IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115199

MACHINE LEARNING APPROACH FOR AQI AND POLLUTANT PREDICTION FOR METROPOLITAN CITIES

Malini R[1], Mallika C[2], Navyashree PN[3], Rukhaiya Badar R[4]

Assistant Professor, Department of Information Science & Engineering, Atria Institute of Technology, Bengaluru, India[1]

Student, Department of Information Science & Engineering, Atria Institute of Technology, Bengaluru, India[2,3,4]

Abstract: The term "air pollution" generally is the process of releasing pollutants into the air which can be harmful to the health of humans and the environment in general. It is one of the greatest challenges humanities has have ever had to face. It can cause harm to crop, animals, and forests, among others. To stop this from happening in the transport sector, it is necessary to identify air quality issues caused by pollution using machine learning techniques. Therefore, air quality assessment and prediction are now an important area of research. The objective is to explore methods based on machine learning to achieve forecasting the air quality of air using predictions that have the highest accuracy. Analysis of data by a supervised machine-learning technique (SMLT) to collect a variety of data points such as, variables identification, univariate analysis bi-variate and multi-variate analyses as well as missing value treatment and examine the data validation as well as data cleaning/preparing, and visualization will be carried out for the entire data set. The analysis we present gives an entire guideline for the analysis of the sensitivity of model parameters with respect to their performance in predicting levels of pollution in the air by accuracy calculations. The aim of this paper is to propose a machine-learning-based method for accurately predicting an accurate Air Quality Index value by predictions in the form of highest accuracy by the comparison of supervised classification machine learning algorithms. Furthermore, to evaluate and analyse the effectiveness of different machine learning algorithms based on the transportation traffic department data with an evaluation classification reports, to identify the confusion matrix, and then categorizing the data according to priority. the outcome shows the efficiency of the proposed machine-learning algorithm method can be evaluated with most accuracy, precision, recall as well as F1 Score.

Keywords: Air Pollution, Air Quality Index, Machine Learning Algorithms, Decision Tree, Support Vector Machine.

I. INTRODUCTION

Modern human life includes energy consumption and its effects. Human-caused sources of air pollution include industrial plants, automobiles, airplanes, burning of straw, coal, and kerosene, as well as aerosol containers. CO, C02, Particulate Matter (PM), N02, S02, 03, NH3, Pb, etc. are among the harmful pollutants released into the air daily. The air pollution can impact the health of humans as well as plants and animals. Air pollution can cause grave illnesses for people, ranging from heart, bronchitis, lung cancer, and on. Air pollution may cause further environmental problems like pollution from acid rains, warming of the planet, reducing visibility, smog and atmospheric pollution and climate change and premature deaths. Scientists have discovered that air pollution can affect the historic monuments in a negative manner.

Air pollution caused by automobiles pollution, factories' emissions, and power plants, as in addition to the exhausts from agricultural processes. can contribute to an increase the greenhouse gas emission. The greenhouse gases adversely affect the climate, and consequently the expansion of plant life [4].

CO2 emissions from carbons inorganic and greenhouse gases also affect the interactions between soil and plants. The consequences of climate change aren't only limited to humans and animals however, the environment of agriculture and productivity are also significantly affected [3].

Economic losses are one of the consequences as well. This is because the Air Quality Index (AQI) an assessment measure that is linked to public health in a direct manner. A higher AQI indicates a higher risk of exposure for the entire population of human beings. So, the need to be aware of the AQI prior to time has spurred researchers to study and forecast the quality of the air. The monitoring and prediction of AQI especially in urban areas is now an important and challenging task due to the rapid pace of motor and industrial development. Research and study work are geared towards developing nations, but there is a significant concentration of harmful pollutants PM 2.5 is found to be more than three times greater in countries that are developing [9].

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022
DOI: 10.17148/IJARCCE.2022.115199

A few researchers attempted to study the prediction of air quality within Indian cities. After studying the literature available, there was a need that has identified to fill this gap by analysing and predict AQI within India. Artificial Machine Learning and artificial intelligence are among the principal methods employed by many start-ups and major platforms. Naive Bayes (NB) is an artificial intelligence classifier that has proven remarkable results even with a tiny amount of training data. While it's true that NB is by far the most effective learning algorithm, it's also the most accurate in the inputs it utilizes to train. Artificial Neural Networks' (ANNs') complexity makes them capable of handling large amounts of data [7].
AQI prediction is a complex process that requires large quantities of data, which is why an ANN is the better choice. Methodologies and tools like Support Vector Machine Logistic Regression, Decision Trees, K-Nearest Neighbour, Random Forest, and Naive Bayes are used to create predictions that is AQI in this research. The remainder portion of the research paper structured according to the following format. In the second section we will review the relevant research. Section 3 explains the issue and Section 4 proposes the suggested design. The section 5 closes out the article, then the discussion on the future research.

II. RELATED WORK

In [11] authors studied a data set that contains parameters of air in relation to the ambient air. Based on this data, a variety of algorithms were employed to determine the rate of emission rate and comparative analysis was conducted.
Authors of [12] obtained historical data on pollutants in the 6 concentrations which affect an air quality indicator. They utilize it as input into an algorithm for neural networks. The authors used the distributed neural network model that incorporates an AQI in the network's distributed structure for short-term forecasts for the AQI.
In [5] researchers tried to predict PM 2.5 pollutant using a bidirectional short-term memory model. This paper is about PM2.5 pollutant. However, there are other pollutants that should be considered in the prediction of air pollution.
In [6] they employed data provided by the department of environmental protection to determine Air Quality Index (AQI) using temperature, wind direction. In the end, AQI Prediction is made with the help of LSTM and analyse the accuracy of the predictions.
Authors of [1] proposed to develop an IoT that is based on monitoring and prediction of air pollution system that is suggested. It can be used to monitor air pollutants within an area, as well as the analysis of air quality may be carried out. Sanjeev (2021) carried studied the dataset and deduced that random forest is better than other algorithm because it is less prone to over-fitting.
In [2] authors used support Vector Regression (SVR) algorithm to analyse the pollutant level in California. The authors claimed they had devised a naïve method to calculate daily air pollutants.
In [8] authors studied twenty literary works in a variety of pollutions that were studied, ML algorithms and their respective performance. The authors found that numerous works utilized meteorological data, like the speed of wind, humidity, and temperature to determine the levels of pollution more precise. They observed that Neural Network and the boosting models performed better than the other top ML algorithms.

III. PROBLEM STATEMENT

Air pollution is an extremely serious issue in large smart cities. The results of certain indicators that measure air quality more than average, and it is a threat to the health of people. The state and local governments have taken certain measures for the prevention of air pollution in urban and rural regions, but they have not completely recovered. The widespread intrusion of water across the Indian cities is expected because of an ongoing western disturbance. When the disturbance in western India results in enough rain, you can expect that National Air Quality Index is likely to be in the poorer levels. Otherwise, a tiny shower or even pre-showers could alter the weather conditions adversely by boosting the AQI. This paper presents a Machine Learning model for prediction of AQI.

IV. PROPOSED DESIGN

A. DATA PRE-PROCESSING

The quality in data collection is the primary and foremost requirement to ensure the effective visualization and development of effective models using ML. The steps that are used to pre-process data aid in reducing the amount of noise in the data. This in turn enhances the processing speed as well as the ability to generalize ML algorithms. The empty values have been filled in with the median values of every feature to address the issue of missing data. Then, a normalisation procedure is used to make the data more uniform, and to ensure that the value of variables is not affected by their sizes or units. Normalization of data helps to bring the various attributes of data to a common size of measurement. This process plays an important role in the stability of the development of ML models and improves performance.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115199

B. EXPLORATORY DATA ANALYSIS

Analysing data in an exploratory manner is the first stage of data analysis that is completed prior to applying any ML model. The most significant elements are analysed: (a) exploring statuses and patterns in air pollution (b) studying the spread in air pollutants across.

C. METHODOLOGY

Fig. 1 shows the model employed for AQI Data Prediction. In the first stage, Machine Learning is used to predict future data the import of data. The data that has been trained is used to train various algorithms in the following stage and the accuracy of the predicted data as well as actual-time data is assessed.

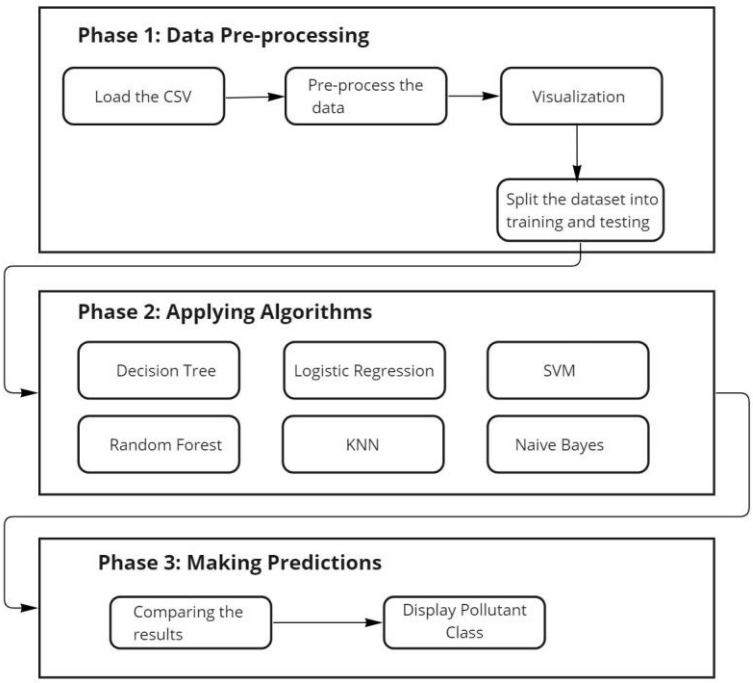

Fig. 1: Methodology of AQI Prediction

1) IMPLEMENTATION OF MACHINE LEARNING (ML) ALGORITHMS:

ML algorithms were applied to the data to determine the AQI's value. To do this the data was separated into two sets: a training set (70 percent) and the test set (30 percent). Accuracy measurements were then made using the forecasted AQI. The average square error was determined for each set to test the accuracy.

2) OPTIMIZATION OF ALGORITHM:

A variety of algorithms were designed to minimize the amount of error and thus increase the accuracy of predictions. Based on this the most efficient algorithm was chosen.

3)PREDICTION OF AQI:

The precision of AQI in Decision Tree is demonstrated in two cases and Logistic Regression in 3. In the Logistic Regression algorithm was also employed and had an accuracy of 98% was observed. The results of the experiments suggest that the highest accuracy is 100% is achievable using the Decision Tree Algorithm.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.115199

Classification report of Decision Tree Classifier Results:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	73
1	1.00	1.00	1.00	175
accuracy			1.00	248

Fig. 2: Decision Tree Classification Report

Classification report of Logistic Regression Results:

	precision	recall	f1-score	support
0	0.97	0.97	0.97	73
1	0.99	0.99	0.99	175
accuracy			0.98	248

Fig. 3: Logistic Regression Classification Report

V. CONCLUSION

In this study, it is proposed that a Machine Learning model for Air Quality Index prediction for smart cities. The model is tested using an examination of Indian cities Air Quality data. Utilizing a variety of methods, the air quality can be accurately predicted by using Decision Tree that has greater accuracy.

REFERENCES

[1] Ayele, T. W. and R. Mehta (2018). Air pollution monitoring and prediction using iot. In 2018 second international conference on inventive communication and computational technologies (ICICCT), pp. 1741–1745. IEEE.
[2] Castelli, M., F. M. Clemente, A. Popovi˘c, S. Silva, and L. Vanneschi (2020). A machine learning approach to predict air quality in california. Complexity 2020.
[3] Fahad, S., O. Sonmez, S. Saud, D. Wang, C. Wu, M. Adnan, and V. Turan (2021a). Climate change and plants: biodiversity, growth and interactions. CRC Press.
[4] Fahad, S., O. Sonmez, S. Saud, D. Wang, C. Wu, M. Adnan, and V. Turan (2021b). Plant growth regulators for climate-smart agriculture. CRC Press.
[5] Jeya, S. and L. Sankari (2020). Air pollution prediction by deep learning model. In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 736–741. IEEE.
[6] Jiao, Y., Z. Wang, and Y. Zhang (2019). Prediction of air quality index based on lstm. In 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), pp. 17–20. IEEE.
[7] Kavitha, P. and M. Usha (2014). Anomaly based intrusion detection in wlan using discrimination algorithm combined with naive bayesian classifier naive bayesian classifier. Journal of Theoretical & Applied Information Technology 62(1).
[8] Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet] 9, 381–386.
[9] Rybarczyk, Y. and R. Zalakeviciute (2021). Assessing the covid-19 impact on air quality: A machine learning approach. Geophysical research letters 48(4), e2020GL091202.
[10] Sanjeev, D. (2021). Implementation of machine learning algorithms for analysis and prediction of air quality. International Journal of Engineering Research & Technology (IJERT) 10(3), 533–538.
[11] Simu, S., V. Turkar, R. Martires, V. Asolkar, S. Monteiro, V. Fernandes, and V. Salgaoncary (2020). Air pollution prediction using machine learning. In 2020 IEEE Bombay Section Signature Conference (IBSSC), pp. 231–236. IEEE.
[12] Wang, W. and S. Yang (2020). Research on air quality forecasting based on big data and neural network. In 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA), pp. 180–184. IEEE.