MACHINE LEARNING ALGORITHMS TO ANALYZE HISTORICAL AIR POLLUTION DATA AND PREDICT THE AIRQUALITY

P. ANUSHA, Student, Department of CSE, Vikas College of Engineering & Technology, AP, India

Mrs. M. PRAMEELA, Assistant Professor, Department of CSE, Vikas College of Engineering & Technology, AP, India

***Abstract*—** Controlling and defensive the higher air greatness has gotten one in everything about first imperative occasions in different creating and metropolitan districts at the present. The greatness of air is adversely contacting collectible to the different styles of tainting influenced through the transportation, power, powers consumptions, and so forth. In our country population is a big problem as day by day population is increasing, so the rapid increasing in population and economic upswing is leading environment problems in city like air pollution, water pollution etc. In some of air pollution and air pollution is direct impact on human body. As we know that major pollutants are arising from Nitrogen Oxide, Carbon Monoxide & Particulate matter (PM), SO2 etc. Carbon Monoxide is arising due to the deficient Oxidization of propellant like as petroleum, gas, etc. nitrogen oxide (NO) is arising due to the ignition of thermal fuel; Sulphur Dioxide (So2) is major spread in air, So2 is a gas which is present more pollutants in air, it's affect more in human body. The predominance of air is overstated by multidimensional impacts containing spot, time and vague boundaries. The goal of this improvement is to take a gander at the AI basically based ways for air quality expectation. In this paper we will predict of air pollution by using machine learning algorithm.

1. Introduction:

The Environment describe about the thing which is everything happening in encircles the Environment is polluted by human daily activities which include like air pollution, noise pollution. If humidity is increasing more than automatically environment is going more hotter. Major cause of increasing pollution is increasing day by day transport and industries there are 75 % NO or other gas like CO, SO2 and other particle is exist in environment.. The expanding scene, vehicles and creations square measure harming all the air at a feared rate. Therefore, we have taken some attributes data like vehicles no., Pollutants attributes for prediction of pollution in specific zone of Delhi.

2. Literature Survey:

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.

The $PM_{2.5}$ problem is proving to be a major public crisis and is of great public-concern requiring an urgent response. Information about, and prediction of $PM_{2.5}$ from the perspectiveof atmospheric dynamic theory is still limited due to the complexity of the formation and development of $PM_{2.5}$. In this paper, we attempted to realize the relevance analysis and short- term prediction of $PM_{2.5}$ concentrations in Beijing, China, using multi-source data mining. A correlation analysis model of $PM_{2.5}$ to physical data (meteorological data, including regional average rainfall, daily mean temperature, average relative humidity, average wind speed, maximum wind speed, and other pollutant concentration data, including CO, NO_2, SO_2, PM_{10}) and social media data (microblog data) was proposed, based on the Multivariate Statistical Analysis method. The study found that during these factors, the value of average wind speed, the concentrations of CO, NO_2, PM_{10}, and the daily number of microblog entries with key words ‗Beijing; Air pollution' show high mathematical correlation with $PM_{2.5}$ concentrations. The correlation analysis was further studied based on a big data's machine learning model- Back Propagation Neural Network (hereinafter referred to as BPNN) model. It was found that the BPNN method performs better in correlation mining. Finally, an Autoregressive Integrated Moving Average (hereinafter referred to as ARIMA) Time Series model was applied in this paper to explore the prediction of $PM_{2.5}$ in the short- term time series. The predicted results were in good agreement with the observed data. This study is useful for helping realize real-time monitoring, analysis and pre-warning of $PM_{2.5}$ and it also helps to broaden the application of big data and the multi-source data mining methods.

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

A Bayesian network classifier can be used to estimate the probability of an air pollutant overcoming a certain threshold. Yet multiple predictions are typically required

regarding variables which are stochastically dependent, such as ozone measured in multiple stations or assessed according to by different indicators. The common practice (independent approach) is to devise an independent classifier for each class variable being predicted; yet this approach overlooks the dependencies among the class variables. By appropriately modeling such dependencies one can improve the accuracy of the forecasts. We address this problem by designing a multi-label classifier, which simultaneously predict multiple air pollution variables. To this end we design a multi-label classifier based on Bayesian networks and learn its structure through structural learning. We present experiments in three different case studies regarding the prediction of PM2.5 and ozone. The multi-label classifier outperforms the independent approach, allowing to take better decisions.

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).

Examining and protecting air quality has become one of the most essential activities for the government in many industrial and urban areas today. The meteorological and traffic factors, burning of fossil fuels, and industrial parameters play significant roles in air pollution. With this increasing air pollution, Weare in need of implementing models which will record information about concentrations of air pollutants(so2,no2,etc).The deposition of this harmfulgases in the air is affecting the quality of people‘s lives, especially in urban areas. Lately, many researchers began to use Big Data Analytics approach as there are environmental sensing networks and sensor data available. In this paper, machine learning techniques are used to predict the concentration of so2 in the environment. Sulphur dioxide irritates the skin and mucous membranes of the eyes, nose, throat, and lungs. Models in time series are employed to predict the so2 readings in nearing years or months.

3. Existing System:

The Air Pollution Forecasting System: Air Quality Index (AQI) is a record that gives the public the degree of contamination related with its wellbeing impacts. The AQI centers around the different wellbeing impacts that individuals may encounter dependent fair and square and long stretches of introduction to the poison concentration. The AQI values are not quite the same as nation to nation dependent on the air quality norm of the country.

The higher the AQI level more noteworthy is the danger of wellbeing related problems. The by and large point of this venture is to make a student calculation that will have the option to foresee the hourly contamination focus. Additionally, an Android application will be built up that will provide the clients about the constant contamination convergence of PM2.5 alongside the hourly forecasted value of the toxin fixation from the student calculation. The Android application will also recommend data of the less dirtied[1].

Disadvantages:

The system is not implemented Stepwise Multiple Linear Regression Method.

The system is not implemented Instance-Linear Regression Model.

4. Proposed System

Data assortment: There is a different method from which we collected data from various dependable sources like Delhi Gov. site.

Exploratory examination: We research and explore examination with various parameter like ID of outliners, consistency check, missing qualities, and so on, it‘s totally occurred in this period of the venture.

Data Manipulation control: In period of data control stage the required missing data need to insert in utilizing the mean estimations of that characteristic of information.[2]

Prediction of boundaries utilizing by gauge model:

For appropriate data indirect relapse we have to keep future qualities for different boundaries just

Advantages

The proposed system implemented Linear Regression is basically use for predicting the real values data y using continuous parameter.

Stepwise Multiple Linear Regression Method is used for continuous data testing and training in effective way.

5. Dataset Description:

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

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

6. METHODOLOGY:

Decision tree classifiers

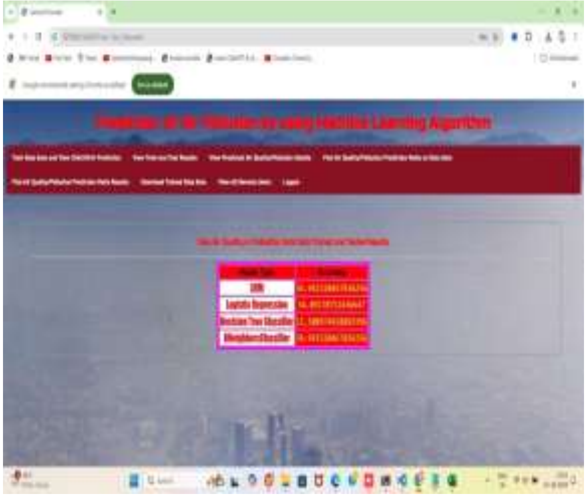
Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class.

Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in Shas one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

Gradient boosting

Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.[1][2] When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest. A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.



K-Nearest Neighbors (KNN)

Simple, but a very powerful classification algorithm

Classifies based on a similarity measure Non-parametric Lazy learning Does not ―learn‖ until the test example is given Whenever we have a new data to classify, we find its K-nearest neighbors from thetraining data

Result & Analysis:

7. 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 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

[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.