

# A Survey on Machine Learning-Based Prediction of Air Quality Index.

Miss. Sampada Bharat Deshmukh  
[sampadad99@gmail.com](mailto:sampadad99@gmail.com)

Miss. Komal Prakash Shirsat  
[komalshirsat84@gmail.com](mailto:komalshirsat84@gmail.com)

Miss. Sakshi Prashant Dhotre  
[Dhotresakshi089@gmail.com](mailto:Dhotresakshi089@gmail.com)

Miss. Pooja Ravsaheb Jejurkar  
[poojajejurkar1403@gmail.com](mailto:poojajejurkar1403@gmail.com)

## Department of Computer Engineering

Shri Chhatrapati Shivaji Maharaj College of Engineering, Ahmednagar

### Abstract

We used machine learning to estimate the air quality index of a certain place in order to forecast the air quality in India. India's air quality index is a standard metric for indicating pollution levels (so<sub>2</sub>, no<sub>2</sub>, rspm, spm, etc.) across time. Many researchers are working on real-time air quality monitoring and forecasting in order to obtain reliable data that will aid in the development of new technologies. putting in place a variety of government policies relating to the For example, the environment or air pollution, as well as making critical judgments. There have been numerous recent advances in air quality forecasting. as well as monitoring strategies The majority of the approaches are automated. As a common analysis tool, machine learning (ML) is used. Because of its unique distinguishing characteristics For our predictive problem, we use cost estimation to improve the model's efficiency. On this data, multiple Support Vector Machine (SVM) Machine Learning (ML) methods were used to estimate emission rate, and a comparison study was performed. These algorithms were written in Python, and the mean square error of each was calculated to ensure that they were accurate.

**Keyword:** Air Quality Predication, Data Processing, Feature extraction, Classification, Machine Learning, Support Vector Machine (SVM) etc.

---

### Introduction

Machine learning is used to forecast the future based on past data. Computer studying (ML) is a type of artificial intelligence (AI) that allows computers to learn without being explicitly programmed. Machine learning focuses on the development of computer applications that can change when exposed to new information, as well as the fundamentals of laptop studying, such as the implementation of a simple laptop learning algorithm in Python. The guidance and prediction process makes use of specialised algorithms. It feeds the training data to an algorithm, which uses this training knowledge to make predictions on brand new test data. The goal is to look into machine learning-based techniques for predicting air quality. The air quality dataset has been pre-processed in terms of univariate analysis, bivariate and multivariate analysis, missing value treatments, data validation, and data cleaning/preparation. Then, using supervised machine learning technique is Support Vector Machines, air quality is predicted.

### Problem Definition

Observing and maintaining a high standard of air quality has become a critical challenge in metropolitan areas with more industries, businesses, and people. As the population grows, so does the demand for transportation, as well as the use of electricity and fuels. We are well aware that there is a lot of waste dumped on the land. The air is also highly contaminated, posing a greater threat to all types of living organisms on the planet. This necessitates the need for monitoring and assessing air quality, and the government should be alerted to take the necessary actions. This research focuses on conducting an effective analysis of all major works done in this area using machine learning algorithms.

## Existing System

The experimental results show that it improves prediction efficiency over basic models. Air pollution has attracted a large density involving men and women's daily lives. It has a terrible impact on human health and daily lifestyles during episodes of severe air pollution. As the number of reasons and types of air pollutants has increased, so has the difficulty of pollutant attention prediction. As a result, it is critical to use ecological analysing data to more accurately predict city air pollution levels. In this type of prediction, conventional prediction methods such as numerical evaluation and machine learning are commonly used. Nonetheless, a few disadvantages of those methods have recently been identified, as listed below.

## Literature Survey :

- 1) Zhongjie Fu, Haiping Lin, Bingqiang Huang and Jiana Yao “Research on air quality prediction method in Hangzhou based on machine learning”

Air pollution has become the subject of many current environmental studies, and the quality of air is directly related to the quality of life and health of human beings. In this paper, the Bayesian network model is used to predict air quality in Hangzhou. Six air pollutants SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, CO, PM<sub>2.5</sub> and PM<sub>10</sub> are used as the evaluation factors of the model, and AQI value is the output of the model, and then the Bayesian network model is established. Finally, the model is used to predict air quality and compare with the actual value. The results show that the accuracy of air quality prediction is over 80%, and the predicted value is close to the actual value in most cases, and this shows that Bayesian network model has a certain practical value as a means of air quality prediction.

Air is the basis of our survival. Air quality index is the index that quantitatively describes the air quality condition[1][2]. The greater the value of the six categories, the more serious the air pollution is, the greater the risk to human health, as shown in Table 1. The factors that affect the ambient air quality index are PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, CO, O<sub>3</sub> TEMP (published every 24 hours by the Ministry of Environmental Protection).

- 2) Mrs. A. Gnana Soundari Mrs. J. Gnana Jeslin, Akshaya A.C “INDIAN AIR QUALITY PREDICTION AND ANALYSIS USING MACHINE LEARNING”

We forecast the air quality of India by using machine learning to predict the air quality index of a given area. Air quality index of India is a standard measure used to indicate the pollutant (so<sub>2</sub>, no<sub>2</sub>, rspm, spm. etc.) levels over a period. We developed a model to predict the air quality index based on historical data of previous years and predicting over a particular upcoming year as a Gradient decent boosted multivariable regression problem. we improve the efficiency of the model by applying cost Estimation for our predictive Problem. Our model will be capable for successfully predicting the air quality index of a total county or any state or any bounded region provided with the historical data of pollutant concentration. In our model by implementing the proposed parameterreducing formulations, we achieved better performance than the standard regression models. our model has 96% accuracy on predicting the current available dataset on predicting the air quality index of whole India, also we use AHP MCDM technique to find of order of preference by similarity to ideal solution.

As the largest growing industrial nation, India is producing record amount of pollutants specifically Co<sub>2</sub>, pm<sub>2.5</sub> etc and other harmful aerial contaminants. Air quality of a particular state or a country is a measure on the effect of pollutants on the respected regions, as per the Indian air quality standard pollutants are indexed in terms of their scale, these air quality indexes indicates the levels of major pollutants on the atmosphere. There are various atmospheric gases which causes pollution on our environment

- 3) Limei Ma, Yijun Gao ,Chen Zhao “Research on Machine Learning Prediction of Air Quality Index Based on SPSS”

Air From autumn to winter, air pollution in North China is very serious. If we can make effective prediction, we can achieve effective prevention. In this paper, SPSS is used to associate a phenomenon with multiple factors through the optimal combination of multiple independent variables. It is more effective and realistic to predict or estimate the dependent variables than to predict or estimate only one independent variable. Regression with two or more independent variables is multivariate linear regression, and multivariate linear regression has more practical significance than univariate linear regression. Through the establishment of multiple linear regression model, the factors affecting air quality were screened and analyzed. The factors influencing air quality index were PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>,

NO<sub>2</sub>, CO and O<sub>3</sub>. Through regression analysis of one year's data samples, the prediction model is obtained. It has been proved that the prediction method is worth popularizing.

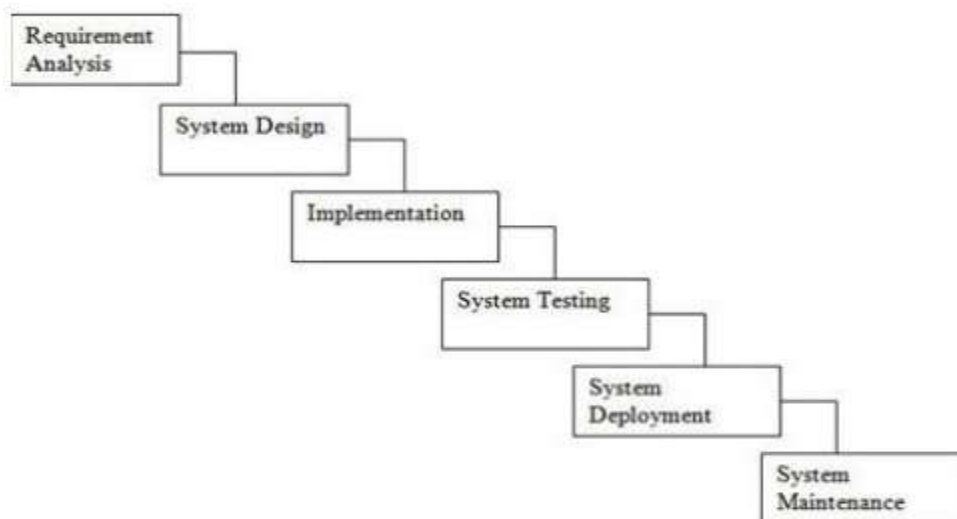
In recent years, due to the rapid development of China's economy, environmental problems have become prominent, especially air pollution. The air quality is directly related to the quality of human life, health and safety [1]. According to the statistics of 2019, China accounts for 7 of the top 10 air pollution cities in the world, which means that China has a long way to go in air pollution control. Therefore, studying the causes of air pollution through big data and predicting the air quality status and change trend in the future can provide scientific decision-making basis for environmental monitoring departments to reasonably control, manage and effectively prevent air pollution [2]. Many scholars have done research on air quality prediction methods. Wu used GM(1,1) model with the fractional order accumulation (FGM(1,1)) to predict the average annual concentrations of SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub> and PM<sub>10</sub> in the Beijing-Tianjin-Hebei region from 2017 to 2020 [3]. Nevin used Fuzzy C-Auto Regressive Model (FCARM) as a prediction model to reflect the regional behavior of weekly PM<sub>10</sub> concentrations in Turkey [4]. Zhu adopted two hybrid models (EMD-SVR-Hybrid and EMD-IMFs-Hybrid) to forecast air quality index (AQI) data, and the AQI forecasting results of Xingtai showed that the two proposed hybrid models are superior to ARIMA, SVR, GRNN, EMDGRNN, Wavelet-GRNN and Wavelet-SVR [5]. Yang proposed a new air quality monitoring and early warning system, including an assessment module and forecasting module [6]. In the air quality assessment module, fuzzy comprehensive evaluation is used to determine the main pollutants and evaluate the degree of air pollution more scientifically.

### Algorithm

- Support Vector Machine (SVM)-
- Support vector machines (SVMs) were first used to solve classification problems in.
- The goal is to find the best separating hyperplane between classes. When a linear separator cannot find a solution, data points are projected into a higher-dimensional space, where previously nonlinearly separable points become linearly separable by using kernel functions.
- The entire task can be formulated as a quadratic optimization problem that can be solved precisely.

### Analysis Models: SDLC Model to be applied

SDLC Models stands for Software Development Life Cycle Models. In this article, we explore the most widely used SDLC methodologies such as Agile ... Each software development life cycle model starts with the analysis, in which the Also, here are defined the technologies used in the project, team load.



**Fig: Waterfall Model**

### Advantages

1. Looks for anomalies in ongoing visitors, recreation, transactions, and behaviour.

2. Capability to discover previously unknown types of attacks.
3. The differences between baseline behaviour and ongoing exercise are catalogued.
4. A sensible method for increasing public awareness of community attacks.
5. To easily predict air quality

### Applications

1. Predict air Quality.
2. In the past few years there have been tremendous advancements in the forecasting models.
3. Simultaneously there is abundant amount of research work in forecasting under almost every scientific applied and in engineering domain like the business and financial, environmental science, medical applications, telecommunication and many more. Recently machine learning techniques have shown admirable performance in predictive analysis.

### Conclusion

The analytical procedure began with information cleaning and processing, then progressed to incomplete records, detailed evaluation, and finally model construction and evaluation. Using SVM algorithm and to analyse air quality.

### References

- [1] Ma Limei, WangFangwei, the Prediction of Haze Based on BP NeuralNetwork and Matlab.
- [2] GaoNing. International Journal of Advanced Network Monitoring and Controls Volume 02, No.2, 2017.
- [3] Ma Limei. Neural network stock forecasting system based on. 2005 in Beijing University of Posts and Telecommunications.
- [4] Ai Hongfu, et al. Forecast of fog and haze based on Bp artificial neural network. Computer simulation, 2015(1): 402:406
- [5] Wang li Forecast and MATLAB implementation of crop pest forecasting based on BP neural network. Anhui Agriculture University, 2003,29 (2): 191-151.
- [6] Zhang mingli Prediction Method of Haze Based on Multiple Linear Regression,computer science,2016(s1):526-528.
- [7] Xiao yangyang ,Prediction of Urban Water Consumption Based on SPSS Multivariate Linear Regression Model,Water Conservancy Science and Technology and Economy,2018(5):6-10.
- [8] Cao Yuru, Research on Qualitative Data Analysis Method Based on SPSS Correspondence Analysis,,Fujian Computer,2018(10):4-6.
- [9] Feng, X., Fu, T.M., Cao, H.S. (2019) Neural network predictions of pollutant emissions from open burning of crop residues: Application to air quality forecasts in southern China. Atmos. Environ., 204: 22-31.
- [10] Wu, Q.L., Lin, H.X. (2019) A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors. Sci. Total Environ., 683: 808-821.
- [11] Sagar, V., Belavadi, S.R., Ranjani, R.(2020) Air quality forecasting using LSTM RNN and wireless sensor networks. Proc. Comp. Sci., 170: 241-248.
- [12] Seng, D.W., Zhang, Q.Y., Zhang, X.F. (2021) Spatiotemporal prediction of air quality based on LSTM neural network. Alex. Eng. J., 60(2): 2021-2032.
- [13] Zhang, L., Liu, P., Zhao, L. (2021) Air quality predictions with a semi-supervised bidirectional LSTM neural network. Atmos. Pollut. Res., 12(1): 328-339.
- [14] Du, X.H., Chen, R.J., Meng, X. The establishment of national air quality health index in China. Environ. Int., 138: 105594.

- [15] Xue, J., Xu, Y., Zhao, L.J. (2019) Air pollution option pricing model based on AQI, *Atmos. Pollut. Res.*, 10(3): 665-674.
- [16] Corani, G., Scanagatta, M. (2016) Air pollution prediction via multi-label classification. *Environ. Modell Softw.*, 80: 259-264.
- [17] Hua, H.D., Wang, C.X. (2018) Prediction and diagnosis of air quality in Dalian city based on Bayesian Networks. *Safety Environ. Eng.*, 25(1): 58-63.

