International Journal of Scientific Research and Engineering Development-– Volume 6 Issue 1, Jan-Feb 2023

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Monitoring and Prediction of Air Pollution Using Machine Learning Models: A Review

Hemangi Tamore, Brinda Temkar, Sanya Wakode, Sunidhi Yadav, Dr.Jyoti Dange

EXTC, Atharva College Of Engineering, Mumbai

Email: tamorehemangi-extc@atharvacoe.ac.in

EXTC, Atharva College Of Engineering, Mumbai

Email: temkarbrinda-extc@atharvacoe.ac.in

EXTC, Atharva College Of Engineering, Mumbai

Email: wakodesanya-extc@atharvacoe.ac.in

EXTC, Atharva College Of Engineering, Mumbai

Email: yadavsunidhi-extc@atharvacoe.ac.in

EXTC, Atharva College Of Engineering, Mumbai

Email: jyotidange2112@gmail.com

----------**********----------

Abstract:

The influence of Machine learning and Data Science is advancing in healthcare, personalized recommendation models, environmental studies as well as in education institutes. It has become an important factor of consideration for both companies as well as individuals. Prediction of air quality is one of the fields that machine learning has given its contributions. Air quality index measures the concentration of various gases like carbon dioxide, carbon monoxide, nitrogen dioxide, sulfur dioxide, particulate matter like smoke, soot, methane that releases after burning natural gas, coal, wood, etc. High concentrations of these substances can cause severe diseases like lung cancer and even premature deaths. Machine learning helps in predicting the air quality so necessary actions can be taken if the pollutants increase more than a certain limit. If pollution of air is not handled carefully and sensibly, some day it can lead to extinction of humans. This paper provides a revision on the results obtained by researchers on monitoring and prediction of air pollution using machine learning and IOT. After thorough reviewing it was observed that the machine learning algorithms used for the analysis were quite effective.

Keywords — **Air pollution, air quality prediction, machine learning, iot, regression, maps**

----------**********----------

INTRODUCTION

Human society is evolving at a great pace. This is evident in every city, including various problems caused due to these evolutions and expansions. The rapid growth in cities naturally increases pollution and pollutants caused due to heavy traffic, industries, etc. Because of this, the quality of air is polluted and has a great impact on human health. There are certain pollutants which affect human health and have a great impact these pollutants are, carbon dioxide (CO2), carbon monoxide (CO), sulfur dioxide(SO2), Ozone (O3), nitrogen oxide (NO) and nitrogen dioxide (NO2) and a complex mixture of solid and liquid droplets called Particulate Matters (PM, e.g., PM2.5, PM10) [2].

There are many studies which show that respiratory disorders are due to air pollution. A study conducted by State of Global Air (SOGA) shows that continuous exposure to air pollution can reduce life expectancy by up to 20 months [2]. Air pollution is accountable for the death of 7 million persons worldwide each year or one in eight

premature deaths yearly [4]. Hence, it is a difficult task to predict the quality of air in a particular region, which makes people less careful about the existing air quality. The World Health Organisation (WHO) set up guidelines and limitations for urban air pollution that should be respected in order to protect the citizens from the pollutants [2]. Almost 570,000 children under the age of five die every year from respiratory infection linked to indoor/outdoor pollution and second-hand smoke [4]. Small children when made to come in contact with air pollution may have a risk of developing respiratory disorders. Therefore, a need to develop a system that predicts air quality accurately is necessary.

Air Quality Index(AQI) is calculated on the basis of concentration of different pollutants like CO2, SO2, CO, NH3, PM2.5, PM10, etc. as shown in Fig. 1 the index uses different colors for better understanding of the level of pollutants in air.Here, green indicates least polluted air while maroon indicates the most polluted air.

AQI Category, Pollutants and Health Breakpoints

AQI Category (Range)	PM_{10} (24hr)	$PM_{2.5}$ (24hr)	NO_2 (24hr)	O_3 (8hr)	CO (8hr)	SO_2 (24hr)	NH_3 (24hr)
Good (0–50)	0–50	0–30	0–40	0–50	0–1.0	0–40	0–200
Satisfactory (51–100)	51–100	31–60	41–80	51–100	1.1–2.0	41–80	201–400
Moderately polluted (101–200)	101–250	61–90	81–180	101–168	2.1–10	81–380	401–800
Poor (201–300)	251–350	91–120	181–280	169–208	10–17	381–800	801–1200
Very poor (301–400)	351–430	121–250	281–400	209–748	17–34	801–1600	1200–1800
Severe (401–500)	430+	250+	400+	748+	34+	1600+	1800+

Fig. 1. Air Quality Index(AQI-index)

This paper includes Section 2 which describes the methodology adopted by different researchers, Section 3 explores the results obtained on the same and finally Section 4 includes the conclusion.

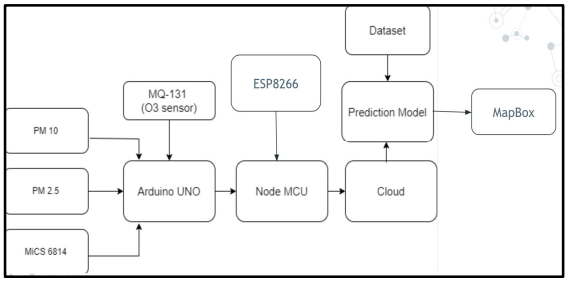

Fig. 2. General steps in predicting Air Quality Prediction

I. Literature Review

In **[1]** the research proposed a Web application based on air pollution monitoring systems using HAPS-based wireless sensor networks that were able to display information about air grade data in graphical and table form. This is in accordance with the design of the website where the air condition menu contains a line-shaped graph that is divided into two sub menus: graphs CO and PM10 levels.

[2] This paper presents an IoT Platform covering the data collection process from the sensing nodes to visualization for the end process. They provide an easy to assemble and duplicate design for research purposes to enable fast sensor deployment and data collection. It is a low level middle- ware, meaning that it can be implemented directly on top of the physical layer of low consumption protocols such as the Zigbee protocol.

In **[3]** the paper proposes that using a multivariate modeling approach enhances the prediction accuracy and reduces error because of the dependency between target gasses and other features included such as temperature, day of the week, and H2S.

In **[4]**this paper the system was designed using an Arduino microcontroller.This system was develped to monitor and analyze real time data air quality and log data to a remote server,the data kept updated over the internet.Air quality measurement were taken using Part per Million matrics(PPM) and analyze using Microsoft Excel.This designed system was taken accurate measurement of Air quality.The accurate result was displayed on the designed hardwares display interface and it could be accessed via cloud on any smart phone service.

[5]This paper has used Machine learning techniques to predict concentration of the so2 in the environment. Sulfur dioxide effects on the skin and mucous membranes of the

eyes,nose,throat,and lung. This system employed models in Time series to predict so2 in the environment.In this paper they have used Time series forecast models for prediction of Air Quality Index in Metro cities.

In **[6]**this SARIMAX and Holt-Winter's models are used to predict the air quality index. These time series forecasting models can be utilized to predict the values of the Air Quality Index(AQI) based on past data. To analyze the performance of models Mean Absolute Percentage Error (MAPE) is used as the score function. The Holt-Winters algorithm can handle seasonality,but results produced by this model have not very accuracy. On the other hand,The SARIMAX can handle seasonality and has much better result accuracy than the Holt-Winters model.

In **[7]**this paper,they have compared the decision tree, linear regression and random forest. Major air pollutants are taken and meteorological conditions are taken using the Arduino Platform. Random forest gives better accurate results due to overfitting that reduces errors But drawback is Random forest uses more memory and high cost. Haotian Jing & Yingchun Wang(2020).

[8]had predicted the air quality index using XG Boost.It uses the weak classifier and shortcoming of the previous weak classifier to form a strong classifier thus reducing the error between predicted and actual values. It uses the K- cross validation.The mean absolute error and coefficient of determination is determined to predict the difference between actual and predicted value. The drawback is that it takes the previous value and is affected by outlier unwanted pollutants in the air.

[9]Their research had predicted the level of air pollutant with Recurrent neural network at any given time and removed the drawback of hourly prediction due to memorization power of algorithm but lacks in working without memory operation.

[10]This paper has determined PM10 level best with Random Forest but does not accurately predict the level of dangerous pollutants but can work with incomplete data sets.

III. Comparative Study And Analysis

In this section we observe the performance of different machine learning models, adopted by researchers in their studies which are as follows:

Table I: Performance of Different ML Models

ML Model	SO2	CO	O3	NO 2	PM 2.5	PM 10
Linear Regression	0.125	0.02	0.09	0.1	0.02	0.02
Decision Tree	0.8060	0.61	0.62	0.64	0.75	0.61
Random Forest regression	0.856	0.79	0.79	0.701	0.86	0.79
SARIMAX	0.863	0.75	0.832	0.866	0.72	0.81

The above comparison was obtained by observing the results and performance parameters of the mentioned algorithms in [6] and [7].

IV. CONCLUSION

In this paper, we have successfully understood the need of machine learning and the huge role that it plays in prediction of air pollution. The papers discussed above proposed Air Pollution Monitoring and Prediction systems using Machine Learning Models out of which a few successfully predicted the pollution levels accurately while

others didn't. After going through these papers we observed that SARIMAX was the best proposed model as it took the parameter of seasonality into consideration which helped in correctly predicting the air quality that would then alert potential users and also help in future predictions.

REFERENCES

[1] Iskandar, Adha, M. B., Hendrawan, & Edward, I. Y. M. (2019). *Design and Implementation of Web Application on Air Pollution Monitoring System Using Wireless Sensor Network Based on HAPS. 2019 IEEE 5th International Conference on Wireless and Telematics (ICWT).* doi:10.1109/icwt47785.2019.897824110.1109/icwt47785.2019.8978241

[2] Ben-Aboud, Y., Ghogho, M., & Kobbane, A. (2020). *A research-oriented low-cost air pollution monitoring IoT platform. 2020 International Wireless Communications and Mobile Computing (IWCMC).* doi:10.1109/iwcmc48107.2020.9148176 10.1109/iwcmc48107.2020.9148176

[3] Bashir Shaban, K., Kadri, A., & Rezk, E. (2016). *Urban Air Pollution Monitoring System With Forecasting Models. IEEE Sensors Journal, 16(8), 2598–2606.* doi:10.1109/jsen.2016.2514378

[4] Okokpujie, Kennedy & Noma-Osaghae, Etinosa & Odusami,Modupe & John, Samuel & Oluwatosin, Oluga. (2018). A Smart Air Pollution Monitoring System. International Journal of Civil Engineering and Technology. 9. 799-809.

[5] C R, Aditya & Deshmukh, Chandana & K, Nayana & Gandhi, Praveen & astu, Vidyav. (2018). Detection and Prediction of Air Pollution using Machine Learning Models. International Journal of Engineering Trends and Technology.

[6] Arora, Himanshu & Solanki, Arun. (2020). Prediction of Air Quality Index in Metro Cities using Time Series Forecasting Models. Xi'an Jianzhu Keji Daxue Xuebao/Journal of Xi'an University of Architecture & Technology. 12. 3052-3067. 10.37896/JXA 12.05/1721.

[7] Venkat Rao Pasupuleti , Uhasri , Pavan Kalyan , Srikanth and Hari Kiran Reddy, Air Quality Prediction Of Data Log By Machine Learning,2020,IEEE.

[8] Haotian Jing & Yingchun Wang, Research on Urban Air Quality Prediction Based on Ensemble Learning of XGBoost, 2020, E3S Web of Conferences.

[9] Xiaosong Zhao , Rui Zhang, Jheng-Long Wu, Pei-Chann Chang and Yuan Ze University, A Deep Recurrent Neural Network for Air Quality Classification, 2018, Journal of Information Hiding and Multimedia Signal Processing.

[10] Nicolás Mejía Martínez, Laura Melissa Montes, Ivan Mura and Juan Felipe Franco, Machine Learning Techniques for PM10 Levels Forecast in Bogotá,2018,IEEE.

[11] Kennedy Okokpujie, Etinosa Noma-Osaghae, Odusami Modupe, Samuel John, and Oluga Oluwatosin, "A SMART AIR POLLUTION MONITORING SYSTEM," International Journal of Civil Engineering and Technology (IJCIET), vol. 9, no. 9, pp. 799–809, Sep. 2018.

[12] C. Santos, J. A. Jiménez, and F. Espinosa, "Effect of Event-Based Sensing on IOT Node Power Efficiency. Case Study: Air Quality Monitoring in Smart Cities," IEEE Access, vol. 7, pp. 132577–132586, 2019.

[13] D. Wei, "Predicting air pollution level in a specific city," 2014.

[14] Kostandina Veljanovska and Angel Dimoski, "Air Quality Index Prediction Using Simple Machine Learning Algorithms," International Journal of Emerging Trends & Technology in Computer Science, vol. 7, no. 1, 2018.

[15] D. Zhu, C. Cai, T. Yang, and X. Zhou, "A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization," Big Data and Cognitive Computing, vol. 2, no. 1, p. 5, Mar. 2018.

[16] A. Masih, "Machine learning algorithms in air quality modeling," Global Journal of Environmental Science and Management, vol. 5, no. 4, pp. 515–534, 2019

[17] Aditya C R, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu;Detection and Prediction of Air Pollution using Machine Learning Models(IJETT)

[18] https://archive.ics.uci.edu/ml/datasets/Air+quality

[19] David A. Freedman (2009). Statistical Models: Theory and Practice. Cambridge University Press. p. 26. A simple regression equation has on the right hand side an intercept and an explanatory variable with a slope coefficient.

[20] Rokach, Lior; Maimon, O. (2008). Data mining with decision trees: theory and applications. World Scientific Pub Co Inc. ISBN 978- 9812771711. [12] BreimanL (2001)."RandomForests". MachineLearning. 45 (1):32. doi:10.1023/A:1010933404324

