International Journal of Innovative Research in Science and Engineering
Volume 7, Issue 08, 2021
www.ijirse.com
ISSN: 2454-9665

Air Quality Measurement using Machine Learning

Sudhanshu[1], Shivam Shubham Chauhan[2], Vivek Jha[3], Trapti Rautaila[4]

[1]*U.G. Student, Department of Computer Science and Engineering, IIMT College of Engineering, Greater NOIDA, Dr. A.P.J. Abdul Kalam Technical University, Uttar Pradesh, India*Sudhanshu22.iimt@gmail.com

[2]*U.G. Student, Department of Computer Science and Engineering, IIMT College of Engineering, Greater NOIDA, Dr. A.P.J. Abdul Kalam Technical University, Uttar Pradesh, India.* Shubhamchauhan.iimt@gmail.com

ABSTRACT:-

Due to the rapid increment of air pollution the examining and protecting air quality has become one of the most essential activities for the government today. The air quality becoming poor and poorer due to the various forms of pollution caused by transportation, electricity,fuel uses etc. The deposition of harmful gases is becoming the serious threat for the people of urban area .With this increasing air pollution, weare in need of implementing models which will recordinformation about concentrations of air pollutants(SO2,NO2,etc). Life of people is to much affected by those deposition gasses, especially in urban areas. Lately, many researchers began to use Big Data Analytics approach as there areenvironmental sensing networks and sensor data available.In this paper, we predict the concentration of SO2 in air using Machine Learning .Sulphur dioxide irritates the skin and mucous membranes of the eyes, nose, throat, andlungs.Models in time series are employed to predict the SO2 readings in nearing years or months.

***Keywords:* Machine Learning, Time Series, Prediction, Air Quality, SO2.**

INTRODUCTION:-

In the developing countries like India, the rapid increase inpopulation and economic upswing in cities have lead toenvironmental problems such as air pollution, water pollution,noise pollution and many more.According to the world's worst polluted places byBlacksmith Institute in 2008 [1], two of the worst pollutionproblems in the world are urban air quality and indoor airpollution. Air pollution has directimpact on humans health .There has been increased publicawareness about the same in our country.Global warming,acid rains, increase in the number of asthma patients are someof the long-term consequences of air pollution. Precised airquality forecasting can reduce the effect of maximal pollutionon the humans and biosphere as well. Hence, enhancing airquality forecasting is one of the prime targets for the society.SO2 is one of the major pollutantspresent in air.It is colorless and has a nasty, sharp smell.Itcombines easily with other chemicals to form harmfulsubstances like sulphuric acid, sulfurous acid etc.SO2 affects human health when it is breathed in. Itirritates the nose, throat,

and airways to cause **coughing,wheezing, shortness of breath**, or a tight feeling around thechest. SO2 is also affect the plants and living of animals.

Air quality evaluation has been conducted usingconventional approaches in all these years. These approachesinvolve manual collection and assessment of raw data.According to Niharika *et al*., [2], the traditional approachesfor air quality prediction use mathematical and statisticaltechniques. In these techniques, initially a physical model isdesigned and data is coded with maths equations. Butsuch methods suffer from disadvantages like:
1) They provide limited accuracy as they are unable topredict the extreme points i.e. the pollution maximumand minimum.
2) Cut-offs cannot be determined using such approach.
3) They use inefficient approach for better output prediction.
4) The existence of complex mathematical calculations.
5) Equal treatment to the old data and new data.

The proposed system is capable of predicting concentration ofSO2 for forthcoming months / years.

AIR QUALITY EVALUTION :-

Air quality evaluation is a way for evaluating andcontrolling air pollution. The characteristics of air supply affectits suitability for a specific use. A few air pollutants, calledcriteria air pollutants, are common throughout the world. These pollutants caused injured health, harm theenvironment and property damage. Thosepollutants are:
1) Carbon Monoxide (CO)
2) Lead (Pb)
3) Nitrogen Dioxide (NO2)
4) Ozone (O3)
5) Particulate matter (PM)
6) Sulfur Dioxide (SO2).
The Air Quality System (AQS) contains ambient air
pollution data collected by EPA, state, local, and tribal airpollution control agencies from over thousands of monitors.AQS also contains meteorological data, descriptiveinformation about each monitoring station (including itsgeographic location and its operator), and data qualityassurance/quality control information. AQS data is used toassess air quality, assist in Attainment/Non-Attainmentdesignations, evaluate State Implementation Plans forNon-Attainment Areas, perform modeling for permit reviewanalysis, and other air quality management functions. AQSinformation is also used to prepare reports for Congress asmandated by the Clean Air Act.

Table : AQI Classification.

AQI	AIR POLLUTION LEVEL
0-50	Excellent
51-100	Good
101-150	Light Polluted
151-200	Moderately Polluted
201-300	Heavily Polluted
Above 300	Severely Polluted

DATASET:-

3.1 Dataset/Source: Kaggle

Structured/Unstructured data: Structured Data in CSVformat.

Dataset Description:

The dataset consists of around 450000 records of all thestates of India.We worked only on Dataset ofUttar Pradesh.So we had 60383 records. This datasetconsist of 13 attributes listed below.

1)stn_code
2)sampling_date
3) state
4) location
5) agency
6)type
7)so2
8)no2
9)rspm
10) spm
11)location_monitoring_station
12)pm2_5
13)date

Station code is a code given to each station that recorded thedata,sampling_date is the date when the data is recorded. Stateand location represents state and cities whose data is recordedand agency is the name of agency that recorded the data.Typestates the type of area where the data was recorded such as industrial,residential,etc.so2,no2,rspm and spm is the amountof sulphur dioxide, nitrogen dioxide, respirable suspendedparticulate matter and suspended particulate matter measuredrespectively. Date is a cleaner version of sampling_date.

PM2.5 refers to atmospheric particulate matter (PM) that havea diameter of less than 2.5 micrometers, which is about 3%the diameter of a human hair.But majority of values in this column are null.

Splitting for Testing : Data Splitting was done as 80% fortraining and 20% for testing.

Preprocessing and Feature Selection:

International Journal of Innovative Research in Science and Engineering
Volume 7, Issue 08, 2021
www.ijirse.com
ISSN: 2454-9665

We only studied and applied algorithms on the data ofUttar Pradesh State .Hence, no. of rows was reduced to 60,383and state column automatically is of no more use.All the values in pm2_5 were null values ,so we dropped thecolumn.The agency's name have nothing to do with howmuch polluted the state is. Similarly, stn_code is also notuseful.
The date is a cleaner representation of sampling_date attributeand so we will eliminate the redundancy by removing thelatter. location_monitoring_station attribute is again unnecessary as it contains the location of the monitoringstation which we do not need to consider for the analysis.So, to summarize we have deleted the following features fromour dataset :state,pm2_5,agency, stn_code,sampling_date and location_monitoring_station.We have simplified the type attribute to contain only one ofthe three categories: industrial, residential, other.For SO2 andNO2, we replaced NAN values by mean.For date, we havedropped NAN values as there were only 3 null values.So after pre-processing our dataset contains 60,380 rows and7 columns.

Fig1: Calculation of S02

The air quality index of a particular data point isthe aggregate of maximum indexed pollutant onthat particular area. That pollutants maxsubindex is taken as the air quality index of thatparticular location. Figure 2 shows the meanAQI calculation of all the gases.

```
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
```

Fig2: AQI calculation

International Journal of Innovative Research in Science and Engineering
Volume 7, Issue 08, 2021
www.ijirse.com
ISSN: 2454-9665

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

Fig3: Mean AQI

EXPOLATORY DATA ANALYSIS:-

•The below graph shows concentration of so2 over the years.It was highest in the years of 1998 and 2013and lowest in the year2007.However,it is bad for the latest years.

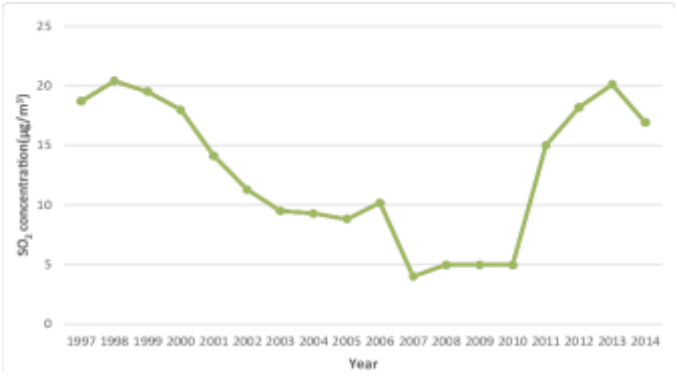

Fig: 4 concentration of SO2 over the years.

•This graph shows that the amount of so2 is highest in the residential areas.

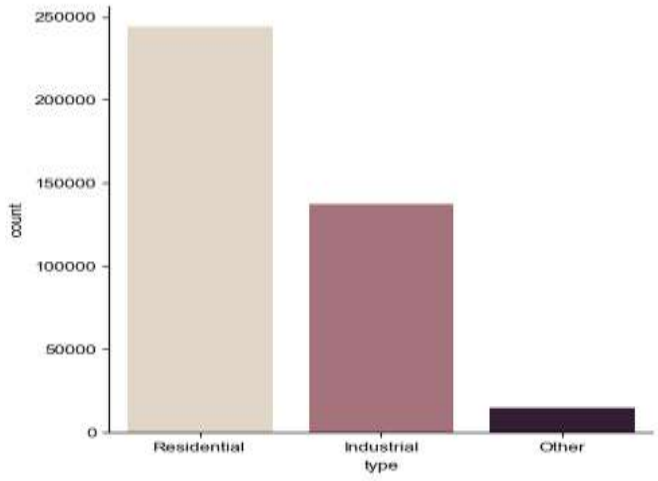

Fig 5: count of SO2 in different areas.

CONCLUSION :-

Based on the bar plots plotted we come to the conclusionthat some cities are highly polluted and need urgentattention. Also for cities like Noida,Ghaziabad whereconcentration of so2 is increasing, we can take measuresfrom now to not face problems later.We used AR model and ARIMA model for predicting values of so2. Featuressuch as location_monitoring_station or station code wereof no use as they have nothing to do with so2 predictions.So2 safe levels are as follows:0.20 ppm (parts per million) averaged over a one hour period.0.08 ppm averaged over a 24 hours period. 0.02 ppm averagedover a one year period.In order to predict air quality, pm2_5 is also an importantattribute. The values of this must be recorded in future asthis particulates are responsible for various health effects including cardiovascular effects such as cardiac arrhythmiasand heart attacks, and respiratoryeffects such as asthmaattacks and bronchitis.This model is not able to show expected output as the data isnot in sequence as per date column.The same is the problemfor cities.If we predict for the entire state, it won't be helpfulSo we will be now calculating AQI and use classificationmodels further.This model further, also makes us aware of the challenges infuture and research needs such as pm2.5,AQI,etc.

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 the 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).

ISSN: 2454-9665

[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] ChallaVenkara 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 IOTbased 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.