past data and improve automatically.
- It is a data-driven technology.
- Machine learning is much similar to data mining as it also deals with the huge amount of the data.

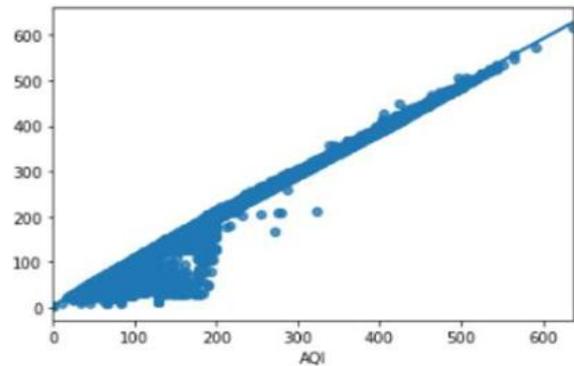
### IV. DATA ANALYSIS AND MODELLING

Its provide details of how the components or modules are integrated and is described with the help of Unified Modelling Diagrams. A system architecture is the conceptual model that defines the structure, behaviour, and more views of a system. An architecture description is a formal description and

representation of a system, organized in a way that supports reasoning about the structures and behaviours of the system. A system architecture can comprise system components, the expand systems developed, that will work together to implement the overall system The goal of this project is to take the publically available weather data from 2013 to 2020 and apply machine learning techniques to see if we can predict the amount of PM2.5 concentration in the air given other environmental features. In this project applying some of the machine learning algorithms (Linear regression). I hope to come up with a predictive model with a high accuracy and a very low Root Mean Square Error (RMSE).

### V. RESULTANDANALYSIS

By applying Linear regression on processed data we plot the below graph



And to classify AQI in different state we use different classification algorithm on previous results.

1. Logistic Regression algorithm 2. Random Forest algorithm

3. Decision Tree algorithm

Table 1 accuracy table for different classifier algorithms

Sr no	Classification algorithm	Accuracy score	Result
1	Logistic Regression	72.78	Low accuracy
2	Random Forest	99.97	High accuracy
3	Decision Tree	99.98	High accuracy

### VI. CONCLUSION

Throughout this project, several models which can predict AQI levels and classify them into different pollution bands were experimented and their

performance was successfully evaluated. The exploratory data analysis and feature engineering methods implemented for the prediction models revealed interesting correlations between weather and pollution data. We obtained several notable outcomes from the predictive models that are worth being discussed. Different approaches to handle null values yielded varied performance from each of the models, however simply dropping the records that had null values seemed to be the best approach. Between obtaining the AQI by predicting the PM2.5 values and using a classifier to predict the AQI band straight away, the classifier seemed to perform better. A regression model could be used for applications in data analytics, but it is concluded that classifier models perform better for air quality prediction.

#### REFERENCE

- [1] Carbajal-Hernández, José Juan "Assessment and prediction of air quality using fuzzy logic and autoregressive models." *Atmospheric Environment* 60 (2012): 37-50.
- [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] Singh Kunwar P., et al. "Linear and nonlinear modelling approaches for urban air quality prediction," *Science of the Total Environment* 426(2012):244-255.
- [4] Sivacoumar R, et al, "Air pollution modelling for an industrial complex and model performance evaluation", *Environmental Pollution* 111.3 (2001) : 471-477.
- [5] Gokhale sharad and Namita Raokhande, "Performance evaluation of air quality models for predicting PM10 and PM2.5 concentrations at urban traffic intersection during winter period", *Science of the total environment* 394.1(2008): 9-24.
- [6] Bhanarkar, A. D., et al, "Assessment of contribution of SO2 and NO2 from different sources in Jamshedpur region, India," *Atmospheric Environment* 39.40(2005):7745-7760.