Air Quality Index Prediction using Machine Learning

Harsh Jaiswal[1], Aman Singh[2], Akhilesh Chauhan[3], Dicksha Sharma[4]

[1,2,3] *Student, Department of Computer Science and Engineering, ITM Gida Gorakhpur, India.*
[4] *Assistant Professor, Department of Computer Science and Engineering, ITM Gida Gorakhpur, India.*

Abstract: *As we know, now a days the quality of air decreasing day by day and is very harmful for humans. Before going to outside of home it is now necessary to know the quality of air to take the necessary precautions. The AQI of a particular area is a measure of pollutants mixed in air of the respected region. As per the air quality standard pollutants are indexed in terms of their scale, these air quality indexes indicates the levels of major pollutants mixed in the atmosphere. Today, major cities experience levels of air pollution that are dangerously higher than the government-set air quality index standard. It significantly affects a person's health. The Machine Learning (ML) algorithms are capable of making predictions about air pollution. We are here to check the air quality index of a particular area to find air pollution level in that area. The classifiers used in our project are Random Forest Regressor. The measured results of our experiments show that the Random Forest Regressor algorithm outperforms other algorithms, and that it reaches the highest accuracy not only in air quality prediction, but also in other detection systems.*
Key Word: Machine Learning Techniques, AQI, Random Forest Regressor

I. INTRODUCTION

It is a system that runs using machine learning algorithms to detect the air quality index [1]. Air pollution is one of the most important problems that the globe is now facing. Due to the economy's tremendous growth, industrial activity is expanding more frequently, which is causing air pollution to increase more quickly. As a one of the developing industrial nation, our country is producing record amount of pollutants on a daily basis. Need of knowing the AQI of any area is rising day by day. There are various gases available in the atmosphere which are the reason for pollution on our environment specifically like Carbon-dioxide, Nitrogen dioxide, Sulphur oxide, Chlorofluorocarbons, Particulate Matter, etc. and other harmful contaminant of air.

We gathered the information from the database, which includes information on pollution concentrations that occur across the country. We begin by figuring out the individual pollutant index for each data point that is provided and then determining the corresponding AQI for the area. In order to estimate the air quality of India in any given place, we have developed a model that can predict the air quality index for each available data point in the dataset.

With this ml model, different information about the data is collected using different methodologies to determine the regions that are most severely affected in a specific location (cluster). By including the suggested parameter-reducing formulations into our model, we outperformed the traditional regression models in terms of performance. On the currently available dataset for predicting the air quality index for all of India, our model has a 96% accuracy rate.

II. LITERATURE SURVEY

Numerous AQI prediction studies have been conducted using various machine learning techniques and algorithms that can be used by different institutions. This paper gives an overview of some of these studies conducted in research papers using the techniques and results they used.

Madan Tanisha et al. [1] In machine learning, dataset can play a significant role in determining the output. The authors of proposed that Machine Learning algorithms plays important role in measuring air quality index accurately. Logistic regression and auto regression, ANN help in determining the level of PM 2.5. Artificial Neural Network is best among all. However, there still lacks a survey that can analyze and summaries the existing approaches. To bridge up the issue, this survey paper details the task of air quality index prediction using ml by summing up the existing datasets and their collection methods. It analyses the existing feature extraction techniques.

Sriram Krishna Yarragunta et al. [2] In this presented paper we try to conduct a survey on several ML algorithms that are used to solve the problem of air quality index try to see the solutions which are provided by several other methods From the results, concluded that the Random Forest algorithm gives better prediction of air quality index. It also summarizes and analyses the existing techniques critically to identify gaps based on two groups: traditional statistical machine learning and deep learning methods. The experimental results on two benchmark datasets show that the performs about 7% better than the state-of-the-art methods in a mixed domain for the deception dataset with the highest accuracy of 91.2% .

Timothy M. Amado et al. [3] have developed a system that provides better results. In this research paper it introduces some machine learning techniques like Naïve-Bayes, Support Vector Machine and Decision Tree for the prediction of air

quality index using the dataset. In this authors proposed model by using BILSTM which is the Deep Learning model to predicted the PM2.5 with improved performance comparing the existing model and produced exceptional MAE, RMSE. Usha Mahalingam et al. [4] developed a model for predicting AQI using machine learning approaches and used several techniques The proposed method classifies by phase-wise processing. In this paper, we are processing different dataset by using different data mining techniques. This can help people to see the exact result. In this authors used the prediction model results were based on Big Data Analytics and Machine Learning, which have helped to evaluate and contrast current assessments of air quality. Among all the algorithms, the Decision Tree algorithm produced the best results. Murugan Rishanti et al. [5] In today's world, research plays a critical role in decision-making. SVR and the LSTM Machine Learning model were used by the authors. It was demonstrated that SVR algorithms are the most suitable in forecasting air pollutants concentrations among the Machine Learning algorithms used for estimating atmospheric pollutants (PM10 and PM2.5).

III.PROPOSED METHOD

Our proposed methodology includes the following steps:
 i. First I will collect the records of pollutants in the atmosphere.
ii. Then I will perform preprocess operation to extract information from the data.
iii. Then I will perform the cleaning operation to solve the problem of missing value and to give it numerical value.
iv. Then select the attribute from the given dataset.
v. Then there is normalization of data is performed to make it in a range.
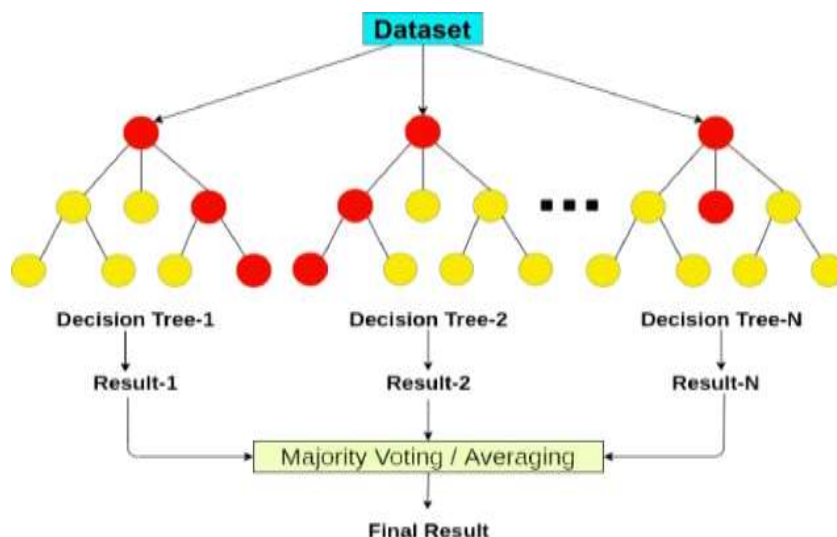vi. The system then displays the results.
The flowchart of the methodology is shown below:

Fig1: [Flowchart of proposed method]

Algorithm Used:

As the name suggests, we use in our Random Forest Regressor for classification and for predicting the result.

The Architecture of Disease Prediction System:

Fig 2: [The architecture of Air Quality Index Prediction System]

IV.CONCLUSION

In our research, we used a Random Forest Regressor Machine Learning algorithm to predict the AQI. And we also tested different algorithms like Linear Regression, Lasso Regression, Decision Tree Regression, KNN Regression etc. Despite testing these algorithms, I have found that the Random Forest Regressor gives higher accuracy than other algorithms. In this research, we found that the potential less accurate result is obtained if there is missing data, but if we can feed the system with one huge amount of data sets, this air quality index prediction can provide up to 96% accuracy.

Collecting a large number of datasets on pollutant is time consuming and cannot be done in a year or two. It takes several years to collect these datasets and to train the system with these data searches. Doctoral students (PhD) can use this system for further research work. With the help of a AQI prediction system, it was possible to save people from polluted air. Components like gases and particulate matter affect the air's quality. When frequently breathed in, these contaminants lower the air quality, which can cause significant ailments. With the use of an air quality index prediction system, it is feasible to detect the presence of these toxics and keep an eye on the environment so that prudent actions can be taken to improve the quality of the air. As a result, there are only fewer health issues left related to air pollution. Machine learning-based prediction models have been found to be more dependable and consistent. Data collection is now straight forward and accurate. Only machine learning (ML) algorithms are capable of handling the meticulous analysis required to provide precise and efficient predictions from such a large body of environmental data. AQI prediction system only provides possible results, it does not guarantee that it will correctly predict the AQI each time, things may differ according to the weather condition. But it is far more accurate in predicting AQI from other systems. In our research we tested the accuracy of this system for 5 different locations value and our accuracy can be up to 96%.

References

[1]. Madan Tanisha, Shrddha Sagar and Deepali Virmani, "Air Quality Prediction using Machine Learning Algorithms-A Review", 2020 2nd International Conference on Advances in Computing Communication Control and Networking (ICACCCN), pp. 140-145, 2020.

[2]. Sriram Krishna Yarragunta and Mohammed Abdul Nabi, "Prediction of Air Pollutants Using Supervised Machine Learning", 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1633-1640, 2021.

[3]. Timothy M. Amado and Jennifer C. Dela Cruz, "Development of machine learning-based predictive models for air quality monitoring and characterization", TENCON 2018–2018 IEEE Region 10 Conference, pp. 0668-0672, 2018.

[4]. Usha Mahalingam, Kirthiga Elangovan, Himanshu Dobhal, Chocko Valliappa, Sindhu Shrestha and Giriprasad Kedam, "A machine learning model for air quality prediction for smart cities", 2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), pp. 452-457, 2019.

[5]. Murugan Rishanti and Naveen Palanichamy, "Smart City Air Quality Prediction using Machine Learning", 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1048-1054, 2021.