INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 05, May 2023, pp : 1366-1368

e-ISSN : 2583-1062

Impact Factor : 5.725

FORECASTING OF AIR QUALITY USING MULTICLASSIFIER APPROACH

Pratik Avhad[1], Dhananjay Kale[2], Saurabh Zanje[3], Yash Thapa[4], Unde S. P[5]

[1,2,3,4]Student, Department of Computer Engineering, RGCOE, Ahmednagar, India

[5]Assistant Professor, Department of Computer Engineering, RGCOE, Ahmednagar, India

ABSTRACT

An key environmental risk factor for the spread of diseases including lung cancer, autism, asthma, low birth weight, etc. is air pollution. In order to protect the health and wellbeing of its citizens, governments in emerging nations must regulate air quality. The amount of air pollution varies from location to location and depends on a variety of factors, including burning fossil fuels, heavy traffic congestion, industrial pollutants, and other factors. In practically all industrial and urban regions nowadays, analyzing and safeguarding air quality has become one of the most important tasks for the government. In this study, the concentrations of air pollutants such as SO2, NOx, PM2.5, O3, and PM10 are examined using machine learning techniques. In order to effectively extract features and make decisions, this research examines the air quality based on various pollutant concentrations during pre-covid and post-covid days. Based on historical air quality data, a machine learning model is created using logistic regression and decision trees to forecast the air quality index. The results of the experiments demonstrate the effectiveness of the proposed model in determining air quality and forecasting future pollution levels.

Keywords: air quality monitoring; machine learning; air quality index, concentration,

correlations, pollution.

1. INTRODUCTION

One of the adverse effects of pollutants released into the air is a decline in air quality. Over the past few decades, there have also been more negative effects such acid rain, global warming, aerosol production, and photochemical smog. Numerous researchers are looking at the underlying pollution-related factors causing COVID-19 pandemics in various countries as a result of the recent rapid spread of COVID-19. Air pollution has been associated to significantly higher COVID-19 death rates, and patterns in COVID-19 death rates resemble patterns in places with both high PM2.5 exposure and high population density. In order to help communities and individuals more effectively reduce the detrimental effects of air pollution, it is vital to foresee and plan for pollution oscillations. To do this, monitoring and regulating air pollution heavily relies on air quality evaluation. Both naturally occurring and man-made particles can exist.Examples include ash, sea spray, and dust. Particulate matter (including soot) is released during the burning of solid and liquid fuels, such as when producing electricity, heating a home, or running a car. The size of a particle, or its diameter or width, varies in particulate matter. PM2.5 is the term used to describe the mass of airborne particles with a diameter of less than 2.5 micrometres (m) per cubic metre of air[13]. PM2.5 (2.5 micrometres or one-hundredth of a millimeter) is also referred to as fine particulate matter. Because fine particulate matter (PM2.5) poses a serious threat to public health when airborne levels are elevated, it is an important component of the pollutant index [1].

2. LITERATURE REVIEW

Radhika M.Patil el.et author represented the Air Quality Index (AQI), which indicates whether or not the air around us is polluted, is the subject of this research review paper. It is crucial to understand the AQI because without understanding the worst effects or dangers of air pollution, people won't be as concerned about it and less likely to take action to lessen it. According to this review, the majority of researchers have focused their research on forecasting AQI and pollutants concentration levels, which will provide an accurate picture of AQI. For the prediction of AQI and air pollutants concentration, several researchers opt for Artificial Neural Networks (ANN), Linear Regression, and Logistic Regression [1].

Aditya C R el.et. author proposed that in the suggested system that will make it easier for regular people and meteorologists to identify and forecast pollution levels and take the appropriate measures accordingly. Additionally, this will assist in creating a data source for tiny communities, which are typically overlooked in favour of large metropolis[2].

Yun-Chia Liang el.et. author proposed that the use of artificial intelligence techniques yields encouraging outcomes for AQI forecasting. This study used information gathered over an 11-year period by Taiwan's EPA and CWB. Three regions of Taiwan were taken into consideration including two locations known for having poor air quality year-round Stacking ensemble and AdaBoost provide the best performance of target predictions based on three separate datasets, with good results for R2. To be more precise, AdaBoost produces the best MAE results, whereas the stacking ensemble produces the best RMSE results. All findings indicate that SVM produces the worst outcomes of all investigated approaches and only offers useful outcomes for 1-h predictions [3].

In this article, we made the new dataset AirNet available to researchers who wish to examine air quality using deep learning techniques. It contains 6 air quality indicators from 1498 monitoring stations, which is at least 40 times greater than most prior datasets compared to other research in the field. Additionally, utilizing the AirNet dataset, we established the baseline method Wipe Net, which produced a CSI score of 0.56 and a 16% point improvement over traditional LSTM techniques [4].

RM Fernando el.et. Air is essential to human survival on a fundamental level. The value that qualitatively represents the state of air quality is the air quality index, or AQI for short. The threat to human health and the environment increases with the air quality index. Human activities are always to blame for air pollution. Poor air quality in Sri Lanka is a major concern, particularly in cities like Colombo and Kandy. The potential for health problems brought on by air pollution will be reduced by accurate air quality forecasting [5].

3. MACHINE LEARNING PREDICTION METHODS

1. Support Vector Machine

The hyperplane that serves as a boundary between different data points is created by the support vector machine, a supervised learning technique for classification, regression, and outlier detection, from which the output can then be inferred, two distinct SVM versions are displayed. The data points closest to the hyperplanes on the classification problem in Figure 1a are regarded as support vectors. The distance separating the classes is the area between these two areas.

2. Randomforest

Random forest is a well-known supervised learning ensemble approach that combines numerous decision trees to create a forest and the bagging concept to introduce randomization into the model-building process. The individual tree is divided using a random selection of features, and each decision tree's training data subset is made using a random selection of instances. The variable from the random number of features is taken into consideration for the optimal split at each decision node in every tree. The most frequent prediction will be used by random forests if the target attribute is categorical. If it is numerical, however, the average of all projections will be used.

3. Adaptive Boosting

The following technique, adaptive boosting, likewise derives from a subset of ensemble methods that combines a number of weak learners but arranges them sequentially rather than in parallel as random forest does. The base models are trained sequentially, one at a time, and the classifiers are given weights based on how well they predict a random sample of input examples. By doing this, the more precise classifiers will contribute more to the final result. Additionally, weights are assigned to each input item according on how challenging it is for all classifiers together to correctly forecast the instance.

4. Artificial Neural Network

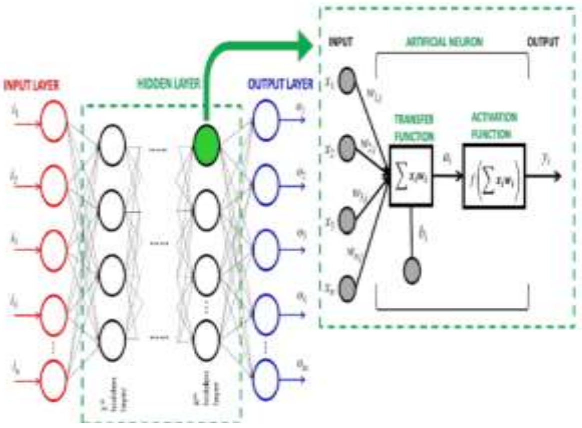


Fig 1. artificial neural network.

The artificial neural network is the following strategy that is favoured in this investigation. Being the first algorithm created, ANN is not only recognized as the "universal approximate" that can accurately estimate any arbitrary function [16], , known as deep learning or deep neural network. In the process of learning new knowledge, the neural network simulates the structure and networks of the human brain.

5. Linear regression

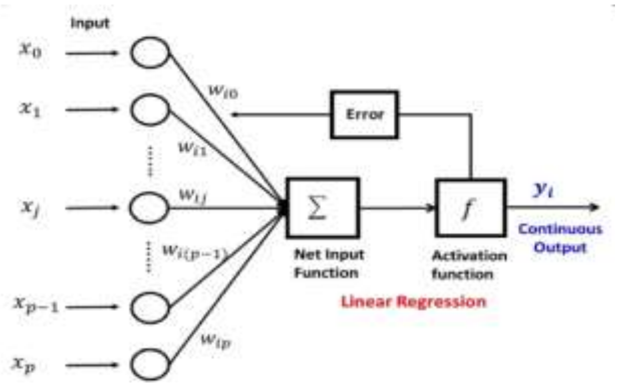
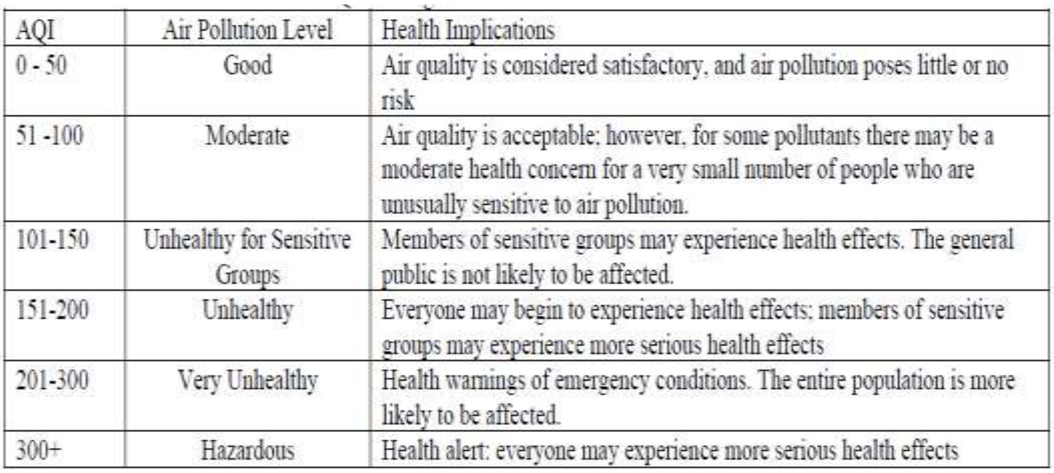


Fig 2. linear regression’s learning process.

Most academicians likely began their initial experiences with machine learning with linear regression. The fitting of one or more independent variables and the dependent variable into a line in n dimensions constitutes the primary working principle. The term "n" typically refers to a dataset's total number of variables. When trying to fit all the occurrences into the line, this line is allegedly constructed to minimize the overall errors. Linear regression is capable of learning constantly under machine learning by refining the model's parameters. These variables include w0, w1, w2, and wm. The technique most frequently used for optimization is gradient descent.

AQI – Range Classification



AQI	Air Pollution Level	Health Implications
0 - 50	Good	Air quality is considered satisfactory, and air pollution poses little or no risk
51 -100	Moderate	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.
101-150	Unhealthy for Sensitive Groups	Members of sensitive groups may experience health effects. The general public is not likely to be affected.
151-200	Unhealthy	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects
201-300	Very Unhealthy	Health warnings of emergency conditions. The entire population is more likely to be affected.
300+	Hazardous	Health alert: everyone may experience more serious health effects

4. CONCLUSION

Nearly 15000 data recordings and 12 air pollution concentrations and meteorological variables, including solar radiation, relative humidity, average temperature, wind direction, wind speed, O3, CO, NO2, SO2, PM2.5, and PM10, were collected for this investigation. The correlation matrix for Figure 2 shows that PM2.5 and PM10 have the highest correlation among themselves and with one another. This correlation matrix was calculated using R.When compared to other air pollution and meteorological factors, the correlation matrix shows that PM10, SO2, NO2, and CO have the highest association with PM2.5. Therefore, we used the parameters of PM2.5, PM10, NO2, SO2, and CO to train the prediction model.

5. REFERENCES

[1] Radhika M.Patil “A Literature Review on Prediction of Air Quality Index and Forecasting Ambient Air Pollutants using Machine Learning Algorithms” ISSN No:-2456-2165 Volume 5, Issue 8, August – 2020.

[2] Aditya C R, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu “Detection and Prediction of Air Pollution using Machine Learning Models” ISSN: 2231 – 5381 http://www.ijettjournal.org Page 204.

[3] Yun-Chia Liang , Yona Maimury , Angela Hsiang-Ling Chen ,and Josue Rodolfo Cuevas Juarez “Machine Learning-Based Prediction of Air Quality” Appl. Sci. 2020, 10, 9151; doi:10.3390/app10249151

[4] Airnet: A Machine Learning Dataset For Air Quality Forecasting” Under review as a conference paper at ICLR 2018.

[5] RM Fernando, WMKS Ilmini, and DU Vidanagama ”Air Quality Prediction Using Machine Learning” #35-CS-18-0001@kdu.ac.lk