Vol. 15 No. 3 (2024):292-300
DOI: https://doi.org/10.61841/turcomat.v15i3.14803

PREDICTION OF AIR POLLUTION BY USING MACHINE LEARNING

[1]Mr.A.Damodar Rao,[2]S.Soumya,[3]T.Renuka,[4]Y. Bhargavi
[1]Assistant Professor,[234]Students
Department Of CSE
Malla Reddy Engineering College for Women

ABSTRACT

Defensive and in charge Nowadays, in many developing and urban areas, the greater air quality has become one of the most important factors in everything. The magnificence of the air is negatively affecting collectibles due to the many tainting methods caused by power consumption, transportation, and other factors. Population growth is a major issue in our nation as it is happening at a rapid pace. This, along with economic expansion, is causing environmental issues in cities, such as water and air pollution. in a portion of the air.

Air pollution and pollutants have a direct effect on human health. As is well known, the main sources of pollution include carbon monoxide, nitrogen oxide, particulate matter (PM), so2; etc. A propellant such as gasoline, petroleum, etc. that has not been properly oxidized is producing carbon monoxide. The burning of thermal fuel releases nitrogen oxide (NO), but sulfur dioxide (So2), one of the main air pollutants, is more prevalent and has a greater impact on human health. Multidimensional collisions with location, time, and imprecise boundaries augment the air's dominance. To examine AI-based approaches for air quality prediction is the aim of this enhancement. In this research, we will use a machine learning system to forecast air pollution.

I. INTRODUCTION

The environment describes how everything that occurs in and around it is contaminated by the everyday actions of humans, such as noise and air pollution. The surroundings will naturally become hotter if the humidity rises higher. The main factor contributing to the increase in pollution is the daily growth of transportation and industry, which releases 75% of NO and other gases including CO, SO2, and other particles into the atmosphere. The ever-growing scene, cars, and inventions are threatening the air quality at an alarming pace.

To anticipate the amount of pollution in a certain Delhi zone, we have thus collected certain characteristic data, such as the number of cars and the attributes of pollutants.

Vol. 15 No. 3 (2024):292-300
DOI: https://doi.org/10.61841/turcomat.v15i3.14803

The only thing that surrounds us is the environment. The most serious kind of environmental pollution is caused by natural disasters and human activity. One example of this is air pollution. The climatic factors of temperature, relative humidity, wind direction, and air speed control the amount of air pollutants in ambient air. We feel significantly hotter when there is more humidity because perspiration does not evaporate into the environment. One of the primary causes of air pollution is urbanization, since an increase in transportation facilities releases more pollutants into the atmosphere. Industrialization is another major contributor to air pollution. Particulate matter (PM), carbon monoxide (CO), nitrogen oxide (NO), sulfur dioxide (SO2), and others are the main pollutants. When propellants like gasoline and petroleum don't oxidize properly, carbon monoxide is created. When thermal fuel ignites, nitrogen oxide is produced; carbon monoxide produces headaches and vomiting; smoking produces benzoene, which causes respiratory issues; nitrogen oxides produce nausea and dizziness; and particulate matter with a diameter of 2.5 micrometers or less has a greater negative impact on human health.It is imperative that actions be done to reduce environmental air pollution. The Air Quality Index (AQI) is a tool for assessing air quality. In the past, air quality was predicted using traditional approaches like probability and statistics, but such techniques are quite difficult. Thanks to technological advancements, obtaining information on air pollution levels via sensors is now quite simple. Analyzing the raw data thoroughly is necessary to identify the contaminants. Deep learning, machine learning techniques, recursive neural networks, convolution neural networks, and other neural network models guarantee the ability to forecast future AQI so that appropriate action may be done. We have used a supervised learning strategy in the suggested study. Numerous algorithms fall under the category of supervised learning algorithms, including Random Forest, SVM, kernel SVM, Naive Bayes, Linear Regression, and Nearest Neighbor. Since Random Forest produces superior results than any other algorithms, our method uses it to predict air pollution accurately.

II. LITERATURE SURVEY

"Using multi-source data, we analyze relevance and make short-term predictions of PM 2.5 concentrations in Beijing."

Huang, H.; Du, W.P.; Ni, X.Y.

A significant public crisis has emerged, with the PM2.5 issue commanding widespread public attention and necessitating immediate action. The complicated generation and growth of PM2.5 means that there is

Vol. 15 No. 3 (2024):292-300
DOI: https://doi.org/10.61841/turcomat.v15i3.14803

still a lack of information and predictions about PM2.5 from an atmospheric dynamic theory standpoint. We used multi-source data mining in this article to try to accomplish relevance analysis and short-term prediction of PM2.5 concentrations in Beijing, China. The Multivariate Statistical Analysis method was used to propose a PM2.5 correlation analysis model that would be applied to both physical data (such as regional average rainfall, daily mean temperature, average relative humidity, average wind speed, maximum wind speed, and other pollutant concentration data like CO, NO2, SO2, and PM10) and social media data (microblog data). The research discovered that there is a strong mathematical relationship between PM2.5 concentrations and the following variables: average wind speed, CO, NO2, and PM10 concentrations; the number of microblog comments per day mentioning "Beijing" and "air pollution." The Back Propagation Neural Network (BPNN) model, which is a machine learning tool for huge data, was used to go further into the correlation study. In correlation mining, the BPNN approach proved to be more effective. Finally, this work used an ARIMA Time Series model to investigate the prediction of PM2.5 in the short-term time series. There was a high degree of concordance between the anticipated and actual outcomes. Expanding the utilization of big data and multi-source data mining techniques, this work is helpful for realizing PM2.5 pre-warning, analysis, and monitoring in real-time.

"Air pollution prediction via multi-label classification,"

Scanagatta and Corani,

The likelihood of an air pollutant exceeding a certain threshold may be estimated using a Bayesian network classifier. However, when dealing with stochastically dependent variables, such as ozone measured at several stations or evaluated using different indicators, it is usual to need numerous projections. The conventional wisdom is that each class variable should have its own classifier, yet this method fails to take interdependencies across classes into account. The reliability of the predictions may be enhanced by correctly representing these interdependencies. Developing a multi-label classifier that can forecast several air pollution indicators at once allows us to tackle this issue. With this goal in mind, we use structural learning to acquire the necessary knowledge to construct a Bayesian network-based multi-label classifier. We provide our findings from three separate case studies that tested various approaches to PM2.5 and ozone forecasting. When compared to the independent method, the multi-label classifier provides superior decision-making capabilities.

"Machine Learning for Indian Air Quality Prediction and Analysis".

Vol. 15 No. 3 (2024):292-300
DOI: https://doi.org/10.61841/turcomat.v15i3.14803

The academics involved include Akshaya A.C., Mrs. J. GnanaJeslin M.E., Mrs. A. GnanaSoundariMtech, and a Ph.D.

By predicting the air quality index of a certain place using machine learning, we are able to anticipate the air quality of India. One common way to track pollution levels over time in India is with the use of the air quality index, which takes into account factors including SO2, NO2, RSPM, spm, and more. Based on past data and forecasting for a certain future year, we created a model to forecast the air quality index as a gradient descent boosted multivariable regression issue. Using cost estimation, we enhance the model's efficiency for our prediction problem. Given pollutant concentration data from the past, our model can accurately forecast the air quality index for a whole county, state, or constrained area. We outperformed the industry-standard regression models with our model after using the suggested parameter-reducing formulas. Our model achieves an impressive 96% accuracy rate when it comes to forecasting the air quality index for the whole country of India using the present information. Additionally, we use the AHP MCDM approach to determine the order of preference based on how close it is to the ideal answer. Outliers, BVA, prediction, dataset, preprocessing, and AQI are some of the keywords.

"Artificial Neural Network-Based Future Air Quality Index Recognition".

Hello, Dr. J.R. Prasad and Ruchi Raturi.

Predicting pollution levels for the next day, next month, or next year is necessary since pollution is increasing at a rapid pace. Making use of certain historical data pertaining to the wind. Air pollution, which includes the release of chemicals, particles, or biological elements into the atmosphere, is a growing problem as a result of human activity. It is harmful to both humans and the environment. Indeed, air pollution is a major issue in urban and industrial areas when it comes to environmental issues. Therefore, it is crucial to anticipate pollution and prevent such issues. Among the most intriguing and difficult challenges is the prediction of air pollution using data mining. Air pollution data storage, inventory management, and basic statistic creation are all areas that have been well-served by various systems. While decision support systems are used by certain systems, their capabilities are severely restricted. With this system, you can ask basic questions like "What is the maximum limit of air pollution?" or "which area has a maximum pollution?" but more involved questions like "Predict next month's air pollution count." or "Give me, tomorrow's pollution details" are beyond their capabilities. Predicting air quality using multilayer perceptrons (MLPs), artificial neural networks (ANNs), linear regression, ozone, RSPM, and other related terms is essential.

Vol. 15 No. 3 (2024):292-300
DOI: https://doi.org/10.61841/turcomat.v15i3.14803

Air Pollution Detection and Forecasting with the Use of Machine Learning Techniques

Authors: Aditya C. R., Chandana R. Deshmukh, Nayana D. K., and Praveen Gandhi The ability to see

Governments in both developed and developing nations see air pollution control as a top priority. Air pollution is caused by a combination of variables, including weather, traffic, industrial characteristics (such as emissions from power plants), and the combustion of fossil fuels. More focus should be directed on particulate matter (PM 2.5), one of the several types of particulate matter that affect air quality. At high altitudes, it poses a significant threat to human health. As a result, it's crucial to maintain control by regularly monitoring its airborne level. This work uses logistic regression to determine whether a data sample is contaminated or not. Using the historical PM2.5 data as input, autoregression is used to forecast the future PM2.5 values. If we know the PM2.5 level for the next year, month, or week, we can bring it down to a safe level. In an effort to forecast PM2.5 levels and identify air quality, this method uses a dataset consisting of daily meteorological conditions in a particular city.

III. SYSTEM ANALYSIS AND DESIGN

EXISTING SYSTEM

The Air Quality Index (AQI) is a record that provides the public with information on the level of pollution and its effects on public health, as maintained by the Air Pollution Forecasting System. The many health effects that people may experience as a result of exposure to the toxin concentration over short and extended periods of time are the focus of the AQI. Depending on the national standard for air quality, the AQI values vary from country to country.

A higher AQI score indicates a more significant risk of health concerns. Ultimately, we want our student-made algorithm to be able to predict the hourly contamination focus. In addition, a mobile app for Android will be developed to inform users about the continuous convergence of PM2.5 contamination levels, as well as the hourly predicted value of the toxin fixation based on the student's calculations. Data that is less dirty will likewise be recommended by the Android app [1].

Disadvantages

- The system has not been put into place. A Multiple Linear Regression Approach with Steps.
- The system has not been put into place. Analysis of Non-Linear Data

PROPOSED SYSTEM

Vol. 15 No. 3 (2024):292-300
DOI: https://doi.org/10.61841/turcomat.v15i3.14803

1) Data collection: We used an alternative approach to gather information from a variety of credible sources, such as the Delhi Government website.

In the second phase of the project, known as "exploratory examination," we investigate and test out a number of parameters, including outline ID, consistency check, missing attributes, and more.

3) Data Manipulation Control: During this step, you will need to insert the necessary missing data using the average estimates of that information characteristic. [2] in

4) Using a gauge model to predict boundaries: Maintaining future attributes for distinct borders is necessary for acceptable data indirect relapse.

Fifthly, use straight relapse: when all boundaries are in active or accessible mode, the air quality index (AQI) may be predicted using the direct relapse formula.

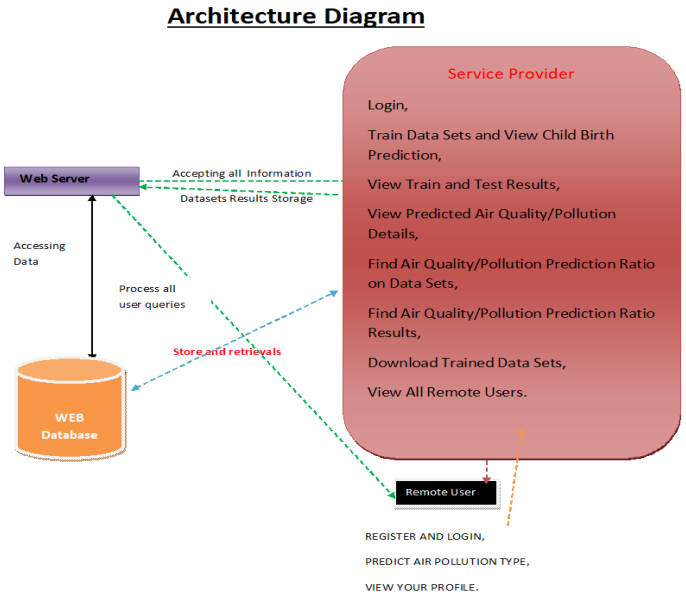
6) Investigating data correctness: We need to determine whether the model we're using is a good match for the data as a whole by comparing the root mean error and the absolute percentage error. Only then can we say that this factor is good for accuracy.

Advantages

- The suggested method put into action Predicting the actual values of data y using continuous parameters is the main purpose of linear regression.
- Continuous data testing and training are both made more successful using the Stepwise Multiple Linear Regression Method.

IV. SYSTEM ARCHITECTURE

Vol. 15 No. 3 (2024):292-300
DOI: https://doi.org/10.61841/turcomat.v15i3.14803



V. SYSTEM IMPLEMENTATION

Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Train Data Sets and View Child Birth Prediction, View Train and Test Results, View Predicted Air Quality/Pollution Details, Find Air Quality/Pollution Prediction Ratio on Data Sets, Find Air Quality/Pollution Prediction Ratio Results, Download Trained Data Sets,

View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to

Vol. 15 No. 3 (2024):292-300
DOI: https://doi.org/10.61841/turcomat.v15i3.14803

login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT AIR POLLUTION TYPE, VIEW YOUR PROFILE.

VI. CONCLUSION

Precision of our model is very acceptable. The anticipated AQI has a precision of 96%. Future upgrades incorporate expanding the extent of district and to incorporate whatever number locales as could be allowed as of now this venture targets foreseeing the AQI estimations of various areas of close by New Delhi. Further, by utilizing information of various urban areas the extent of this venture can be exhausted to anticipate AQI for different urban communities also.

REFERENCES

[1] Ni, X.Y.; Huang, H.; Du, W.P. "Relevance analysis and short-term prediction of PM 2.5 concentrations in Beijing based on multi-source data." Atmos. Environ. 2017, 150, 146-161.

[2] G. Corani and M. Scanagatta, "Air pollution prediction via multi-label classification," Environ. Model. Softw., vol. 80, pp. 259-264,2016.

[3] Mrs. A. GnanaSoundariMtech, (Phd) ,Mrs. J. GnanaJeslin M.E, (Phd), Akshaya A.C. "Indian Air Quality Prediction And Analysis Using Machine Learning". International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 11, 2019 (Special Issue).

[4] Suhasini V. Kottur , Dr. S. S. Mantha. "An Integrated Model Using Artificial Neural Network

[5] RuchiRaturi, Dr. J.R. Prasad ."Recognition Of Future Air Quality Index Using Artificial Neural Network".International Research Journal ofEngineering and Technology (IRJET) .e-ISSN: 2395-0056 p-ISSN: 2395-0072 Volume: 05 Issue: 03 Mar-2018

[6] Aditya C R, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu ." Detection and Prediction of Air Pollution using Machine Learning Models". International Journal o f Engineering Trends and Technology (IJETT) - volume 59 Issue 4 - May 2018

[7] Gaganjot Kaur Kang, Jerry ZeyuGao, Sen Chiao, Shengqiang Lu, and Gang Xie." Air Quality Prediction: Big Data and Machine Learning Approaches". International Journal o f Environmental Science and Development, Vol. 9, No. 1, January 2018

[8] PING-WEI SOH, JIA-WEI CHANG, AND JEN-WEI HUANG," Adaptive Deep Learning-Based Air Quality Prediction Model Using the Most Relevant Spatial-Temporal Relations," IEEE ACCESSJuly 30, 2018.Digital Object Identifier10.1109/ACCESS.2018.2849820.

Vol. 15 No. 3 (2024):292-300
DOI: https://doi.org/10.61841/turcomat.v15i3.14803

[9] GaganjotKaur Kang, Jerry Zeyu Gao, Sen Chiao, Shengqiang Lu, and Gang Xie,"Air Quality Prediction: Big Data and Machine Learning Approaches," International Journal of Environmental Science and Development, Vol. 9, No. 1, January2018.

[10] Haripriya Ayyalasomayajula, Edgar Gabriel, Peggy Lindner and Daniel Price, "Air Quality Simulations using Big Data Programming Models," IEEE Second International Conference on Big Data Computing Serviceand Applications,2016.