INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 4 - V7I4-1499)
Available online at: https://www.ijariit.com

Forecasting air quality using Machine Learning techniques

Rajeena R. R.
rajeenabangalore@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

Dr. Usha S
sakthivelusha@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

Shalini.G
shalini.jg12@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

Sharanya B
sharanyab1999@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

Rekha S
rekhasriram308@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

ABSTRACT

Air pollution, in general, implies to discharge of impurities into atmosphere which are harmful in human life and in universe. It has the potential to be most destructive menace humanity has always faced. It gives problem to animals, birds, yield, crops and forest among other things. To overcome the difficulty, machine learning techniques must be used to anticipate air quality from pollutants. As a result, forecasting quality of air monitoring and prediction of it has been a significant study area. The actual objective is looking for a machine learning-based solutions for forecasting air quality with the highest level of accuracy. The entire dataset will be analysed using machine learning technique to calculate large bits of data which are variable identification, uni-variate analysis and multi- variate analysis, also bi-variate analysis, missing value treatments, and data validation, data cleaning/preparing, and visualization. Our paper offers a detailed guide to model parameter risk assessment in terms of performance in forecasting air quality pollution calculating prediction accuracy. For presenting a technique using machine learning for reliably predicting the Air Quality Index value by comparing various classifying machine learning methods and producing calculation results in the type of highest accuracy. In addition, the outputs of several machine learning methods from the provided datasets will be compared and discussed, along with estimation of user interface which is GUI based for forecasting air quality using attributes present in the dataset.

Keywords: *Machine Learning Algorithms, Pollutant, Air Quality Index, Datasets, Data Visualization, Accuracy, Air Pollution, Prediction.*

1. INTRODUCTION

The contaminants entering to the air environment which can cause problems to healthy human and also to air environment is known to as air pollution. particle matter (PM), Ozone (O3), sulfur dioxide (SO2), nitrogen oxides (NO), carbon monoxide (CO), volatile organic compounds (VOCs), fertilizers, pesticides, and metals too are pollutants present in outdoor. Increased levels of polluted air has been linked with higherrates of death and morbidity. The particle size plays a big role in deciding where particles go into respiratory system. PM2.5, which refers to pollutants with diameter are lesser than 2.5m or equals to 2.5 m, is a growing source of concern since many particles which can be put down in alveoli, the lungs exchange zone of gas. According to 2016 data from a research, minimum 140 million Indian people breatheair which is 10 times more polluted air which is WHO acceptable guideline, and in the world 13 out of 20 city with the best yearly Indian pollution levels. Economic pollution contributes 51% of pollution, whereas automobiles contribute around 27%, burning the crops accounts to 17%, and fireworks for 5% respectively.Two million Indians die prematurely each year as a result of air pollution. Emissions come from automobiles and industry, but biomass burning for cooking and heating is a major source of pollution in rural regions. Forecasting Air pollution which has a number of negative consequences in the society, including soil depletion, affects to wildlife habitat, depletion of ozone layer, many crop and forest are damaging, and universal climate alterates. Air pollution is causing a manmade phenomenon in environmentknown to as Global warming. As it means global warming of the land and sea temperature. The temperature is raising at the minimum of highest due to a growth in greenhouse gases, which generate heat as more of Earths heat escapes into space it increases heat in Earth's atmosphere.

The popular greenhouse gas is Carbon dioxide that has highest effects on global warming. When fuels are burned, many gases are released in atmosphere which one among them is carbon dioxide. Humans are been confident that fossil-fuels are necessary to run automobiles and flights smoothly, to operate factories. CO2 is released into the atmosphere as a result of doing this. Methane,

laughing gas, and fluoridated gases are some of the greenhouse gas which are generated by natural and other artificial resources. Coal plants and agricultural operations may emit a lot of methane. Nitrous oxide is a common byproduct of industrial processes, agricultural.

The Fluorinated gases which are emitted by industries, like hydrofluorocarbons. Fluoridated gas are often utilized rather than Chlorofluorocarbons (CFC's) gases. Major nations have banned Chlorofluorocarbons considering they destroy the layers in ozone. World Health Organization established updated Air Quality Guidelines in 2006. The WHO's recommendations are more stringent than those of most individual countries. According to recent studies, machine learning appears to outperform the other two strategies for anticipating pollution. It's becoming more common to forecast air quality. The data mining differs only not just from one to the next but Depending on the classification algorithms, one thing leads to another, butalso in terms of the features used. Some of them take into account a long list of meteorological factors, while others make a cautious selection or don't use climatic data at all. The suggested system can forecast air quality indexes by assessing various types of air pollutants existing in the air and applying various machine learning methods.

2. EXISTING SYSTEM

To forecast PM2.5 concentrations, three key methodologies are now used: statistic models, electro- chemical transports, and the machine learning. Different climatic elements (wind, precipitation, and temperature) exhibit a negative connection with PM concentrations, according to statistical models made on linear regression using single variable(PM10, PM2.5). Chemical transports and the atmosphere spreadsystem are both numeric approaches, hence WRF-Chem and CMAQ are the most sophisticated. These models are frequently used to forecast pollution of air, but their effeciency is dependent to up-to-date resource list, which is harder to obtain. Addition to this, due to the intensity of airflow (wind speed and direction) surrounding environmental features, complex geophysical structures of complex areas interfere with the implementation of these climates and predict pollution. A machine learning approach, unlike a pure statistical procedure, can take multiple attributes during the single system-model. Artificial Neural Networks are the most used classifiers for forecasting pollution from meteorological data. Other successful research include hybrid models, which incorporate numerous Artificial Intelligence technologies, like logics which are downy and neural networks, component principle analysis and with support vector machines, numeric methods, and machine learning.

3. OBSERVATION

The process of observing the outcomes during the practical period is called observation. In this study, wewill examine the results that we acquired during the experiment and what errors and accuracies we drew from them. The wide range and the major distribution of attributes in an input data are important to many machine learning techniques. Input data contains outliers which can mislead and skew machine learning algorithms' in modelling processes, outcome in prolonged training time, which has less accuracy models, and eventually poorer outcome. In applied statistics and machine learning, data perception is an important skill. Statistics is concerned with quantitative data descriptions and estimations. Data visualization is a valuable set consisting instruments to acquire a qualitative knowledge of data. It is useful for evaluating and learning about the datasets, as well as spotting patterns, corrupt data, outliers, and other anomalies.

Visualization of data are utilized to indicate and demonstrate crucial links in graphs and charts which is more intuitive instead estimates association with a little topic-knowledge. Visualization of data and investigational data analysis are the disciplines in themselves, and it will be recommended that you read some of the books indicated at the end for further information.

4. RESULT AND DISCUSSION

Machine learning methods for AQI prediction obtain greater than sufficient accuracy after effective training of the dataset using so2, no2, PM10, PM 2.5 , and ozone like pollutants data. Following a successful prediction, the predicted value for each algorithm was compared individually, and the predicted value was compared to the original value. Prior to the prediction value, the graph in the analysis section was shown, which included the effect of various pollutants on the air quality index. The accuracy has been demonstrated in both training and testing, indicating that there is no overfitting.

And after all the testing, we can observe that the decision tree produces the most efficient outcomes, since its accuracy and mistakes are superior to any of the algorithms we tested on this datasets. As a consequence of comparing all of the methods, we can conclude that the machine learning decision tree produced the best outcome.

5. CONCLUSION

Before prediction, the data is evaluated and preprocessed in this paper. Different machine learning techniques are employed to estimate and forecast the city's air quality. The expected and forecasted data are plotted using the matplot package. Finally, the following conclusion can be drawn: Using the algorithms outlined, a model for forecasting AQI for a specific city is trained and evaluated with various input factor. With the help of the matplot toolkit, some graphs are displayed as a graphical interface for the expected and forecasted data for all inputs.

6. REFERENCES

[1] Rahul Ahuja, Ishan Verma, Lipika Dey and Hardik Meisheri "Air pollutant severity predictions using Bi-directional LSTM Network" in IEEE 2018, New-Delhi, India.
[2] Rutvik Mehta, Temesegan Walelign Ayele "Air Pollution monitoring and the Prediction using IOT" Published by IEEE 2018 Second International Conference.
[3] Patricia Smith-Wills, Evrin Fenyves, William Rea, and Luke Curtis "Adverse health effects of outdoor air pollutants", Volume-

32, Issue-6, United states received 6th January 2006, Accepted 24 March 2006.
[4] Abdullah Kadari, Khaled Bashir Shaban and Eman Rezk "Urban Air Pollution Monitoring Systrem with Forecasting Models", Published by IEEE 2016 Sensors Journal, Volume-16, Issue-8.
[5] N Jaiswal, P Goyal and AT-Chan, "Statisytical models for Predictability of suspended respiratory system in the urban areas ", Atmospheric Environment, pp. 2068- 2077, vol 40, no.11,2006.

