



**IJIRCCCE**

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

**Volume 9, Issue 6, June 2021**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.542**



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

# Air Quality Prediction Using Machine Learning Model

Harshita M Raj, Dr. Amutha S

PG Student, Department of CSE, Dayananda Sagar College of Engineering, VTU, Bengaluru, Karnataka, India

Professor, Department of CSE, Dayananda Sagar College of Engineering, VTU, Bengaluru, Karnataka, India

**ABSTRACT:** Urban air quality monitoring has been a constant challenge with the advent of industrialization. Air pollution has become a major challenge for government and public all over the world. The government and other related departments are trying to do early precautions by processing the data and know the future effects. In prediction concept machine learning technique is playing a vital role to solve the issues and the problem.

In our system which has been proposed we are concentrating on the air pollution happening in the smart cities. We proposes a machine learning technique to minimize the pollution which is one of the major issue and concern in the smart cities. The machine learning algorithms used in our proposed system are : Linear Regression (LR) , K-Nearest Neighbor (KNN) and Random Forest. Dataset here we used are Delhi AQI. The output of our proposed system has given better result than the existing system.

**KEYWORDS:** Air Pollution, Machine Learning Algorithms, Linear Regression, KNN, Random Forest.

## I INTRODUCTION

Internet of things is interconnected to the each and every person in there day today life. Which will help to improve quality of humans life by connecting to n number other devices. Till now as per the nominal check there 50 billion devices has been connected to the IoT. Different domains are wanted to connected each other to improve the accuravy and prediction in the era of smart city. Produced information are heterogeneous, conveyed and consistently changing and expanding step by step, henceforth boosting the information examination effectiveness in such information is significant.

Smart city concept itself is a big deal which helps to monitor the city in a better way and also it makes the domains more interactive and efficient. From this the interaction between common people and the government will be more and also helps the urban life. Here we need to concentrate on data which is essential for the smart solutions, it is one of the important aspect and challenging task also. This is basically a result of volume, combination and components of the delivered data. In this way, viable methodology ought to be made to gauge and gain significant information from SC data.

Pollution of air in smart cities is pulling peoples lives into threat and it is the domain where we need to concentrate on, it has become a hit issue and topic to the media and other weather monitoring stations. Many research has been taken place on the domain of air pollution to solve it with different approaches. In our proposed system we are concentrating on air quality prediction using machine known technology. This technology plays a vital role in this critical applications like mining of the data, image processing and other system expert. Linear classifier has been displayed to perform shockingly well with limited quantities of preparing information.

AQI assumption needs colossal proportion of data consequently ANNs would be better choice. Different estimations and devices like Neural Networks, feathery a system, Support Vector Machine, Support Vector Machine for backslide, soft reasoning, Decision Trees, K-Nearest Neighbor has been used in advance for different use cases. Three Machine Learning algorithms used in proposed system is Linear Regression, KNN and Random Forest.

## II. RELATED WORK

In this section research work related to ANN, Air Quality Index Monitoring and Support Vector Machines are explained in detail.

**Artificial Neural Networks** The Artificial Neural Network (ANN) showing is presented by researchers for handling issues using AI. One such work is the steam gasification of palm piece shell using CaO adsorbent and coal base flotsam and jetsam as a stimulus. The effect of the limits like temperature, CaO/biomass extent and Coal base flotsam and jetsam weight rate are shown using ANN. The ANN structure is a trademark model, which is intriguing from information dealing with perspective since it figures and acknowledges choices and terminations a similar the human cerebrum. The Human cerebrum has billions of neurons which are interconnected each other and give over electrochemical signs.

ANNs, as often as possible insinuated as 'Neural Networks' or 'Neural structures' work like a duplicate of the human cerebrum to manage complex issues in AI part.

**AQI Monitoring** An air quality record (AQI) is a quantitative measure used to reliably examine the air idea of different constituents with respect to human prosperity. Kumar et al. proposes a continuous AQI Monitoring System using various limits, for instance, CO, CO<sub>2</sub>, Humidity, PM 2.5, Temperature and pneumatic power. The system is attempted in Delhi and the assessments are differentiated and the data given by the close by regular control authority.

**Sponsorship Vector Machines** :Support Vector Machine (SVM) is an overseen estimation, which is used to describe data into in any event two classes . Part techniques are a class of AI methodology that have become a verifiably standard device for learning tasks, for instance, plan affirmation, gathering or peculiarity acknowledgment. This reputation is on a very basic level an immediate consequence of the achievement of the SVM, likely the most well known bit methodology, and to the way that part machines can be used in various applications as they give a framework from linearity to non-linearity.

### III.PROBLEM STATEMENT

Air Pollution is major problem in smart cities like Bangalore ,delhi. Certain parameters of air quality have higher value than the average which causes the damage to the human life. Governmnet has to take propoer actions of this regard in rural and urban areas.

**Model for AQI Prediction** : The below mention figure shows the model for air pollution monitoring and prediction, At the beginning the data has to be collected and create a dataset. Once it has been done Data has to be splitted to testing and training part. After the training of the data model will get generated That model will be used for further prediction purpose.

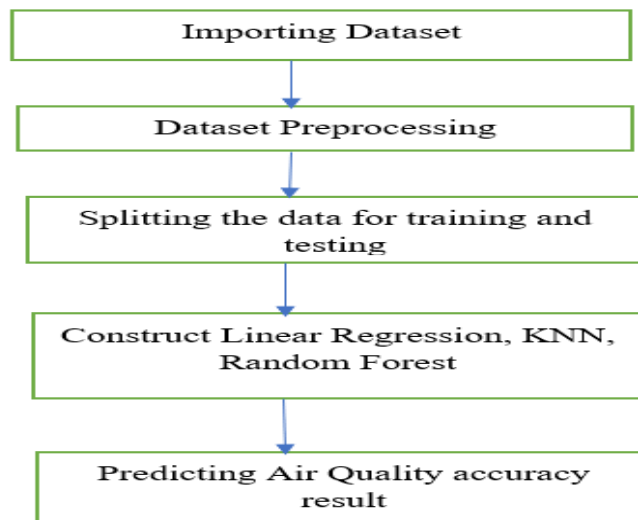


Figure 1 : AQI data prediction Model

### IV.ALGORITHMS USED

The algorithms used here are Linear Regression, KNN and Random Forest.

1. **Linear Regression** : Which is one of the supervised machine learning model in which the model will check the best fit linear line between the dependent and independent variable. Which means the linear algorithm will generate the relationship between dependent and independent variable.

There are two types of linear algorithms are there simple and multiple. Where we are using simple linear algorithm, in which only one independent variable is there and model has to find linear relationship of the dependent variable as shown in below equation.

$$y = b_0 + b_1x$$

2. **K-Nearest Neighbor (KNN)** : KNN is also a supervised machine learning model which assumes the the similarity between new data and available data and put the new data into the category that is most similar to the available categories. KNN is a non-parametric algorithm, that is it does not make any assumption on underlying data. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies the data into category that is much similar to the new data.
3. **Random Forest** :Random Forest is a supervised classification algorithm. We can see it from its name, which is to create a forest by some way and make it random. There is a direct relationship between the number of trees in the forest and the results it can get the larger the number of trees, the more accurate the result.

#### V.STEPS TO IMPLEMENT

- A. **Dataset**: Inside the degree of the endeavor, road traffic, pollution, environment, social, library and preventing data were accumulated from the metropolitan networks of Aarhus and Brasov in Denmark and Romania independently some place in the scope of 2013 and 2015. In this examination, tainting dataset of CityPulse EU FP7 Project is used to comprehend the proposed system. The dataset contains 8 features including ozone, particulate matter, carbon monoxide, sulfur dioxide, nitrogen dioxide, longitude, extension and timestamp was used for attempt. The dataset has 17568 models that are assembled at five-minute ranges. Every model regard is given as EPA's AQI standard. In this examination, ozone and nitrogen dioxide poisons are picked for air quality estimate.
- B. **Model Training** :The proposed model is trained on python idle tool using linear regression model primarily ozone and nitrogen dioxide data with other parameters is separated into two parts as training and testing data. Later Training are carried out by Linear Regression for each type of gas.
- C. **Edges** In this section, AQI levels are isolated into three edge regards by thinking about AQI fundamental level (100).

Air Quality Prediction accuracy is checked for Linear Regression, KNN and Random Forest algorithm. Linear Regression gives the best accuracy and air quality prediction is checked for Linear Regression technique. Experimental results shows the evaluated result of air quality prediction model is working properly and evaluated using LR technique.

Data Sources and Classification Standard The entire experiment is carried out using the dataset collected from the Government official website named Central Pollution Control Board (CPCB), Ministry of Environment, Forest and Climate change Government of India. Selection of Experimental dataset:

- Dataset of Delhi
- Real-time AQI values
- In Delhi city AQI value is integrated of nearly eight pollutants namely T, TM, Tm, SLP, H, VV, V, VM.

Descriptive study for the dataset After collecting the data, most of the psychology researchers move to summarization of data by using different methods. In this research work, being a programming language python is selected to carry out the descriptive statistics. The library function used for the data statistics and data manipulation is the pandas and NumPY. Descriptive statistics using pandas.



## VI.RESULT

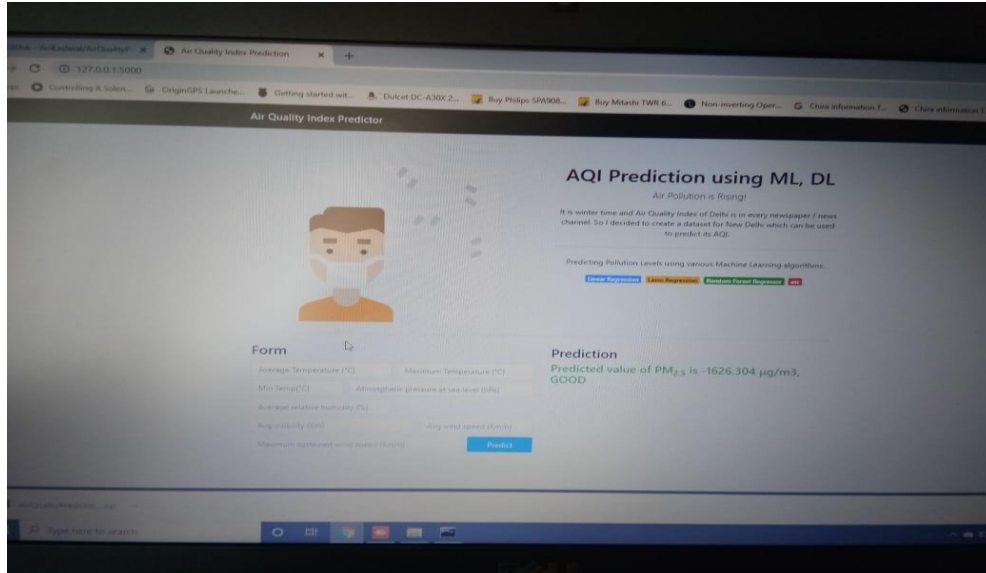


Figure 2 : Final output after prediction

The figure 2 shows the output of the proposed method where linear regression is the algorithm which is implemented in our proposed system.

## VII.CONCLUSION AND FUTURE WORK

In this assessment work, a Machine Learning model for Air Quality Index assumption for splendid metropolitan regions is proposed. The model is attempted with the Delhi Air Quality data.

Among three techniques used that is Linear Regression, Knn and Random Forest. Linear Regression gives us the best accuracy and air quality prediction is checked for Linear Regression technique. Data mining estimations could be used to assemble the exactness level later on work.

## REFERENCES

- [1] N. McCrea, An Introduction to Machine Learning Theory and Its Applications: A Visual Tutorial with Examples.
- [2] <https://www.forbes.com/sites/davidteich/2018/12/26/machine-learningand-artificial-intelligence-in-business-year-in-review2018/#980755b2041c>.
- [3] P. Kavitha, and M. Usha, "Anomaly Based Intrusion Detection In WLAN Using Discrimination Algorithm Combined with Naive Bayesian Classifier," Journal of Theoretical and Applied Information Technology, vol. 62, no. 3, pp. 646– 653,2014.
- [4] T. Ensari, M. Günay, Y. Nalçakan, E. Yildiz, Overview of Machine Learning Approaches for Wireless Communication. InNext-Generation Wireless Networks Meet Advanced Machine Learning Applications 2019 (pp. 123–140). IGI Global.
- [5] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, H. Liu, Feature selection: A data perspective. ACM Computing Surveys (CSUR), vol. 50, no. 6, 94, 2018 Jan 12.
- [6] S. Ledesma, G. Cerda, G. Aviña, D. Hernández, M. Torres, Feature selection using artificial neural networks. InMexican International Conference on Artificial Intelligence 2008 Oct 27 (pp. 351–359). Springer, Berlin, Heidelberg.
- [7] B. Ghaddar, and J. Naoum-Sawaya, High dimensional data classification and feature selection using support vector machines. European Journal of Operational Research vol. 265, no. 3, 993–1004, 2018 Mar 16.
- [8] M. Shahbaz, S. A. Taqvi, A. C. Loy, A. Inayat, F. Uddin, A. Bokhari, S. R. Naqvi, Artificial neural network approach for the steam gasification of palm oil waste using bottom ash and CaO. Renewable Energy vol. 132, 243–254, 2019 Mar 1.
- [9] B. C. Liu, et al., "Urban air quality forecasting based on multidimensional collaborative Support Vector Regression (SVR): A case study of Beijing-Tianjin-Shijiazhuang", PLOS, 2017.



- [10] C. M. Bishop, "Neural Networks for Pattern Recognition", Oxford, 1995.
- [11] Q. Feng, "Improving Neural Network Prediction Accuracy for PM10 Individual Air Quality Index Pollution Levels", Environmental Engineering Science, vol. 30, no. 12, 725–732, 2013.
- [12] S. Kumar, A. Jasuja, Air quality monitoring system based on IoT using Raspberry Pi. In Computing, Communication and Automation (ICCCA), 2017 International Conference on 2017 May 5 (pp. 1341–1346). IEEE.
- [13] D. Van Le, C. K. Tham, Machine Learning (ML)-Based Air Quality Monitoring Using Vehicular Sensor Networks. In Parallel and Distributed Systems (ICPADS), 2017 IEEE 23rd International Conference on 2017 Dec 15 (pp. 65–72). IEEE.
- [15] L. Wang, Y. P. Bai, "Research on Prediction of Air Quality Index Based on NARX and SVM", Applied Mechanics and Materials (Volumes 602- 605), 3580–3584, 2014.
- [16] <https://timesofindia.indiatimes.com/life-style/healthfitness/health-news/top-8-main-causes-for-air-pollution-indelhi/articleshow/61626744.cms>.



**INNO**  **SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 7.542**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
**INDIA**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details