AIR QUALITY ANALYSIS

EXPERIMENTAL ANALYSIS

A.DATA SOURCES

- To predict the air quality index of a particular region , we need the pollutant concentration of all the gases which will be available in the cpcb.nic.in website, which holds all the data that pollutes the cities every year.
- The AQI formulae will be applied in order to calculate the AQI by using the linear regression algorithm for a particular year.
- Several datasets will be imported inside the directory and null values will be set to the infinite data. The predicted and actual values will be represented using the Box-Plot analysis in order to remove the outliers.

BYRE-PROCESSING THE DATA

- In this dataset the outliers are mainly Of faulty sensor or transmission errors, these errors have huge variation than the normal valid results.
- We know the standard range of pollutants occurs on a particular area so to remove the outliers from the data we use boundary value analysis.
- By using BVA we found the upper quartile range and lower quartile range Of a given data.

SIMULATION AND CALCULATION

- We acquired the dataset with various columns Of sensor data from various places in India. we have
- the average readings of ambient air quality with respect to air quality parameters, like Sulphur dioxide (S02), Nitrogen dioxide (N02), Reparable Suspended Particulate Matter (RSPM) and Suspended Particulate Matter (SPM).
- Data acquired from the source has more noisy data since few of the data from the stations have been shifted or closed the period were marked as NAN or not available.so we have to pre-process the data in order to remove the outliers. Each individual pollutant indexes, gives the relationship between the pollutant concentration and their corresponding individual index.

```
#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)*(100/300)
    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()
```

	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 S02

- The air quality index of a particular data point is the aggregate of maximum indexed pollutant on that particular area. That pollutants maxsub index is taken as the air quality index of that particular location.

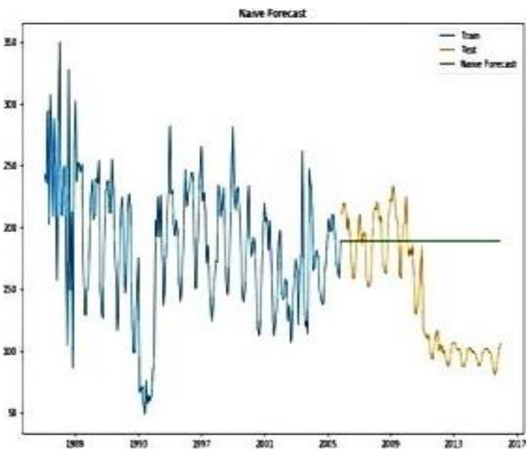
AQI calculation of all the gases

```
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
```

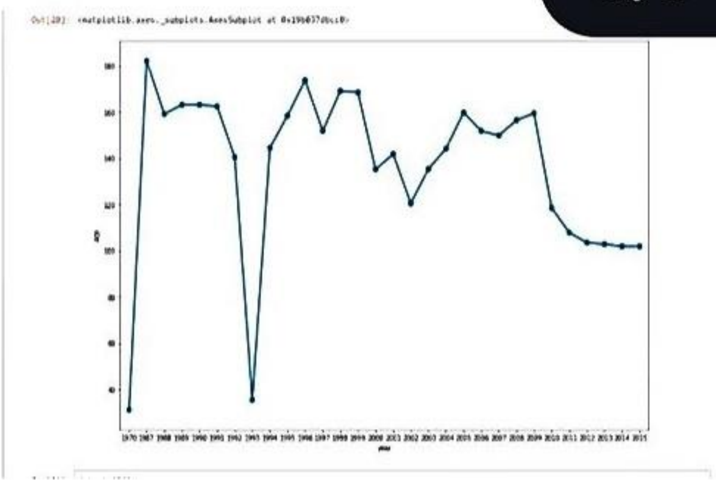
Figure 4 AQI Calculation

D.PREDICTION OF AIR QUALITY INDEX

- Using Naïve Forecast approach, we 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.
- We calculated the moving average of our data points and plotted the moving average. We identified the moving average varies one the year (2010-2011) i.e. before 2010 there are variations at x minimum and x maximum and after 2011 the variations are y minimum and y maximum.
- Plotted the graph of train and test dataset with their moving average and analyzed the moving average.



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



3/6

Figure 5 Mean AQI