Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 10 Issue: 05 | Oct-2021

INDIAN AIR QUALITY PREDICTION AND ANALYSIS USING MACHINE LEARNING

Vivek Singh [1]

Department of Electrical Engineering & K. V. N. Naik Collage of Engineering, Nashik

---***---

Abstract -We read the air quality of India by using machine literacy to prognosticate the air quality indicator of a given area. Air quality indicator of India is a standard measure used to indicate the contaminant (so2, no2, rspm, spm.etc.) situations over a period. We developed a model to prognosticate the air quality indicator grounded on literal data of former times and prognosticating over a particular forthcoming time as aGradient decent boosted multivariable retrogression problem. we ameliorate the effectiveness of the model by applying cost Estimation for our prophetic Problem. Our model will be able for successfully prognosticating the air quality indicator of a total county or any state or any bounded region handed with the literal data of contaminant attention. In our model by enforcing the proposed parameter- reducing phrasings, we achieved better performance than the standard retrogression models. our model has 96 delicacy on prognosticating the current available dataset on prognosticating the air quality indicator of whole India, also we use AHP MCDM fashion to find of order of preference by similarity to ideal result.

Key Words: AQI, dataset, preprocessing, outliers, BVA, vaticination

1. INTRODUCTION

As the largest growing artificial nation, India is producing record quantum of adulterants specifically Co2, pm2.5 etc and other dangerous upstanding pollutants. Air quality of a particular state or a country is a measure on the effect of adulterants on the reputed regions, as per the Indian air quality standard adulterants are listed in terms of their scale, these air quality indicators indicates the situations of major adulterants on the atmosphere. There are colorful atmospheric feasts which causes pollution on our terrain. Each pollution has individual indicator and scales at different situations. The major adulterants Similar as (no2, so2, rspm, spm) indicators AQI is acquired, with this individual AQI, the data can be distributed grounded on the limits. We collected the data from the Indian government database, which contains contaminant attention being at colorful places across India. We start by calculating the individual indicator of the contaminant for every available datapoints and find their separate AQI for the region. We've designed a model to prognosticate the air quality indicator of every available data points in the dataset, our model is able of vaticinating the air quality of India in any given area. By prognosticating the air quality indicator, we can annul the major pollution causing contaminant and the position affected seriously by the contaminant across India. With this soothsaying model, colorful knowledge about the data are uprooted using colorful ways to gain heavily affected regions on a particular region (cluster). This give further information and knowledge about the cause and senility of the adulterants.

2. AIR QUALITY INDEX PREDICTION MODEL

A. SYSTEM ANALYSIS

The Fine material (PM2.5) could be a important bone as a result of it's a giant concern to people's health once its position within the air is comparatively high. PM2.5 refers to little patches within the air that gauge back visibility and beget the air to look hazy once situations are elevated. But in the proposed system we calculate the air quality indicator of all the adulterants using the AQI formulae to know the air quality position in a particular megacity using grade descent and Box-Plot analysis. In the proposed system the air quality indicator of the forthcoming times can be prognosticated using the present AQI values.

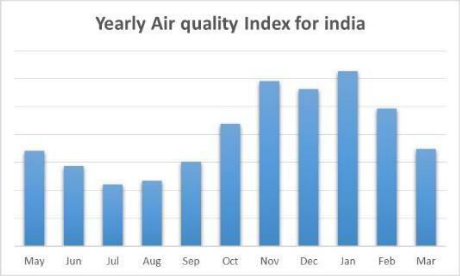


Figure 1 Air quality index

BACK PROPAGATION

Back propagation is a fashion employed in fake neural systems to figure an inclination that's needed in the count of the loads to be employed in the network. Back propagation is longhand for"the retrogressive proliferation of miscalculations,"since a boob is reused at the yield and appropriated in reverse all through the system's layers. It's regularly habituated prepare profound neural networks.

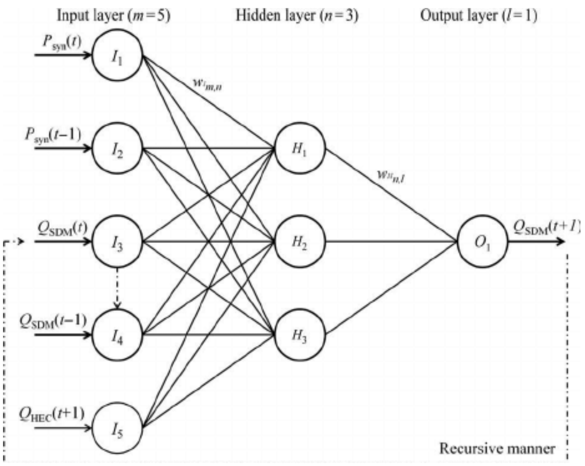


Figure 2 Neural networks

Back spread is a enterprise of the delta guideline tomulti-layered feed forward systems, made conceivable by exercising the chain principle to iteratively register angles for each subcaste. It's forcefully linked with the Gauss – Newton computation and is a piece of pacing with exploration in neural reverse spread

3. EXPERIMENTAL ANALYSIS

A. DATA SOURCES

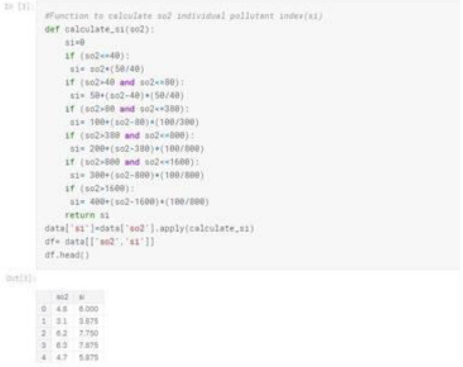
To prognosticate the air quality indicator of a particular region, we need the contaminant attention of all the feasts which will be available in thecpcb.nic.in website, which holds all the data that pollutes the metropolises every time. The AQI formulae will be applied in order to calculate the AQI by using the direct retrogression algorithm for a particular time. Several datasets will be imported inside the directory and null values will be set to the horizonless data. The prognosticated and factual values will be represented using the Box- Plot analysis in order to remove the outliers.

B. PRE-PROCESSING THE DATA

In this dataset the outliers are substantially of defective detector or transmission crimes, these crimes have huge variation than the normal valid results. We know the standard range of adulterants occurs on a particular areaso to remove the outliers from the data we use boundary value analysis. By using BVA we plant the upper quartile range and lower quartile range of a given data.

C. AQI SIMULATION AND CALCULATION

We acquired the dataset with colorful columns of detector data from colorful places in India. we've the average readings of ambient air quality with respect to air quality parameters, like Sulphur dioxide (So2), Nitrogen dioxide (No2), Respirable Suspended Particulate Matter (RSPM) and Suspended Particulate Matter (SPM). Data acquired from the source has further noisy data since many of the data from the stations have been shifted or closed the period were marked as NAN or notavailable.so we've topre-process the data in order to remove theoutliers.Each individual contaminant indicators, gives the relationship between the contaminant attention and their corresponding individualindex.Figure 3 shows an illustration of the individual AQI computation of SO2



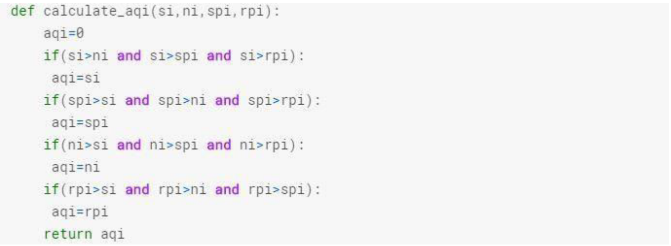
```
In [1]:
#Function to calculate so2 individual pollutant index(si)
def calculate_si(so2):
    si=0
    if (so2<=40):
     si= so2*(50/40)
    if (so2>40 and so2<=80):
     si= 50+(so2-40)*(50/40)
    if (so2>80 and so2<=380):
     si= 100+(so2-80)*(100/380)
    if (so2>380 and so2<=800):
     si= 200+(so2-380)*(100/800)
    if (so2>800 and so2<=1600):
     si= 300+(so2-800)*(100/800)
    if (so2>1600):
     si= 400+(so2-1600)*(100/800)
    return si
data['si']=data['so2'].apply(calculate_si)
df= data[['so2','si']]
df.head()
```

Out[1]:

	so2	si
0	4.8	6.000
1	3.1	3.875
2	6.2	7.750
3	6.3	7.875
4	4.7	5.875

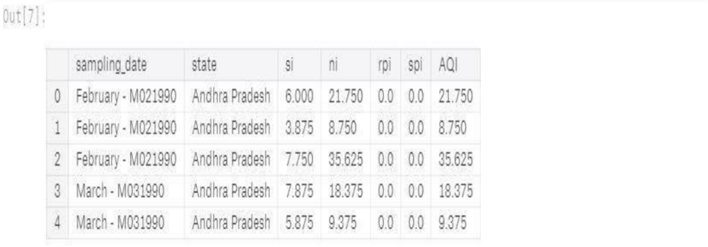
Figure 3 Calculation of SO2

The air quality indicator of a particular data point is the total of maximum listed contaminant on that particular area. That adulterants maxsub indicator is taken as the air quality indicator of that particular position. Figure 4 shows the mean AQI computation of all the feasts



```
def calculate_aqi(si,ni,spi,rpi):
    aqi=0
    if(si>ni and si>spi and si>rpi):
     aqi=si
    if(spi>si and spi>ni and spi>rpi):
     aqi=spi
    if(ni>si and ni>spi and ni>rpi):
     aqi=ni
    if(rpi>si and rpi>ni and rpi>spi):
     aqi=rpi
    return aqi
```

Figure 4 AQI Calculation



Out[7]:

	sampling_date	state	si	ni	rpi	spi	AQI
0	February - M021990	Andhra Pradesh	6.000	21.750	0.0	0.0	21.750
1	February - M021990	Andhra Pradesh	3.875	8.750	0.0	0.0	8.750
2	February - M021990	Andhra Pradesh	7.750	35.625	0.0	0.0	35.625
3	March - M031990	Andhra Pradesh	7.875	18.375	0.0	0.0	18.375
4	March - M031990	Andhra Pradesh	5.875	9.375	0.0	0.0	9.375

Figure 5 Mean AQI

In this graph AQI is the average value of AQI of each time across India.

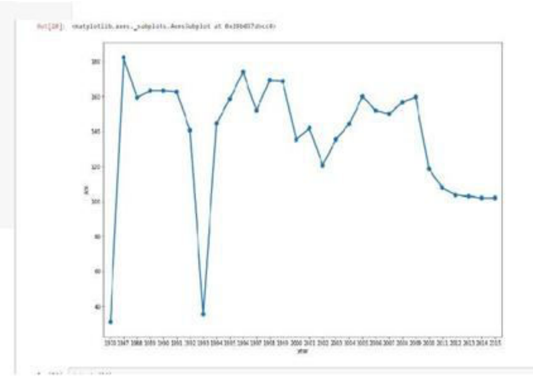


Figure 6 Graph between average AQI and sample data

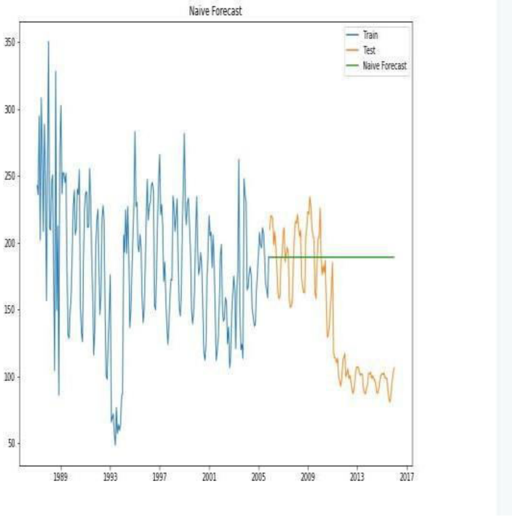
D. PREDICTION OF AIR QUALITY INDEX

Using Using Naïve Forecast approach, we pecked the dataset into two corridor of first 75 and rest 25 data into test and train datasets to identify the huge seasonal variations and trend.

We calculated the moving normal of our datapoints and colluded the moving normal. We linked the moving average varies one the time (2010-2011) i.e. before 2010 there are variations at x minimum and x outside and after 2011 the variations are y minimum and y outside.

Colluded the graph of train and test dataset with their moving average and anatomized the moving normal. Figure 7 shows the moving average graph.

Figure 7 Moving average graph



E. RESULTS ANALYSIS

Box plot is one of common graphical systems employed inEDA.A jalopy plot or boxplot is a helpful system for graphically portraying gatherings of numerical information through their quartiles. Box plots may likewise have lines broadening vertically from the holders (bristles) demonstrating inconstancy outside the upper and lower quartiles, hereafter the terms box-and- hair plot and box-and-stubblegraph. Exceptions might be colluded as individual focuses.

Box Plot gives abecedarian data about a dissipation. It graphically delineates a gathering of numerical information as indicated.

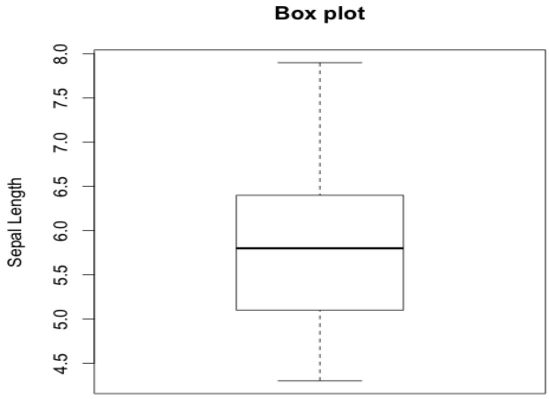


Figure 8 Box-Plot analysis

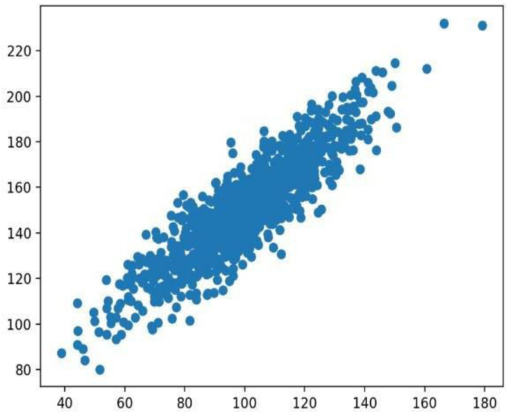


Figure 9 testing the dataset

By this data analysis we came to know that there are seasonal variations and trend, in order to reduce these criteria, we etest the data month wise to prognosticate it month wise. By etesting the data, we can reduce the outlier more efficiently than raw data. After removing the outlier's direct retrogression is applied to the filtered data and to fit the trent line on the data points grade descent hyperactive parameters are used to optimize the model.

LINEAR REGRESSION

While doing straight fall our thing is to fit a line through the dispersion which is closest to the maturity of the focuses. Latterly lessening the separation (mistake term) of information focuses from the fitted line.

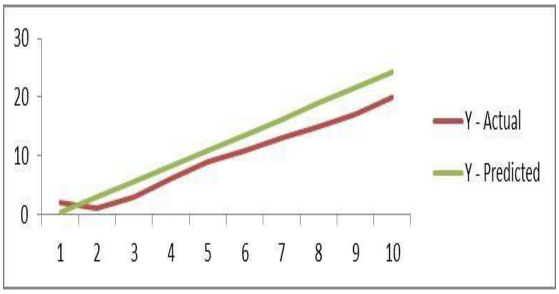


Figure 10 Linear regression graph

Y=mx +c denotes the equation of regression line

GRADIENT BOOST ALGORITHM

The principle issue told by individualities is air impurity since air contains multitudinous substances which might be made by manmade or regular procedure. The Air substances present most organic tittles, points of interest and dangerous material into the air. Boosting Algorithm is a victor among the most current literacy perceptivity showed over the most recent twenty times.

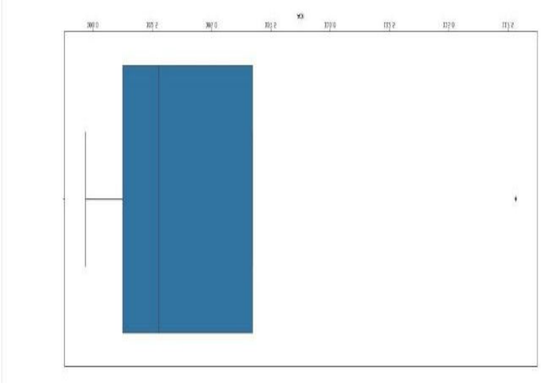


Figure 11 Outlier removal using BPA

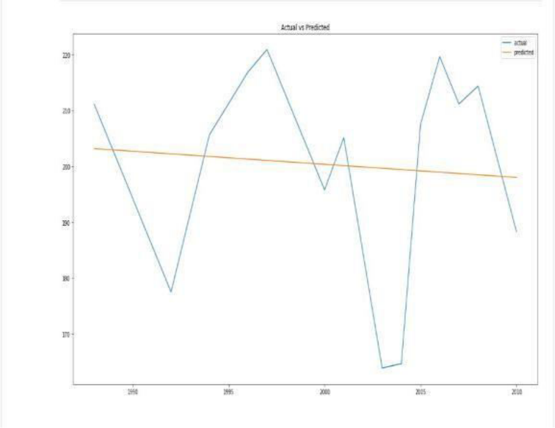


Figure 12 Actual and predicted values

4. CONCLUSIONS AND FUTURE ENHANCEMENTS

The Since our model is able of prognosticating the current data with 95 delicacy it'll successfully prognosticate the forthcoming air quality indicator of any particular data within a given region. With this model we can read the AQI and alert the reputed region of the country also it a progressive literacy model it's able of tracing back to the particular position demanded attention handed the time series data of every possible region demandedattention.The air quality information employed in this paper originates from the demitasse air quality checking and disquisition stage, and incorporates the normal every day fine particulate issue (PM2.5), inhalable particulate issue (PM10), ozone (O3), CO, SO2, NO2 obsession and air quality record (AQI). The essential perspectives that should be viewed as with respects to guaging of the bane focus are its different sources alongside the factors that impact its obsession.

REFERENCES

[1] Dragomir, Elia Georgiana. "Air quality index prediction using K-nearest neighbor technique no. 1 (2010): 103-108.

[2] Carbajal-Hernández, José Juan "Assessment and prediction of air quality using fuzzy logic and autoregressive models." Atmospheric Environment 60 (2012): 37-50.

[3] Kumar, Anikender and P. Goyal, " Forcasting of daily air quality index in Delhi", Science of th Total Environment 409, no. 24(2011): 5517- 5523..

[4] Singh Kunwar P., et al. "Linear and nonlinear modelling approaches for urban air quality prediction, " Science of the Total Environment 426(2012):244-255.

[5] Sivacoumar R, et al, " Air pollution modelling for an industrial complex and model performance evaluation ", Environmental Pollution 111.3 (2001) : 471-477

[6] Gokhale sharad and Namita Raokhande, "Performance evaluation of air quality models for predicting PM10 and PM2.5 concentrations at urban traffic intersection during winter period", Science of the total environment 394.1(2008): 9- 24.

[7] Bhanarkar, A. D., et al, "Assessment of contribution of SO2 and NO2 from different sources in Jamshedpur region, India, "Atmospheric Environment 39.40(2005):7745-India." Atmospheric Environment 39.40 (2005): 7745-7760.

[8] Singh Kunwar P., Shikha Gupta and Premanjali Rai, " Identifying pollution sources and prediction urban air quality using ensemble learning methods", Atmospheric environment80 (2013): 426-437.

[9] Wang Jun, and Sundar A. Christopher, "Intercomparison between satellite derived aerosol optical thickness and PM2. 5 Mass: Impliances for air quality studies", Geophysical research letters30.21(2003).

[10] Sharma M E A McBean and U.Ghosh, "Prediction of atmospheric sulphate deposition at sensitive receptors in northern India", Atmospheric Environment 29.16(1995): 2157- 2162.

[11] Russo Ana Frank Raischel and Pedro G.Lind, "Air quality prediction using optimal neural networks with

stochastic variables”, Atmospheric Environment 79(2013): 822-830.

[12] Challa Venkara Srinivas et al ,” Data Assimilation and performance of Wrf for Air Quality Modeling in Mississippi Gulf Coastal Region “

[13] Hutchison Keith D., Solar Smith and Shazia J. Faruqui, “ Correlating MODIS aerosol optical thickness data with ground-based PM2.5 observations across Texas for use in a real time air quality prediction system, “ Atmospheric Environment 39.37(2005) :7190 – 7203

[14] Wang Z et al , “ A nested air quality prediction modelling system for urban and regional scales : Application for high high-ozone episode in Taiwan “ Water, Air and Soil Pollution130.1-4(2001):391-396

[15] Nallakaruppan, M. K., and U. Senthil Kumaran. "Quick fix for obstacles emerging in management recruitment measure using IOT- based candidate selection." Service Oriented Computing and Applications 12.3-4 (2018): 275-284.

[16] Nallakaruppan, M. K., and Harun SurejIlango. "Location Aware Climate Sensing and Real Time Data Analysis." Computing and Communication Technologies (WCCCT), 2017 World Congress on. IEEE, 2017.