# Air Quality Prediction Using Machine Learning Algorithm

**A1: YUVARAJA O H**

*M.C.A. Student, Dayananda sagar College of Engineering*
*CO -AUTHOR: Chandrika M*
*Assistant Professor, Department of MCA, Dayananda sagar College of Engineering*

---

## Abstract:

Nowadays pollution is most important problem. There are air pollution, soil pollution, Thermal Pollution, water pollution, light pollution, Radioactive pollution, and noise pollution. Air pollution is most important and it will directly affect to all the living organisms. It can impact kids, animals, individuals and their health, e.g. asthma patients can be affected by air pollution. Air is mixture of many gases and some dust particles. Air will lost all its recourses and it will lead to bad atmosphere and create unhealthy environment so we need to check its quality day to day and we need to predict the air atmosphere

## Introduction

Air is Earth atmosphere which is mixture of many gases and some dust particles. Air is a mixture of oxygen, nitrogen, and carbon dioxide which has the percentage of these mixtures those are 78%, 21%, and 0.06% respectively. Where humans breathe the oxygen from air and leave the carbon dioxide where the trees and plants breathe the carbon dioxide from air and leave oxygen, but plants and trees take oxygen while preparing food which human needs and wants. So we need to check the quality of air day to day by graphical representation to understand where the quality is standardizes the levels of all country and city. The harmful gases are creating major issues and serious threat to the quality of life in smart cities.

## Air Quality Prediction

We all must and should need to implement efficient air quality monitoring models which collects information about the air pollutants and provide assessment of air pollution in each area.

## What is the importance of maintaining quality of air?

It provides air quality standards and key objectives for air pollutants, which are designed to protect human health, animal's lives, and the environment. Air pollution can cause both long term and short term effects on health and lot of people are concerned about pollution in air that they can breathe.

## How does air pollution affect?

When people breathe polluted air pollutants get into our lungs; they can enter the body and be carried to our internal organs like brain, heart. An adult breathes more than 15,000 liters of air every day. Polluted air affect to agricultural productivity, damages ecosystem, and loss of biodiversity. This causes severe health problems such as cardiovascular, asthma, cancer diseases. Pollution reduces the living organism lives. Air

pollution in India causes a serious and most number of health issue. Of the most polluted cities in the world, 21 out of 30 were in India in 2020.

## What is the actual status of air pollution in India?

In India 47% of pollution is caused by the industries, 29% by vehicles, 16% by crop burning and 8% by burning crackers. In India air pollution causes more effect and deaths of 2 million every year. When we consider throughout world death rate India will be increases 2% every year. According to research the total number of deaths is 7million people just by air pollution. Air pollution causes 19% of cardiovascular deaths, 23% of lung cancer deaths, 21% of stroke deaths, 24% of heart disease deaths.

## What to do when the air quality is bad in your locality?

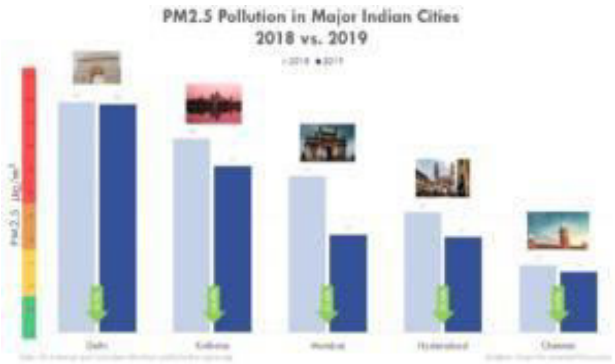
A wide range of pollutants are emitted by the human activity on factories and industries, like CO, $CO_2$, NO, $NH_3$, lead, hydrocarbons, organic compounds and other chemicals.

**Fig1: Air pollution from factories & industries**

Pollution is harmful for our health and clear air is important. Sometimes we can feel powerless when it comes to live in an area with pollution, since we can't hideaway inside with our air purifier forever. Thankfully, a resource called the Air Quality Index (AQI) was created by the US Environmental Protection Agency. It helps to monitor air quality so we can understand the impact it may affect our health.

## How to proceed with the prediction task?

Prediction of air quality not only involves the forecasting or other measures like speed of wind, temperature of wind. But it also includes some other criteria's like pollutant concentrations and emissions from distant locations, movements and




possible transformations of pollutants, prevailing wind.

**Fig2: Air pollution in major Indian cities 2018 v/s 2019**

## What are the main Sources of Air pollution?

Dust&constructions, waste burning, transport, diesel generator, industries, domestic cooking, fuel

burning.

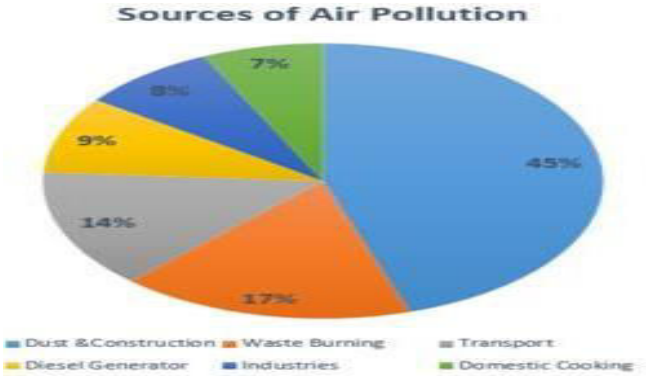



**Fig3: Sources of air pollution in India**

A Smart Air analysis of new air quality data found that pollution improved in 2019 in all major citiesin India. Small particles, dust in the air decreasedfrom 2018 to 2019 by as much as 15.5%.

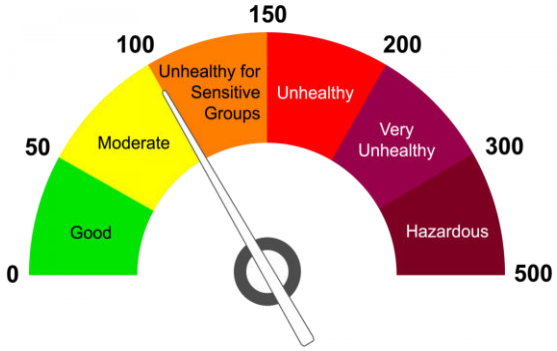



**Fig4: Air quality index, ozone watches/alerts, and health advisories.**

## How does the AQI work?

The AQI reports on what effects breathing the outdoor air may have on your health and how clean or polluted the air in your area is. The AQI forecast is available in India. There are 199 stations and we can view the regional maps that rate the quality of air across the India. Many factors are predicting air quality result that is both subjective and objective in air pollution forecasting. These models are made up of algorithms. The algorithms perform statistical analysis and data mining, determining trends and patterns in data. Predictive analytics software solutions will have built-in algorithms that can be used to make predictive models. Most commonly used predictive models are Decision tree, Regression (linear and logistic).

- **Decision tree**

Decision tree are simple, but powerful form of multiple variable analysis. They produced by algorithms that splits data into branches like segments. Decision trees partition data into subsets based on categories of input variables.

- **Regression (linear and logistic)**

Regression is one of the most popular methods in statistics. Regression analysis estimates relationships among variables, finding key patterns in large and diverse data sets and how they relate to each other.

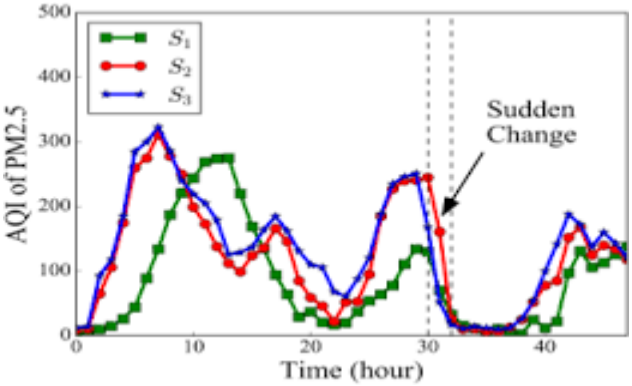
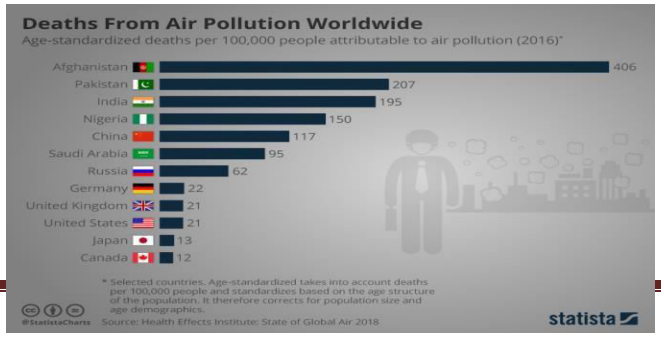



**Fig5: prediction models visualization**

For example, a huge dataset for the spam filtering would contain spam messages as well as not-spam messages. In a supervised learning problem, we know



that in the training dataset which message is spam or not-spam, and use this information train.

**Fig6: Deaths from Air pollution in worldwide.**

## Analyze the data

Predictive analytics is driven by predictive modeling. Predictive analytics and predictive models typically include a machine learning algorithm. Bigdata and machine learning is go hand-in-hand. These models are trained over time to respond new data or values, delivering the results the business needs. Predictive modeling heavily overlaps with machine learning.

## Two types of predictive models

**Classification models**:- Predict class membership.

**Regression models**:- Predict a number

- **Advantages of air quality prediction**

1. Air quality has many investments on the multiple values of the forecasting where it invested on the global, community etc.
2. Air quality forecasting mainly helps in the planning decreasing the effects on health and the costs associated.
3. Awareness has the potential to create a cleaner environment and a healthier population002E

- **Disadvantages of air quality prediction.**

1. Air quality has biggest problem of pollution which produce by vehicles, dust etc.

| Air Quality Index (AQI) Values | Levels of Health Concern | Colors |
|---|---|---|
| 0 to 50 | Good | Green |
| 51 to 100 | Moderate | Yellow |
| 101 to 150 | Unhealthy for Sensitive Groups | Orange |
| 151 to 200 | Unhealthy | Red |
| 201 to 300 | Very Unhealthy | Purple |
| 301 to 500 | Hazardous | Maroon |

2. By the pollution many creature, animals are Suffering.
3. Depositions of harmful gases are creating serious threat issues for the quality of life in smart cities.
4. The quality of air is affected by multidimensional factors including time location and uncertain variables.

**Fig7: categorization of the air quality**

The six levels of health concern Hazardous means air they breathe, the effect of health will decrease.

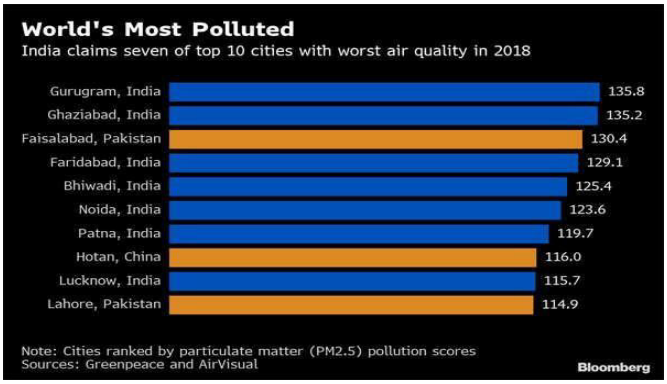



**Fig8: Top 10 World's Most Polluted Cities**

There are 10 cities with highest pollution, seven are in India, that means India is most air polluted country in the world, while one is in China and two are in Pakistan. Among the Indian cities are Gurgaon, Ghaziabad, Faridabad, Bhiwadi, Noida, Patna and Lucknow. The other three are Hotan in China and Lahore and Faislabad in Pakistan.Delhi was ranked at number 11 on the pollution chart.

## Conclusion

This paper reports our recent literature study, survey, reviews and compares current research work on air quality evaluation based on big data

analytics, machine learning models and techniques. The advancement of IoT, Big Data, Artificial Intelligence and machine learning technologies, Real-time air quality monitor and evaluation is desirable for future smart cities. Statistical models have a wide application and require less time to build models, but they require a large amount of historical data and have a high dependence on the data time series approach. AI methods have good performance and can solve nonlinear data, but the models are unstable. AI methods have a high dependence on data. Moreover, most optimization algorithms are easy.

## References

[1] Published in: 2017 IEEE International Conference on Big Data Date of Conference: 11-14 Dec. 2017 Date Added to IEEE Xplore: 15 January 2018

[2] Mrs. A. GnanaSoundariMtech, (Phd) ,Mrs. J. nanaJeslin M.E, (Phd), Akshaya A.C. "Indian Air Quality Prediction And Analysis Using Machine learning". International Journal of Applied Engineering Research ISSN 0973-4562 Number 11, 2019 (Special Issue)

[3] Suhasini V. Kottur , Dr. S. S. Mantha. "An Integrated Model Using Artificial Neural Network (Ann) And Kriging For Forecasting Air Pollutants Using Meteorological Data". International Journal of Advanced Research in Computer and Communication Engineering ISSN (Online) : 2278-1021 ISSN (Print) : 2319-5940 Vol.4, Issue 1, January 2015

[4] RuchiRaturi, Dr. J.R. Prasad ."Recognition Of Future Air Quality Index Using Artificial Neural Network".International Research Journal of Engineering and Technology (IRJET) .e-ISSN: 2395-0056 p-ISSN: 2395-0072 Volume: 05 Issue: 03 Mar-2018

[5] Aditya C R, Chandana R Deshmukh, Nayana D K, Gandhi Vidyavastu ." Detection and Prediction of Air Pollution using Machine Learning Models". International Journal of Engineering Trends and Technology (IJETT) – volume 59 Issue 4 – May 2018

[6] Gaganjot Kaur Kang, Jerry ZeyuGao, Sen Chiao, Shengqiang Lu, and Gang Xie." Air Quality Prediction: Big Data and Machine Learning Approaches". International Journal of Environmental Science and Development, Vol. 9, No. 1, January 2018.

[7] E. Kalapanidas and N. Avouris, "Applying machine learning techniques in air quality prediction," in Proc. ACAI, vol. 99, September 2017