International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
©2021 INPRESSCO®, All Rights Reserved
Available at http://inpressco.com/category/ijcet

Research Article

Prediction of Air Quality using Machine Learning

Pooja Kulkarni and Prof.Reena Kharat

Department of Computer Engineering Pimpri Chinchwad college of Engineering Nigdi ,Pune

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

We speculate the pleasing air of India via the usage of machine gaining knowledge of to predict the air great index of a given area. Air pleasing index of India is a general measure used to indicate the pollutant (so2, no2, rspm, spm. etc.) degrees over a time. We developed a model for predicting the air quality index, based primarily on historical facts from previous years and predicting multivariable regression over a particular upcoming 12 month issue. Examining and protecting excellent air has become one of the most essential things the authorities need to do today in many industrial and urban areas .Factors of meteorology and tourists, fossil fuel burning, and manufacturing parameters play a major role in air pollution. With this growing contamination of the air,We want to propose models that document records of air pollutant concentrations (so2,no2,etc)

Keywords: *AQI; dataset; preprocessing; ML algorithm; pickle file*

Introduction

India, as the largest growing industrial nation, produces record amounts of pollutants specifically Co2, pm2.5 etc., and other hazardous aerial contaminants. The air quality of a specific state or country is a measure of the impact of pollutants on the valued regions, since pollutants are measured in terms of their scale according to the Indian air quality standard, these air quality indexes indicate the rates of major pollutants in the atmosphere. The rapid population growth and economic upswing in cities in developing countries such as India have led to environment al problems such as air pollution, water pollution, noise pollution, and many more[9]. The AQI may be a way of showing changes within the amount of pollution within the air. Air quality may be a measure of how clean or polluted the air is. we choose the algorithm in machine learning with high accuracy and training such as, ensemble method like bootstrapping or bagging.

Literature Survey

Mrs.A.GnanaSoundari,Mrs.J.GnanaJeslin,AkshayaA.C.[1] They used AHP MCDM technique to find of order of preference by similarity to ideal solution with regression model.They calculated the moving average of their data points and plotted the moving average. They used gradient boosting algorithm to remove the outliers Using Naïve Forecast approach, they spitted the dataset into two parts of first 75% and rest 25% data into test and train datasets to identify the huge seasonal variations and trend.

RuiJun YANG*, HaiLong ZHOU, DanFeng DING[2] They performed experiments on the data set of residential district In Tianhe, Guangzhou city, by using machine algorithm like SVM, Naive Bayesian and KNN, establish the internal mapping relationship between feature variables that include prices of housing and air quality. The result can meet the actual needs of home buyers in real estate.

Ms.Varsha Hable-Khandekar, Dr.Pravin Srinath[3] this paper summarizes air quality forecasting models as well as real time monitoring tools and techniques based on real-time and historical data. It has discussed the merits and demerits of every methodology used for air quality forecasting and monitoring used in recent research along with their comparative analysis and limitations, challenges.

Chavi Srivastava, Shyamli Singh, Amit Prakash Singh [4] They implemented different classification and regression techniques like Linear Regression, SDG Regression, Random Forest Regression, Decision Tree Regression,SupportVector Regression, Artificial Neural Networks, Gradient Boosting Regression and Adaptive Boosting Regression to forecast the Air Quality Index of major pollutants like PM2.5, PM10, CO, NO2, SO2 and O3.The techniques are then evaluated using Mean square error, Mean absolute error and R2.

Timothy M.Amado, Jennifer C.Dela Cruz[5] The proposed methodology is implemented by building a prototype for the integrated sensors using DHT 11

temperature and relative humidity sensor, MQ2, MQ5 and MQ135 gas sensors. Five predictive machine learning models are developed in this study like, *k*-nearest neighbors (KNN), support vector machine (SVM), Naïve-Bayesian classifier, random forest and neural network. They selected best predictive model in

Proposed Methodology

Air excellent prediction brings nonverbal signals that play a necessary role in interpersonal relationships. Great pair automated predictions can be a vital component of natural human interfaces, and can also be used in behavioral science and clinical practice. An automatic system of air quality forecasts needs to explain the following estimates of problems and the first-rate climate index of classification of datasets.

A. Architecture

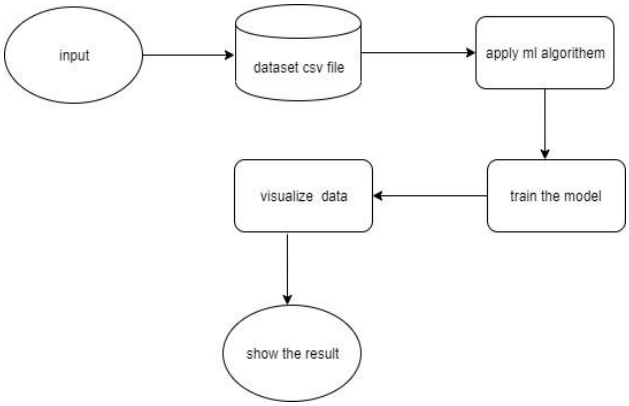

Fig.1 System Architecture

Description

A back-end interface (python, ml algorithm, pickle) and a Front end interface (JavaScript, css, jquery, html web browser) Composes the structure of a web based. Pickle file is used to Store data file and used for the code Reusability.

1) Input: input data in which that input the feature Of data value
2) Dataset(csv file): csv file in which that include the Features for example temperature, humidity, wind speed, time, pollutant's information etc.
3) ML algorithm: Apply to ensemble algorithm method Train Model, ml algorithm apply after 20 percentage data in which that train tests and split the model..
4) Visualize data: visualize the data using matplotlib and seaborn
5) Result: Air quality index prediction to predict the air quality ratio graph and predict the result and performance analysis.

B. Algorithm

Ensemble strategies combine multiple timber classifiers of choice to produce better overall predictive results than a single collection tree classifier. The main principle behind the ensemble model is that collectively a group of poor freshman comes to form astrong learner, thus increasing the version's accuracy.If we try to predict the target variable to use anysystem control strategy, the main reasons for discriminating between actual and predicted values are noise, uncertainty and bias.Ensemble helps all variables to be minimized (except fornoise, which is a sirreducible error). Using techniques suchas bagging and boosting, the uncertainty can be reduced andthe model's robustness multiplied.Combinations with lower variance in more than oneclassifier, especially in the case of unstable classifiers, andmay produce an extra reliable classification than an unmarried classifier.

1) Bootstrap:
Bootstrap refers to Replacement Random Sampling. Bootstrap allows one to further understand the disparity and conflict with the dataset Bootstrap requires random sampling of small subsets of data from the dataset. Bootstrap refers to Replacement Random Sampling. Bootstrap allows one to further understand the disparity and conflict with the dataset. Bootstrap requires random sampling of small subsets of data from the dataset. We know that our sample is little which our mean has errorin it. We can improve the estimate of our mean using thebootstrap procedure.

i) Create many (e.G. M) random sub-samples of our dataset with replacement (meaning we can pick the identical cost multiple times).
ii) Calculate the suggest of each sub-sample
iii) Calculate the average of all of our accrued approach use that as our expected mean for the data.
2) Bagging:
Bootstrap aggregation (or short, Bagging) is a very strong and simple ensemble process. Bagging is the bootstrap system's utility to high variance computer that learns information algorithm. Usually trees for decision making. Suppose there are N observations and M features. A sample from observation is selected randomly with Replacement(Bootstrapping).A subset of features are selected to create A model with pattern of observations and subsets of the features. Features from the subset is selected which offers the satisfactory cut up at the schooling data.

i) This is repeated to create many models and every model is trained in parallel.
ii) Prediction is given based on the aggregation of predictions from all the fashions.
Feature from the subset is selected which offers the

satisfactory cut up at the schooling data.
iii) This is repeated to create many models and every model is trained in parallel.
ii) Prediction is given based on the aggregation of predictions from all the fashion.
c. Advantages:
1) Proposed methodology performs extremely fast features Computation.
2) It performs efficient feature selection.
3) Improve the air quality.
4) Secure and efficient System

Existing System

The existing system involves an automated mannequin using Artificial Neural Networks and Kinging to forecast air pollution at various locations in Mumbai and Navi Mumbai using previous information available from the meteorological division and the Pollution Control Board. The proposed mannequin is implemented and MATLAB's use for ANN an R for Kriging is checked and the findings are reported. This method used the Linear Regression Protocol to forecast Subsequent day emissions. The gadget helps to forecast low print emissions from previous days based primarily on basic parameters and inspect important pollution points and predict future pollution.

Advantages:
1) By using Linear Regression, we can find nature of relationship between variables.
2) Neural networks are simple to implement.

Performance of Proposed System with Existing System

In existing system, has no deeper sample dataset used and the performance of classification algorithm is less [7].and need further study forecasting sudden pollution earlier and improve the forecasting accuracy in longer time[8]. In proposed system providing solution for prediction. Take a dataset and train them by using ensemble methods like bagging or boosting for more stable model and better prediction.

Result and Discussion

Specific statistical metrics are used in model evaluation, such as mean square error, mean absolute error,R2 etc.Boosting (XGboost) algorithms are used in proposed system for air quality prediction.in the dataset it checks all element one by one and handle missing values efficiently. XGboost uses the distributed weighted Quantile Sketch algorithm to find the optimal spilt points among the weighted datasets effectively. The algorithm comes with integrated method of cross validation at every iteration. Making predictions the air quality index to forecast the graph of the air quality ratio and simulate the outcomes and the performance analysis. View data through matplotlib and seaborn.

Conclusion

We are in a position to take air quality index steps for the outcome expected. For predicting values we used machine learning algorithm and prediction model. Apps that include or station code were useless. As they had nothing to do with so2 predictions. This model is no longer capable of showing anticipated production, as the statistics are not in order as they are in line with the date column. The same is the problem for cities. If we predict it will be useful for the whole state, so we can now calculate AQI and use classification

Future Work

Our next study aim could be to improve the overall performance of the class algorithms properly for higher prediction accuracy. At the same time, to check the algorithm intensively, we can use deeper sample facts and try to develop a few software applications.

References

[1]. Mrs. A. GnanaSoundari, (PhD) ,Mrs. J. GnanaJeslin M.E, (PhD),nAkshaya A.C. “Indian Air Quality Prediction And Analysis Using Machine Learning”. International Journal of Applied Engineering ResearchISSN 0973-4562 Volume 14, Number 11, 2019 (Special Issue).

[2]. RuiJun YANG*, HaiLong ZHOU, DanFeng DING Air Quality Prediction Method in Urban Residential Area 2018 11th International Symposium on Computational Intelligence and Design

[3]. Ms. Varsha Hable-Khandekar, Dr. Pravin Srinath Machine Learning Techniques for Air Quality Forecasting and Study on Real- Time Air Quality Monitoring

[4]. Chavi Srivastava, Shyamli Singh Estimation of Air Pollution in Delhi Using MachineLearning Techniques 2018 International Conference on Computing,PowerandCommunicationTechnologies*(GUC ON)*Galgotias University, Greater Noida, UP, India. Sep 28-29, 2018

[5]. Timothy M. Amado, Jennifer C. Dela Cruz Development of Machine Learning-based Predictive Models for Air Quality Monitoring andCharacterization Proceedings of TENCON 2018 - 2018 IEEE Region 10 Conference (Jeju, Korea, 28-31 October 2018)

[6]. Sharma M E A McBean and U.Ghosh, “Prediction of atmospheric sulphate deposition at sensitive receptors in northern India”, AtmosphericEnvironment 29.16(1995): 2157-2162.

[7]. Saba Ameer1, Munam Ali Shah1, Abid Khan1, Houbing Song2, Carsten Maple3, Saif ul Islam4, Muhammad Nabeel Asghar, “Comparative analysis of machine learning techniques for predicting air quality in smart cities”, 10.1109/ACCESS.2019.2925082, IEEE Access

[8]. Bo Liu[1,2], Shuo Yan[2], Jianqiang Li[2], Guangzhi Qu[3], Yong Li[2], Jianlei Lang[4,5], Rentao Gu[6].” An Attention-Based Air Quality Forecasting Method”. 2018 17th IEEE International Conference on Machine Learning and Applications..

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