IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISN NO: 2347 – 3657

Volume 10, Issue 1, 2022

Prediction and Analysis of Air Particulate matter in Delhi

Che Shu, Ma Jiawe, Han Qi,

Drum Tower Hospital of Traditional Chinese Medicine,
Department of Traditional Chinese Medicine General Hospital of the Air Force PLA,
Department of Dalian Medical University, Dalian

Abstract. Human health has become a serious concern due to the rise in air pollution. In order to make informed decisions on air pollution management, air pollution analysis and forecast are critical. Pollutants smaller than 2.5 micrometres (PM2.5) are the primary indicator of air quality in a region. In this study, we used a variety of machine learning methods to construct a comprehensive model for predicting Delhi's air quality. Air quality levels were predicted using the Historic meteorological data which covers seven meteorological factors including wind speed, wind direction, solar radiation, ambient temperature, relative humidity, and PM2.5. Models for predicting PM2.5 levels are examined, and the MLP is shown to be the most accurate.

.Keywords: **Analysis of Particulate Matter, Naive Bayes, SVM, Logistic Regression, Correlation, and Quality Index**

Introduction

PM2.5 is the most harmful pollutant to human health, causing emphysema, lung damage, bronchitis, and asthma, among other ailments. Around 627,000 Indians die each year as a result of pollution, according to the WHO's Global Burden of Disease research from 2014 [7]. PM2.5 has been shown to be Delhi's main source of pollution. The government of India use an air quality index (AQI) to standardise the measurement of air quality. As the AQI rises, so does the pollution in the surrounding area, which has a number of harmful effects on human health, especially in children. As a result, forecasting air quality in advance has become more difficult, making it more difficult for government agencies to take action to reduce pollution before it becomes more dangerous. In this study, we used several machine learning methodologies, such as calculating the correlation coefficient between meteorological parameters and pollution levels, to investigate the relationship between these two variables. Then, using machine learning methods, we examine each parameter in relation to the pollution level and construct a prediction model for predicting the PM2.5 level in Delhi.

Types of Pollution

Air pollution may be divided into two categories:

Primary Pollutant

Pollutants emitted directly through the processes of fossil fuel use, volcanic eruptions, and manufacturing are considered primary pollutants. Sulfur dioxide, nitrogen dioxide, carbon dioxide, Particulate Matter, Methane, Ammonia, Chlorofluorocarbons, etc. are among the most common main pollutants. Coughing, wheezing, a tightening sensation in the chest, and other symptoms that indicate a high concentration in the air are all examples of airborne pollutants that may be hazardous to human health.

Secondary Pollutant

ISN NO: 2347 – 3657
Volume 10, Issue 1, 2022

A secondary pollutant is one that is formed in the atmosphere as a result of the reactions of other pollutants, such as primary pollutants, rather than being released directly into the atmosphere. When hydrocarbons and nitrogen oxides (NOx) mix in the presence of sunlight, they generate ozone. When NO2 reacts with oxygen in the air, it becomes acid rain; and when sulphur dioxide or nitrogen oxides react with water, they form acid rain.

Different Sources of Air Pollution: In addition to vehicular emissions, industrial emissions, and construction and demolition, there are many other causes of air pollution. You'll find them listed below.

Vehicular Emission: Because urban traffic, which includes commercial vehicles, private automobiles, two- and three-wheelers, buses, and heavy-duty trucks, is increasing at an alarming rate, vehicles are seen as a significant source of pollution in the atmosphere.

Industrial Emission: New factories are being built at a rapid pace to keep pace with the rapid advancement of new technology. Air quality has been harmed as a result of this by generating hazardous smoke, gases, and other pollutants.

Road dust, Construction and Demolition: Particulate pollution is thought to be exacerbated by urban activities such as sweeping up road dust and building new roads.

Related work:

With more and more individuals working on Air Pollution Analysis and Forecasting every single day, the government will be able to regulate the levels of Air Pollution more effectively in the future. [1] uses machine learning to forecast the PM10 level in Delhi based on nine different meteorological parameters. There were three models tried to forecast Delhi PM10: Nave Bayes, Support Vector Machine, and Multilayer Perceptron. MLP was shown to be the most accurate. In [2], they use the WRE-pollution, model's weather, and chemical component forecasts to improve the predictive model's performance. Machine learning methods and various feature combinations are used to create a complete model for predicting pollution levels in many Chinese cities. Proposed a new grid-based mobile source emission inventory utilising the Ive model that recognises automotive pollution as a major source of air pollutants, which negatively affects the overall quality of the air. [3]. HazeEst is a machine learning model that utilises sparse and fixed data to predict Sydney's air quality for a certain time period on a specific day. Fixed station data is sparse, but mobile sensor data is abundant. Five air pollutants (SO2,CO,NO2,NO3,O3) and their hourly values are predicted in [5] for up to eight hours in advance in the Bilbao region (spain). Based on historical data, they created 100 models using various kinds of neural networks and then selected the model with the highest degree of confidence as the best one.

III Proposed Work:

The dataset for this experiment is made up of 640 occurrences of each of the eight meteorological parameters at Delhi's Anand Vihar station between August 23, 2015, and August 23, 2017. Naive Bayes, Logistic regression, support vector machine, and Multilayer perceptron are some of the machine learning methods we used to analyse Delhi's air quality.

Dataset

Central pollution control board (CPCB) data was used in this experiment[6]. Under the Ministry of Environment and Forests, the Central Pollution Control Board (CPCB) monitors air quality in India. For the last many years, CPCB has kept a record of every meteorological parameter. For this experiment, we've simply collected seven weather variables for the Anand Vihar station. Wind speed, wind direction, ambient temperature, relative humidity, solar radiation, and PM 2.5 levels were among the parameters considered.

ISN NO: 2347 – 3657
Volume 10, Issue 1, 2022

A training set and a test set are derived from the whole dataset. The training set has 479 cases, whereas the test set includes 120 instances. This is a representation of the meteorological parameter and its accompanying PM2.5 levels and Label.

Ambient Temperature: It is the air temperature around you that determines the ambient temperature. The air temperature in a room is what we're talking about here. In certain cases, the air temperature in a room differs significantly from that in a room's temperature reading. The ambient temperature is much lower or higher than the room temperature if the space is unbearably chilly or heated.

Wind Direction: Wind Pollutants migrate from one location to another based on their direction of travel.

Average Wind Speed: Pollutants are diluted to a large extent by wind speed. Pollutants in the air are more likely to be dispersed by high winds, while pollutants tend to grow in areas with weak breezes.

Relative Humidity: If the air were more concentrated, the percentage of atmospheric moisture in the air would be higher than it is now.

Atmospheric Pressure: As a rule of thumb, atmospheric pressure and PM 2.5 are linked. AP and PM 2.5 Pollutant have a positive association; hence, a region with a high AP is more likely to be polluted than one with a low AP.

Solar Radiation: Solar radiation is the term used to describe the many forms of solar energy and light that are emitted by the sun. The temperature of a planet is influenced by the amount of solar energy it receives.

Particulate Matter 2.5: It is made up of all the airborne solid and liquid particles. Dust, smoke, and liquid are all components of this strange brew, which includes both biological and inorganic constituents. PM2.5 and PM10 are the two types of PM2.5 particles, which have a range in size from 2.5m to 10 micrometres. "High" (> 60 ug/m3) and "low" (60 ug/m3) PM2.5 concentrations were identified.

IV Prediction Models:

Four Machine Learning approaches have been utilised to create a predictor: Naive Bayes Classifier, K Nearest Neighbor, and Multilayer Perceptron (MLP). Those that illustrate the association between PM 2.5 levels and the input parameter. There were two categories of PM2.5 concentrations: "high" (60 ug/m3) and "low" (less than 60 ug/m3). It has a mild pollution threshold of 60 micrograms per cubic metre. Machine Learning models were generalised by doing a 10 fold cross validation exercise. The Multilayer Perceptron model is also implemented using the Keras library with tensor flow as the backend.

Support Vector Machine: It is possible to analyse data using regression or classification by utilising SVM models, which are supervised learning techniques. The marginal hyper plane is used to classify all of the data points in an SVM model's space representation. Classification results are more accurate when plane margins are larger

International Journal of Information Technology & Computer Engineering

ISN NO: 2347 – 3657

Volume 10, Issue 1, 2022

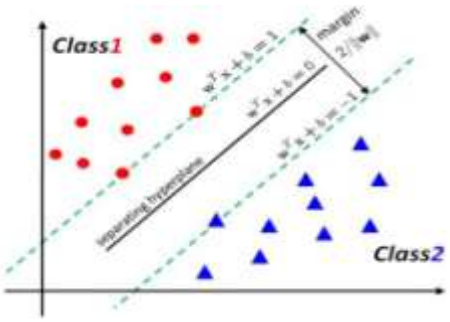
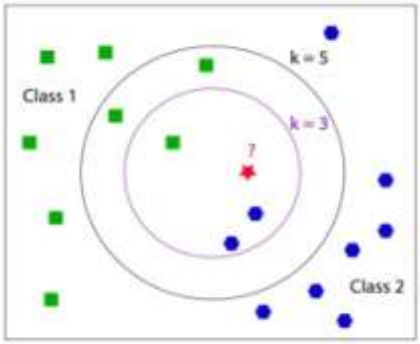


Fig. 2. SVM Classifier

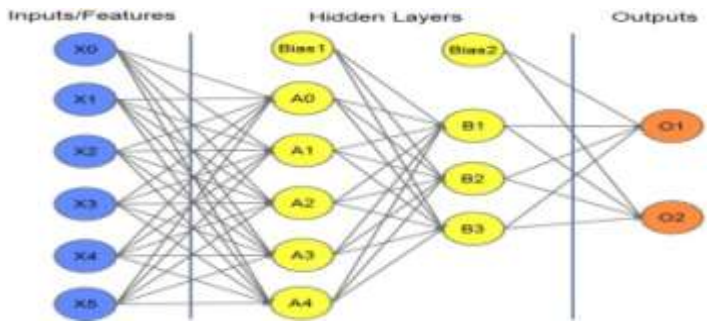
To identify the class that maximises P(Ci |X), m is the number of classes. For example, the probability of P(Ci | X) is computed by using the Bayes theorem (X).

K-Nearest Neighbors Method: An method known as K Nearest Neighbor (KNN) is often utilised in both classification and regression problems. Because it does not assume anything about the underlying data distribution, KNN is a non-parametric learning method. It also uses training data points to generalise only when necessary.

Fig. 3. KNN Classifier



Multilayer Perceptron: Classification and regression problems may be solved using this supervised method. An artificial neural network with several hidden layers, such as MLP, uses input data from one layer and output data from that layer to feed into input data from a subsequent layer, and so on. To develop the network, it makes use of the back propagation method.



International Journal of Information Technology & Computer Engineering

ISN NO: 2347 – 3657

Volume 10, Issue 1, 2022

Fig. 4. Multilayer Perceptron Architecture

Correlation: Correlation is a statistical measure of the connection between two or more variables. Many correlation methods exist, but we utilised the most often used one — Pearson Coefficient — to discover the association between PM 2.5 and each input variable.

V. **Analysis of Data and Correlation**: Pollutants are dispersed in large part due to the speed of the wind. Mean wind speed and temperature both have an adverse effect on PM2.5 concentrations; as these variables rise, so does PM2.5 concentration. Covariance matrix shows that the correlation coefficient is -0.375 as seen in figure 5. (a). In this study, we sought to determine the impact of a meteorological parameter on the PM 2.5 particle size distribution.

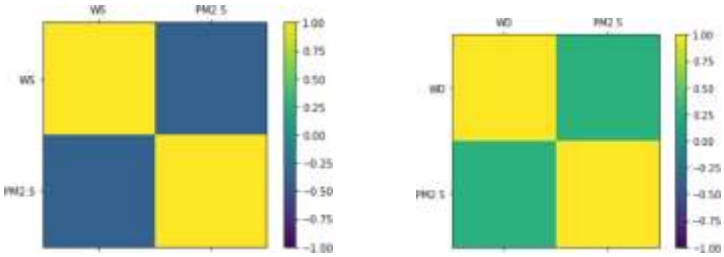


Fig. 5(a) **Fig. 5(b)**

The utmost negativity The wind and PM2.5 concentrations are shown to be correlated. Effort and maximum beneficial impact. Windy conditions provide a correlation. The correlation coefficient for this direction is 0.275. Figure 5 shows the correlation matrix between PM2.5 and wind direction (b).

VI. All seven parameters were correlated with PM2.5 in this study, and those that had a weak or non-existent relationship were eliminated. PM2.5 has just a modest relationship with Atmospheric Pressure and Relative Humidity. As a result, these factors were omitted, and just five parameters were taken into account.

VII. Results:

Accuracy is one of several metrics that can be used to evaluate the success of a classification-based prediction model, but it's not the only one; we've also computed precision, recall, and f-measure values to get a more accurate picture of the model's performance, as well.

Technique	MLP	SVM	Naïve Bayes	KNN
Accuracy	0.983	0.81	0.93	0.95
Precision	0.98	0.67	0.94	0.96
	0.98	0.82	0.93	0.96

ISN NO: 2347 – 3657

Volume 10, Issue 1, 2022

F1-measure	0.98	0.73	0.93	0.96

Table 1. Results obtained from the various methods

VIII. **Conclusion:** We used a variety of Machine Learning techniques to examine the impact of various climatic variables on Delhi's pollution levels. Our experiments have demonstrated that the Multilayer Perceptron model has the highest accuracy of all of the models tested, at 98.33 percent. Multilayer Perceptron networks have been utilised to obtain the greatest accuracy in showing the influence of meteorological parameters on the PM2.5 level and thereby forecasting pollution levels.. We have used numerous networks of Multilayer Perceptron Thus, we may conclude that MLP is the most accurate method for predicting the PM 2.5 concentration.

IX. References:

[1] In "Prediction and Analysis of Pollution Levels in Delhi Using Multilayer Perceptron," Springer publication vol 542, A.Aly, M. Sarfaraz, G. Chaitanya, and M. Adil (2017).

[2] "Prediction Improvement by a Machine Learning" IEEE International Conference on Service Operations and Logistics and Informatics, X.Xia, W.Zhao, X.Rui, Y.Wang, X.Bai, W.Yin, J.Don (2015)

[3] This year's Springerplus

[4] "Machine Learning Based Metropolitan Air Pollution Estimation From Fixed and Mobile Sensors," IEEE Sensors Journal, Volume 17, Number 1, 2017, by H. Ke, R. Ashfaqur, B. Hari, and S. Vijay.

[5] "Short-Term Prediction of Air Pollution Levels using Neural Networks," ACTEA 2009.

[6] www.cpcb.gov.in is the official website of the Central Pollution Control Board.

[7] There are a number of resources on outdoor air quality from the World Health Organization and the Central Pollution Control Board (CPCB) that may be found at http://www.who.int/phe/health topics/outdoorair/databases/ cities/en.