International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. 04 Issue 02 |2020

Air Quality Index Prediction

Suresh Aneesh Jain
Masters in Computer Application (MCA), Department of Computer Application, Jain (deemed to be)University, Bangalore, India

ABSTRACT

As per world health organization the air pollution is more bad as compare to the world environment, so it has become a big issues which effects the people in the world. So in help of knowing the quality of air is good or not, I developed an model which helps in predicting the air quality. In this project I had used the machine learning and python technology, and some machine learning algorithms. Here in this model we had used the previous year data set to predict the accuracy value of the air quality. Here In this model we predict the pm2.5 (particular matter)value (its an air pollution value).
KEY WORDS: - Air quality , Machine Learning , Python ,pm2.5(particular matter)

1. INTRODUCTION

Now a days the air pollution problems are been increased for example cities like Delhi,Bangalore has the most air pollution which may effect the humans with health issues like cough , respiratory diseases , irritation of the eyes , etc… and as well its effects to the animals and the environment as the air quality is where is low. To predict the accuracy rate and quality of the air I had developed a model which makes us help to predict the air quality value which is pm2.5. The prediction is done by the help of using machine learning models and python libraries. The models used in this model are liner regression ,ridge regression , lasso regression , random forest regression , extreme gradient boosting . and we had used some python libraries for the graphs representation and to call the data frame, the libraries are numpy,pandas,matplotlib, seaborn .

2. REVIEW OF LITERATURE :-

2.1 Air Quality Prediction using Machine Learning Algorithms

(Pooja Bhalghat; Sejal Pitale and Sachin Bhoite) The main causes of air, water, land and various other pollutions in the developing countries are the rapid increase in the population and growth of cities in certain countries. Such developed and developing countries collectively cause the threat of pollution. It does not end with just the causing of air pollution but also gives rise to health issues of the population and have to deal with the long term consequences of this air pollution. This paper deals with a study to enhance air quality forecasting to reduce the pollution maximization that has been a threat to the environment. If a control is kept and monitored there are less chances for the explosive growth in air pollution which will minimize dangerous effects later on. Sulphur dioxide considered as a major pollutant has to be controlled and predictive concentration must be kept in mind. The system that is proposed in the paper is capable of such predictions.

2.2 A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization

(Dixian Zhu; Changjie Cai; Tianbao Yangand Xun Zhou)This paper studies the ways to tackle air quality forecasting by techniques of machine learning in order to predict the air pollutants and sulphur dioxide on an hourly basis. Efficiently training a model on big data is one of the popularly known techniques of machine learning. It mainly uses large scale optimization algorithms. In this paper, certain models are refined to predict air pollution concentration at hourly basis. Most of the studies are restricted to the using of models and data that are based on the standard regression models and old data used for air pollution concentration check. Though there are works that apply machine learning to the prediction of air quality, there is a repetitive usage of old models and data for such predictions. This paper examines the usage of meteorological data of the previous days for the prediction of air quality check for about 24 hour basis. This paper results the parameter reducing formulations provide better performance than the regression models and data existing methods.

2.3 Air Quality Prediction by Machine Learning Methods

(Huiping Peng) In this study, the author uses the numerical and observational data of the air quality to produce the results of the concentrated gases which cause air pollution. This data was collected for about a period of 48 hours for six stations across Canada. Air pollution as per the author is explained as a complex mixture of harmful and toxic gases that can cause high impact on the human health. Thus to keep a forecast of such pollution and keep a air quality check becomes a necessity and precaution. This would in return also help

in improving the life quality of the population. This paper has used multiple linear regression and multilayer perception neutral networks to forecast the various models used for concentration of pollutant. Machine learning becomes a major part to overcome the disadvantages of linear methods and huge demand of computational data. When the forecasting is operational, the models must be kept updated frequently and the data arrival is continuous. This paper provides the updated and latest models for the air quality index using methods of machine learning.

2.4 Indian Air Quality Prediction and Analysis Using Machine Learning

(A.GnanaSoundari; J.Gnana Jeslin) In this paper, the author has tried to objectify the method that is used to predict the air quality in India and various results that are extracted from the same. The air quality index of India is defined as a standard measure that is used for the indication of the pollutant level for a certain period of time. In India, the air quality is forecasted with the use of machine learning for a certain area. This study has developed a model so as to predict the air quality index on the basis of previous year dates and predictions of the succeeding years as a multivariable problem of regression. Cost estimation is applied for the improvement of efficiency of the model. This model will help to estimate the air quality index of either the country, state, area or any region bounded with the air pollution concentration. After the implementation of the proposed formulation model, a better performance was observed with an efficiency of 96 percent on the prediction of the air quality index for the whole of India.

3. OBJECTIVE OF THE STUDY

To build a model which helps in prediction of the air quality index.

4. RESEARCH DESIGN

4.1 Proposed system

The proposed system consists the role of finding the accuracy rate of the air quality (pm2.5) with the help of the previous year data set . this prediction is done with the help of the machine learning algorithms and python libraries.

4.2 Linear regression

This model is an machine learning algorithm which is used to find the relationship between the variable and forecasting . This algorithm is based on the supervised learning.

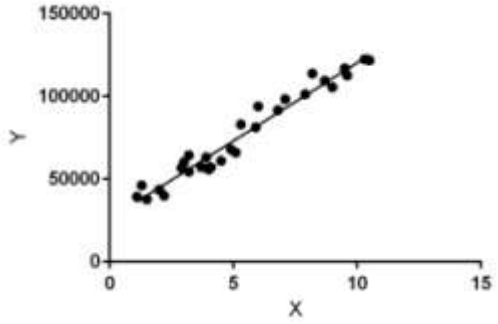


Figure 1 - example graph of linear regression

This algorithm performs a job to predict the variable value which is y on the base of the dependent value x. so this algorithm finds outs the relation between the x and y value where x is a input and y is a output . the linear regression algorithm describe the dependent variable with the help of the straight line which is known as y=mx+c.

4.3 Ridge Regression

The ridge regression is machine learning algorithm which is an advanced version of the linear regression algorithm . This algorithm takes the model further and penalize the data for the better prediction.

4.4 Lasso regression :-

The lasso regression is one of the machine learning algorithm , which is used to find the subset value of the variable. The lasso algorithm shrinkage the data set to find the best accuracy value. By the help of shrinking process the error rate will be reduce as compare to the normal regression.

4.5 Random Forest

The random forest is one the machine learning algorithm. It is an classification model which but its can be used in both way classification and regression way. In this algorithm the input values are been classified in the tree form and it will have a voting process to find the accuracy .

5. SYSTEM ARCHITECTURE

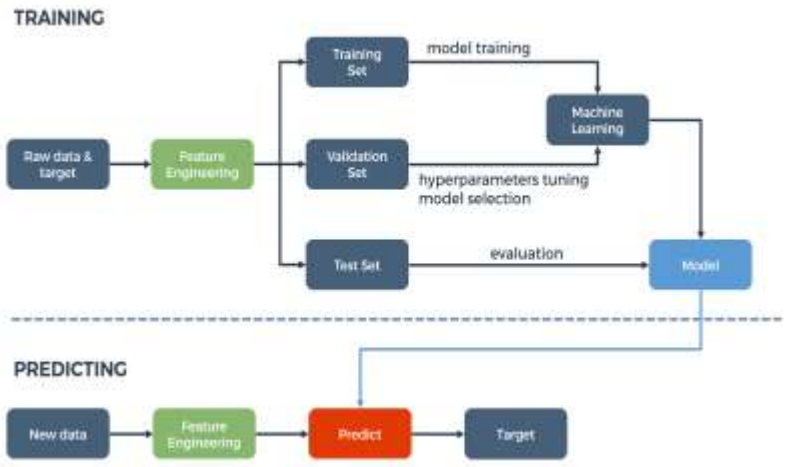


Figure 2. System architecture

As per the figure 2, there are some object which helped us to reference the process steps. so in figure 2 we use a object names Raw data and target this is a step where we insert the input values from the data set . after inserting the input data it calls the next step which is Feature engineering . The feature engineering job is to divide the input values of data set into three sets they are Train set , validation set , Test set . here we take only 80% input values from the data set and split into the three different set by the help of the feature engineering.

5.1 Feature engineering

The feature engineering is a process which is used to extract the important or useful features of data set so that our model performance gets increased.

5.2 Training Set

This is the set where we pass the 80% of input values from the data set. This set is used to train the model with the some input values like temperature, wind, humidity and the train the model to predict the future air index value.

5.3 Validation Set

This is the set which helps in training the model with the help of the parameters. This set is optional set. Its used to avoid the over fitting problem.

5.4 Test Set

The test set model is used to evaluate the performance of the model. it also helps us to check whether the model is predicting correctly or not. In this process we had used the Extra Tree Repressor. The extratreeregressor is a package of scikit -learning library to extract the top 5 features from our data set.

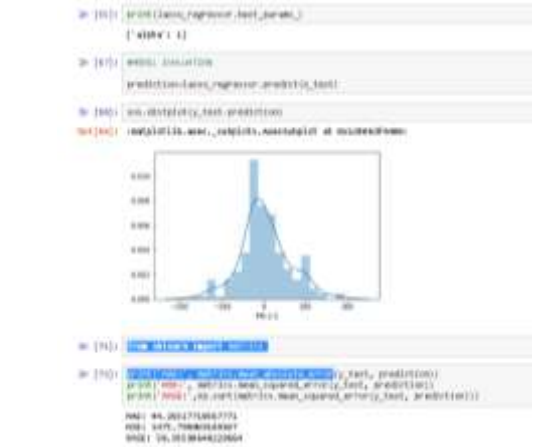
So in the figure 2, prediction process is used after the model is trained and tested through the sets. here in this prediction process we use 20% of data to check whether the model is working properly and to check whether the model is predicting correctly or not .

6. RESULT

6.1 Linear Regression

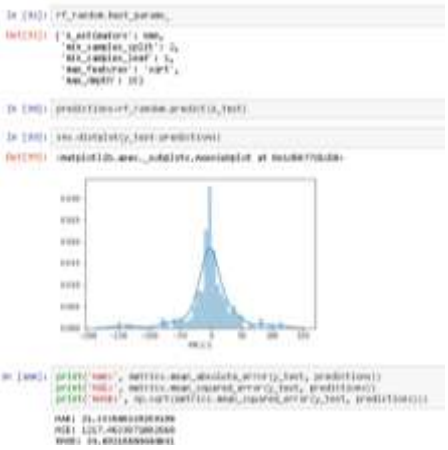
6.2 Ridge Regression

6.3 Lasso Regression

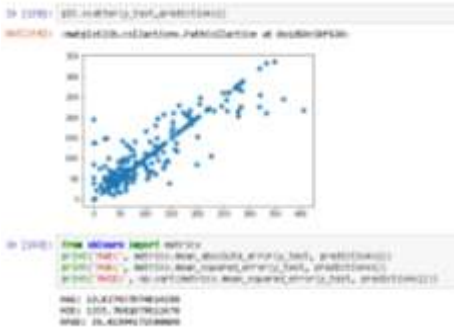


6.4 Random Forest Regression

6.5 Re Hyper Parameter Tuning



6.6 XG Boost Repressor



7. CONCLUSION:

In this project, we have developed efficient machine learning methods for air pollutant prediction. We have formulated the problem as regularized MTL and employed advanced optimization algorithms for solving different formulations. We have focused on alleviating model complexity by reducing the number of model parameters and on improving the performance by using a structured regularise. Our results show that the proposed light formulation achieves much better performance than the other two model formulations and that the regularization by enforcing prediction models for two consecutive hours to be close can also boost the performance of predictions. Here in this project we had shown the importance of the advanced optimization techniques for improving the performance and speed up the data. For future work, we will further consider the commonalities between nearby meteorology stations and combine them in a MTL framework, which may provide a further boosting for the prediction.

8. REFERENCES:

- Environmental Protection Agency (EPA). CFR Parts 50, 51, 52, 53, and 58-National Ambient Air Quality Standards for Particulate Matter: Final Rule. Fed. Regt. 2013, 78, 3086–3286.
- Jacob, D.J.; Winner, D.A. Effect of climate change on air quality. Atmos. Environ. 2009, 43, 51–63.
- https://www.researchgate.net/publication/335911816_Air_Quality_Prediction_using_Machine_Learning_Algorithm
- https://www.mdpi.com/2504-2289/2/1/5/pdf
- https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-017-4914-3
- https://scholar.google.com/scholar?oe=utf-8&gcc=in&ctzn=Asia/Kolkata&ctf=0&v=11.7.11.21.arm&fheit=0&biw=360&bih=640&ntyp=1&ram_mb=2815&devloc=0&ampcct=4044&client=ms-android-samsung&wf=pp1&padt=200&padb=640&hl=en-GB&cds=0&psm=0&dbla=1&um=1&ie=UTF-8&lr&q=related:8wJC2tPqKvFJwM:scholar.google.com/#d=gs_qabs&u=%23p%3D8wJC2tPqKvEJ
- https://scholar.google.co.in/scholar?q=air+quality+index+using+machine+learning+research+papers&hl=en&as_sdt=0&as_vis=1&oi=scholart#d=gs_qabs&u=%23p%3DuA0DzzTX_RoJ
- https://scholar.google.co.in/scholar?q=air+quality+index+using+machine+learning+research+papers&hl=en&as_sdt=0&as_vis=1&oi=scholart#d=gs_qabs&u=%23p%3DkliQhsF9rQsJ
- https://www.sciencedirect.com/science/article/abs/pii/S0048969720316910
- https://www.intechopen.com/books/machine-learning-advanced-techniques-and-emerging-applications/regression-models-to-predict-air-pollution-from-affordable-data-collections
- https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.mdpi.com/2504-2289/2/1/5/pdf&ved=2ahUKEwi20tTg9LTpAhWz4jgGHZLTBOIQFjAAegQIBBAB&usg=AOvVaw30q5XbK7lPs0AmRgr-hieX
- https://www.google.com/url?sa=t&source=web&rct=j&url=https://pdfs.semanticscholar.org/65b4/6801d18d66eb2f15dc5ef6c92433d31d5853.pdf&ved=2ahUKEwi20tTg9LTpAhWz4jgGHZLTBOIQFjABegQICRAB&usg=AOvVaw1e3cy2c87JwKbTf0GWiZIx
- https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.ripublication.com/ijaerspl2019/ijaerv14n11spl_34.pdf&ved=2ahUKEwi20tTg9LTpAhWz4jgGHZLTBOIQFjALegQIDRAB&usg=AOvVaw0RXY_DMWLEZ4EcrbmKnFLw
- https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.ijrte.org/wp-content/uploads/papers/v8i1/A3492058119.pdf&ved=2ahUKEwi20tTg9LTpAhWz4jgGHZLTBOIQFjANegQIAxAB&usg=AOvVaw0_IxR2_3SEpT1cZgUQCHcd