IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Monitoring and Prediction of Air Pollution using Machine Learning Models

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

Professor, Student, Student, Student, Student
Dr. Jyoti Dange, Professor, Dept. of EXTC Engineering, ACE, India.
Atharva College of Engineering, Mumbai, India

Abstract: As advancements in technology continue, machine learning and IoT has proved to be beneficial in healthcare systems, recommendation systems, educational institutions, environmental impact analysis, etc. It plays an important role in consideration of companies as well as individuals both. One of the fields where Machine Learning has given its contributions is prediction of air quality. With the help of air quality index, measurements of the concentration of various gasses can be done. These include carbon dioxide (CO2), carbon monoxide (CO), nitrogen dioxide (NO2), sulphur dioxide (SO2). Particles such as dust, dirt, smoke, soot, methane, etc. are released while burning coals, woods, and natural gasses. Liberation of such substances in high concentration may cause severe diseases like lung carcinoma or early deaths. Machine learning is useful in predicting air quality. Hence, necessary efforts can be taken if the pollution increases beyond a limit. If air pollution is not controlled carefully and prudently, it can lead to human extinction.

Index Terms **– air pollution, air quality, iot, machine learning**

1. INTRODUCTION

Human society is advancing swiftly with pace. With all the advancements, various problems arise which are evident in every city nowadays. The rapid progress in cities leads to an increase in pollution and release of pollutants in the environment due to heavy traffic, emissions, industrial wastes, etc. Due to this, the air quality is affected and which results in impact on human health. Various common pollutants released in the air are carbon dioxide (CO2), carbon monoxide (CO), nitrogen oxide (NO), nitrogen dioxide (NO2), sulphur dioxide (SO2), ozone (O3), droplets called Particulate Matter (PM), etc.

Respiratory disorders are mainly caused due to air pollution. A study conducted by State of Global Air (SOGA) proclaims that continuous exposure to air pollution can reduce the life expectancy up to 20 months. There are 7 million deaths every year which are caused due to air pollution. It is difficult to predict the air quality of a particular region. Hence, people are less aware about the air quality. Some guidelines have been set up by the government and The World Health Organisation (WHO) for air pollution in urban areas which should be followed and respected by the citizens in order of protection from the harmful pollutants. Due to air pollution, a large number of premature deaths, deaths of children below the age of five years is being observed. Small children as well as old aged people when encounters pollutants in air, develop a higher risk of respiratory diseases. Therefore, a system which monitors as well as predicts the accurate air pollution in a particular region is necessary to be developed.

II. NEED OF THE STUDY

There is a growing need for more accurate and reliable air pollution monitoring systems to protect the public health and the environment. The existing air quality monitoring systems have limitations in terms of their spatial and temporal resolution, as well as their ability to accurately predict pollution levels. Machine learning techniques have been developed which can help overcome some of these limitations by providing more accurate and real-time predictions of the air quality. The study on air pollution monitoring system can be helpful in:

2.1 Accuracy and Real-time monitoring

Machine learning techniques can help improve the accuracy of air pollution monitoring systems by processing large volumes of data and identifying complex patterns and relationships in the data. Machine learning techniques can enable real-time monitoring of air quality, which is important for identifying and responding to pollution hotspots quickly.

2.2 Spatial and temporal resolution

Machine learning techniques can help improve the spatial and temporal resolution of air pollution monitoring systems, allowing for more detailed and accurate monitoring of pollution levels.

2.3 Cost effectiveness and policy-making

Machine learning techniques can help reduce the cost of air pollution monitoring by reducing the need for manual data collection and analysis. Accurate and reliable air pollution monitoring systems can provide policymakers with the information they need to make informed decisions about air quality regulations and policies.

III. RESEARCH METHODOLOGY

The methodology section outlines the plan and method that how the study is conducted. This includes Universe of the study, sample of the study, Data and Sources of Data, study's variables, and analytical framework. The details are as follows;

3.1 Data and Sources of Data

Air quality monitoring stations provide data on various air pollutants, including particulate matter, Nitrogen oxides, Sulphur dioxide, and ozone. These stations are often run by government agencies and provide real-time or historical data on air quality. Satellite imagery can provide a global view of air quality by measuring the concentration of certain gases in the atmosphere, such as nitrogen dioxide. This data can be used to identify hotspots of air pollution and track the movement of pollutants over time.

Mobile sensors, such as air quality monitors mounted on cars or drones, can provide data on air quality in specific locations. These sensors can be used to gather data in areas where there are no air quality monitoring stations to supplement existing data. Weather data such as temperature, humidity, wind speed and wind direction can impact air quality by influencing the formation and dispersion of pollutants. This data can be used to develop predictive models that take weather conditions into account.

Demographic and socio-economic data such as population density, income, educational levels can provide insights into the social and economical factors that contribute to air pollution. This data can be used to develop targeted policies and interventions to reduce air pollution in vulnerable communities.

3.2 Theoretical framework

Air pollution is a complex problem that involves the release of various pollutants into the atmosphere, including particulate matters, nitrogen dioxides, sulfur dioxide and ozone. Air pollution results from a wide range of human activities and natural sources. Transportation is a major source of air pollution, particularly in urban areas. Vehicles emit various harmful pollutants that can contribute to respiratory problems and cardiovascular diseases. These pollutants that can also damage crops and forests and contribute to climate change. Industrial activities such as manufacturing, mining and power generation are another significant source of air pollution. These activities can release a range of pollutants causing various diseases, damage to vegetation and wildlife and contribute to climate change. Agricultural activities such as livestock farming and fertilizer use can also contribute to air pollution. Livestock farming produces ammonia and methane, which contribute to climate change and respiratory problems. Use of fertilizers can release nitrogen oxides and other pollutants into the air. Natural sources of air pollution include dust storms, wildfires, and volcanic eruptions. These events can release large amounts of particulate matter, nitrogen dioxide, etc. Indoor sources of air pollution include cooking, heating, and smoking. The effects of air pollution on human health and the environment are complex and varied. Short-term exposure to high levels of pollutants can cause respiratory problems, cardiovascular disease, and other health effects. Long term exposure to lower levels of pollutants can also lead to chronic respiratory problems and other health effects. Air pollution can also damage crops and forests, harm wildlife and contribute to climate change. By understanding the sources of air pollutants and their effects on human health and the environment.

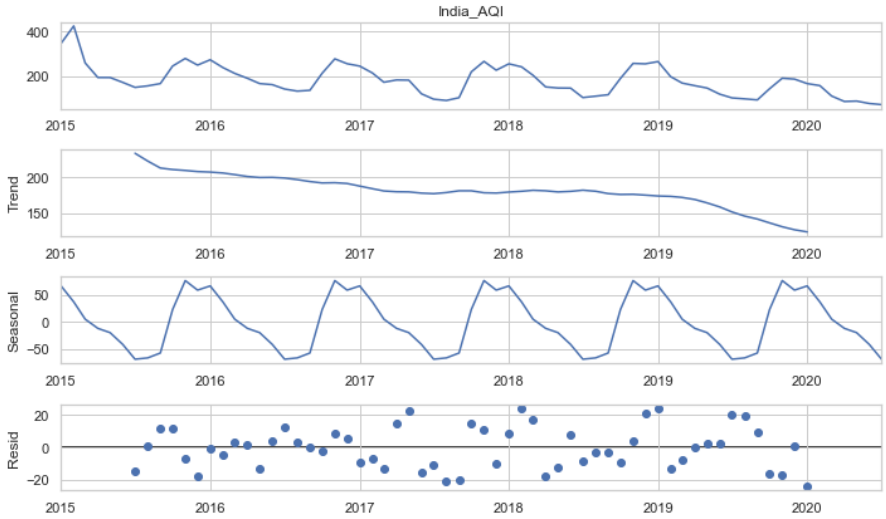
Machine learning is a subfield of artificial intelligence that involves training algorithms to make predictions or decisions based on data. There are different machine learning algorithms that can be used in predicting air pollution viz. Regression Algorithms such as Linear Regression, Multiple Regression, etc., classification algorithms such as Decision Tree, Random Forest and Support Vector Machine. Neural networks and Ensemble methods.

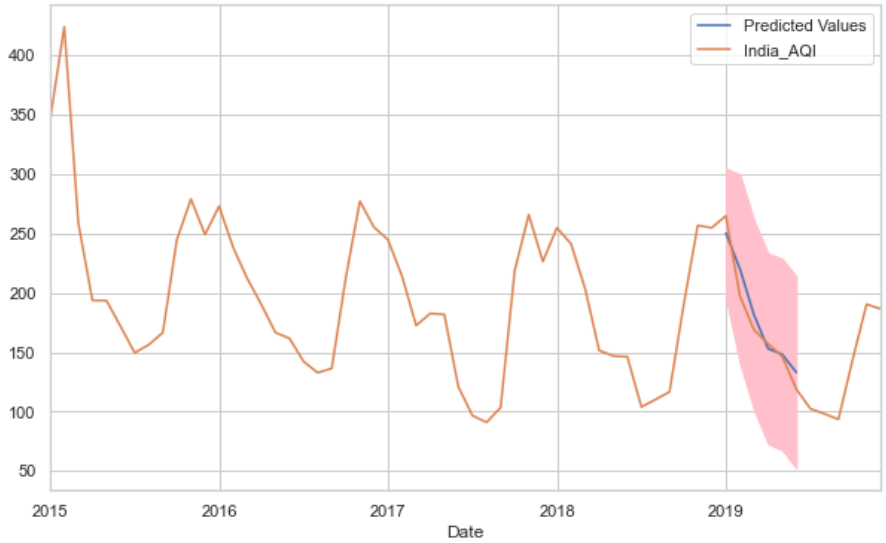
3.4.3 Comparison of the Models

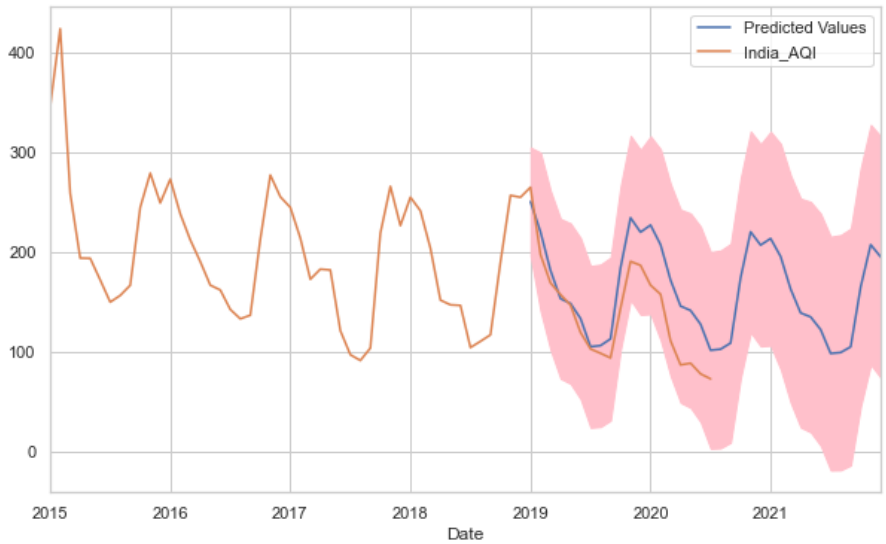
Type of Algorithm	SO2	CO	O3	NO2	PM2.5	PM10
Linear Regression	0.125	0.02	0.09	0.1	0.02	0.02
Decision Tree	0.8060	0.61	0.62	0.64	0.75	0.61
Random Forest regression	0.856	0.79	0.79	0.701	0.86	0.79
SARIMAX	0.863	0.75	0.832	0.866	0.72	0.81

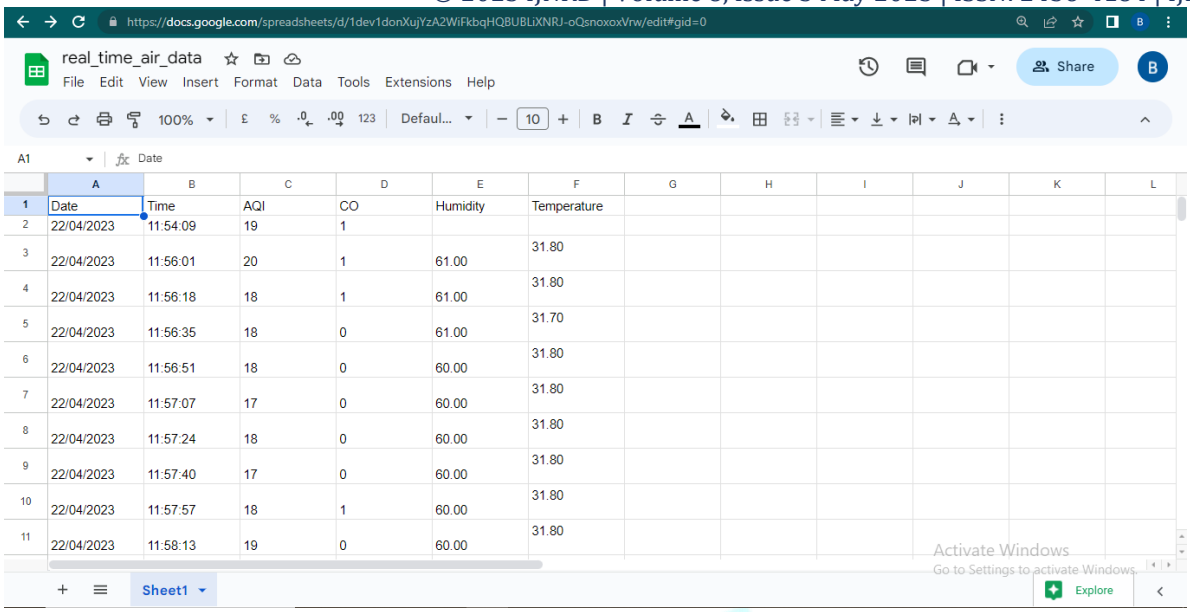
The above comparison was obtained by observing the results and performance parameters of the mentioned algorithms in [6] and [7].

IV. RESULTS AND DISCUSSION









real_time_air_data

Date	Time	AQI	CO	Humidity	Temperature
22/04/2023	11:54:09	19	1		
22/04/2023	11:56:01	20	1	61.00	31.80
22/04/2023	11:56:18	18	1	61.00	31.80
22/04/2023	11:56:35	18	0	61.00	31.70
22/04/2023	11:56:51	18	0	60.00	31.80
22/04/2023	11:57:07	17	0	60.00	31.80
22/04/2023	11:57:24	18	0	60.00	31.80
22/04/2023	11:57:40	17	0	60.00	31.80
22/04/2023	11:57:57	18	1	60.00	31.80
22/04/2023	11:58:13	19	0	60.00	31.80

The model uses Machine Learning algorithm viz. SARIMAX to learn different values that can be present in the dataset. Once the learning is done the model is ready to make predictions and draw conclusions on different pollutants.

The accuracy metrics will be calculated using Mean Absolute Error and Root Mean Squared Error. The results of the algorithm which provides the highest accuracy is displayed on the website. A feature called Maps is included in the website where the current pollution levels can be monitored in the area of a particular city. The Flask API is a lightweight web application framework that is used to create a RESTful API. The Flask API serves the air quality data to the website. In this, a Flask API is created to provide access to the air quality data stored in the database.
Once all the integration is done, the air quality data is visualized using the maps and graphs. The air quality data is displayed on the maps using markers that show the real-time air quality parameters at different locations.

After the whole system is developed, it must be deployed on a server. The flask application is hosted on the server and it can be accessed by the users using a web browser.

Overall, the air pollution monitoring and predicting system using Arduino and Machine Learning algorithm involves collecting the data, pre-processing the data, using SARIMAX model for time-series predicting and fore-casting, creating a Flask API to serve the air quality data, visualizing the data using maps and graphs, and lastly deploying the system on a server for public access.

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/icwt47785.2019.8978241
10.1109/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