Air Quality Index (AQI): Prediction and Optimization

Dr. Jitendra Saturwar[1], Baljit Thakur[2], Nayan Siddhapura[3], Dilip Sahani[4]

[1]Dept. Computer Engineering, Universal College of Engineering, Maharashtra, India
[2,3,4]Dept. Computer Engineering, Universal College of Engineering, Maharashtra, India

---***---

Abstract - — *People's health has been affected in the worst way possible with the development of industrialization and urbanization. Air pollution in many countries have become more serious problem. The air quality has been continuously concerned by environmentalist as well as the common people. The objective of this project is to show the Air Quality Index on Azure maps to indicate deterioration of air quality so that people should be aware of their environment. The Proposed system can predict the Air quality of upcoming days and it can also identify the places whose AQI index can possibly become worse in upcoming time. Acting as a warning system for the people to take appropriate measures to prevent it from further deterioration. It is a web-based Platform so that people can have easy access with an interactive map visualization. The prediction model used in the project is Sessional Autoregressive Integrated Moving Average (ARIMA), which is widely used for forecasting methods.*

***Key Words*: AQI, SARIMA, Forecasting, Map visualization Azure, Air pollution.**

1. INTRODUCTION

Air pollution causes serious harm to human and animal health. Air pollution is caused by various factors that might occur naturally or result from specific human activities. Air Quality Index (AQI) is used to measure pollution present in Air. The index reflects a scale that ranges from 0 to 500. The higher the value of AQI, the greater the health risk associated. An AQI value that's lower than 50 indicates little to no risk, but a value that's 300 or higher means that the air is hazardous to everyone.

Azure Maps is a collection of geospatial services that use fresh mapping data to provide geographic context in web and mobile applications. The services include APIs for maps, vehicle routing, weather, and geofencing. Azure Maps also has a web SDK that you can use to display a map on a webpage. The Azure Maps web SDK has a wide variety of tools you can use to visualize spatial data on an interactive map on a webpage. Before we get started adding a map to a webpage, let's look at some of the capabilities of the SDK. The Azure Maps web SDK has a wide variety of tools you can use to visualize spatial data on an interactive map on a webpage. The project is about getting the data from third party API and displaying it on the webpage. The map will be displayed and the different AQI values will be marked with different colors.

In this project we will take a look at 24 Indian cities air pollution levels over the years as well as forecast the air pollution levels at the current rate of pollution for the entire country. We will also try to explain how it varies season to season. from the data given. We have used standard AQI - Air quality Index, as our measure for the air pollution levels

We will be predicting values of AQI for India through SARIMA model (Seasonal Autoregressive Integrated Moving Average). It is one of the most widely used forecasting methods for univariate time series data forecasting. The dataset will be taken from Kaggle to process the model and predict the AQI.

2. LITERATURE SURVEY

In this study Dong-Her Shih, Ting-Wei Wu, Wen-Xuan Liu and Po-Yuan Shih will be completed on Azure cloud computing platform using cloud services , according to the characteristics of air quality monitoring data , with Microsoft Azure Machine Learning service. The hourly updated data provided by the government is used to generate an alert / warning for deteriorating AQI. The data used in this study is from January 2016 to May 2018 in Taiwan and creates a prediction model of AQI pollutant concentration. Using this data, the model predicts the AQI warning for next six hours. It is expected that this study can help public to avoid approaching the areas with serious air pollution ,and to reduce the health hazards to individuals.[1]

This study has been done by Huan Li ,Hong Fan and Fei Yue Mao In this study A visual exploration method was proposed to analyze air pollution data, which enables rapid processing and multi - perspective exploration of air pollution data to reveal spatio - temporal patterns and basic relationships among multiple variables , The outcome of this project suggests a strong correlation existed between pollutants and wind speed and Temporal characteristics were found through visual analytics . From October to March and timely from 6:00 p.m. to 4:00 a.m. the concentrations are more serious. [2]

This paper presented by Appasani Geeta, Pasumarthi Sai Ramya chenikala and Sravani M Ramesh This chapter presents a study of analyzing the Air Quality Index using individual methods and tools and to interpret the significance of AQI in atmosphere. This empowers us to know the perception of various creators and gives a clear view of data interpretation and comparison between the various tools.[3]

This paper published by Vineeta Ajit Bhat, Asha S Manek and Pranay Mishra In this paper, They have proposed a novel machine learning based system to estimate dense air pollution using data collected from both from government monitoring sites and wireless sensor . They choose three regression models (Linear Regression, Random Forest Regression and Decision Tree Regression) and compared the estimation performances. They selected Random Forest Regression (RFR) as their machine learning algorithm because of low error index rate. They applied RFR algorithm in their system for the optimal air pollution levels estimation performance.[4]

In this study Pranav Shrirama and Srinivash Malladi proposed that Air pollution is a major problem for health, sometimes healthy people also experience health impacts from air pollution. Air pollution can cause respiratory irritation, asthma, breathing problem during outdoor work. Air polluted related health risks is depending on the current health problem and impact is also varies from person to person. The different health problem may cause by different air particles present in the air also consider the factor of pollutant type, the time duration of exposure, and air particles concentration. It is essential to measure the air quality parameter to identify the level of air pollution and every country has its own standard measuring parameters and guidelines of air quality index. [5]

In this Study Nimmathi Satheesh, N. Vallileka, S.Jai Ganesh, R.SivaKumar, N.Muthu Lakshmi has presented sharing the messages of Air Quality Index (AQI) in Metropolitan areas. Air pollution is monitored using air pollution sensors for outdoor environments and Cloud technology is used to display to common people via private cloud. Prototype development was aimed to create awareness and people engagement in restoring the environment back to its healthy state. By using this system, the people can interact with the environment sensors in the field of view to access and view latest and historic environment measurements. [6]

3. PROPOSED SYSTEM

We have shown the detailed system Architecture in Fig 2. AQI Prediction and optimization is a web app so that everyone can access it on hand. It has two main elements i.e. visualization of AQI data on azure maps which provides an interactive platform for the people to view air quality index of their region as well as the rest of the world. The second phase deals with prediction of Future AQI using SARIMA model. SARIMA is Seasonal ARIMA, or simply put, ARIMA model with an additional seasonal component. As mentioned, ARIMA is a statistical analysis model which uses time-series data to better understand the data set and to predict future trends. The SARIMA model equation looks like the following – Fig.1

$$SARIMA\underbrace{(p,d,q)}_{non-seasonal}\underbrace{(P,D,Q)_m}_{seasonal}$$

Fig -1: Sarima Notation

The parameters for these types of models are $\boldsymbol{p}$ and seasonal $\boldsymbol{P}$ indicate the number of AR terms (lags of the **stationary series**), $\boldsymbol{d}$ and seasonal **D** indicate differencing that must be done to **stationary series,** $\boldsymbol{q}$ and seasonal $\boldsymbol{Q}$ indicate the number of MA terms (lags of the forecast errors), **lag** indicates the seasonal length in the data.

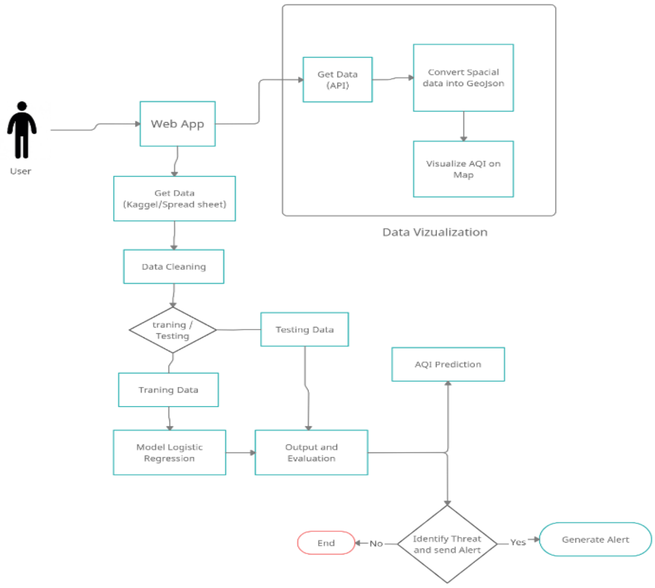

Fig. 2: System Architecture

The modules are:

3.1. Data Visualization.

For visualization process we will get the data through third party API (i.e.waqi.info)[7]. The data which we will get from the API will be in JSON format that will be converted to GeoJSON before plotting . Then we will use Azure maps to plot data on an interactive map . The AQI will be displayed using bubble layer. Different AQI levels will be plotted using different colors as per the standard chart.

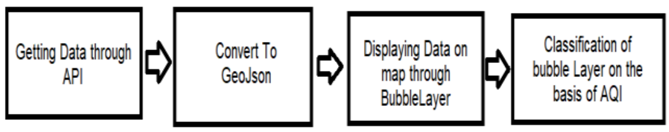

Fig. 3: Data Visualization

3.2. AQI Prediction

The data taken from Kaggle. Data is then sent for preprocessing where all the ambiguities are removed. Notice that the data is daily data. We will convert it into monthly data for our ease by averaging a month's data. After preprocessing the data is trained using SARIMA (Seasonal Autoregressive Integrated Moving Average) model which produces forecasting data as output. The output will be used as prediction for future possible AQI.

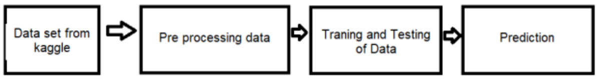

Fig. 4: Prediction model

3. Results and Discussion

The Pre-defined buckets of AQI levels are classified on the basis of the following table:

Good (0-50)	Minimal Impact	Poor (201-300)	Breathing discomfort to people on prolonged exposure
Satisfactory (51-100)	Minor breathing discomfort to sensitive people	Very Poor (301-400)	Respiratory illness to the people on prolonged exposure
Moderate (101-200)	Breathing discomfort to the people with lung, heart disease, children and older adults	Severe (>401)	Respiratory effects even on healthy people

Fig 5: Standard AQI

After converting Data from API into GeoJSON format the data is being displayed on the map. User can navigate on the map and check the AQI levels of their region as well as the whole world and can understand as well as take appropriate measures to prevent it from deteriorating or can get insightful information from the map visuals. Fig 6. displays the map that users will be able to access.

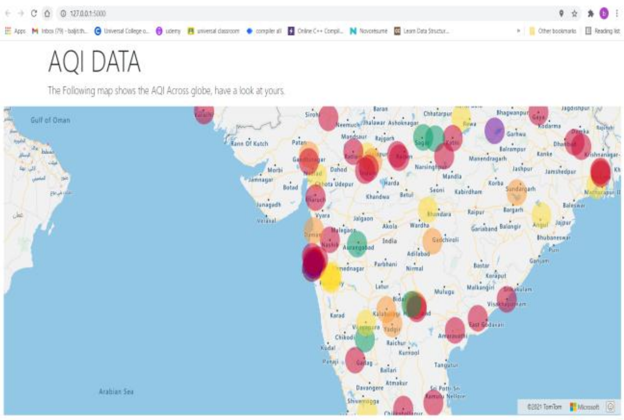

Fig.6: Map Visualization

In preprocessing data, we have removed all the ambiguities such as null values Fig. 7 represents Ambiguous data and Fig.8 represents processed clean data.

```
sns.heatmap(df.isnull(),yticklabels=False,cbar=False,cmap='viridis')
<matplotlib.axes._subplots.AxesSubplot at 0x7fd0ed4eb7d0>
```

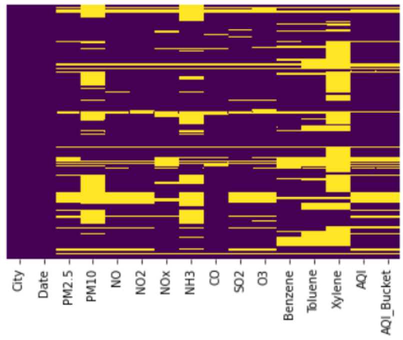

Fig. 7: Raw Dataset

```
sns.heatmap(df.isnull(),yticklabels=False,cbar=False,cmap='viridis')
<matplotlib.axes._subplots.AxesSubplot at 0x7fd0eae73950>
```

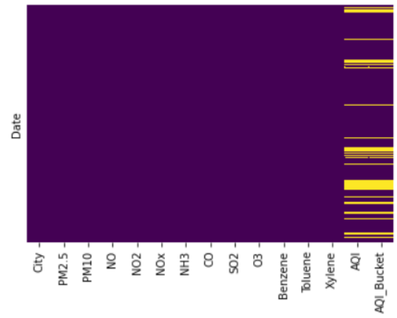

Fig.8: processed clean data

AQI calculation uses 7 measures: PM2.5(Particulate Matter 2.5-micrometer), PM10 (Particulate Matter 10-micrometer), SO2, NOx, NH3, CO and O3(ozone). For calculating the AQI PM2.5, PM10, SO2, NOx and NH3 the average value in last 24 hours is used and it must contain at least 16 values. For CO and O3 the maximum value in last eight hours is considered and then each measure is converted into a Sub-Index based on pre-defined groups.

The final AQI is the maximum sub-index with the condition stating that at least one of PM2 and PM10 should be available and at least three out of the seven pollutants should be available.

Fig 9. Shows the graph of India's AQI over the years.

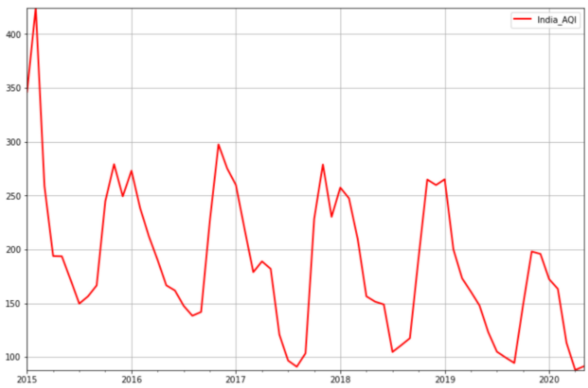

Fig.9 : India's AQI

We start by running auto ARIMA to find out the parameters of the model required. We can manually do it, however, it is much easier for us let the notebook do the work for us. Our next step is to forecast using this model into the future. However, since we do not have information regarding future values, we will split the data into a training data and testing data and try to predict 1 year into the future. We will use the years 2015-2020(till June) as our train dataset and July-June the next year as our test dataset. Further, we will predict into the year 2022.

We have provided our model with the training data and the necessary parameters. The next step is to forecast the next 12 months AQI values. To obtain the value of error we have used root mean square error (RMSE) for comparison between the models. We have got an RMSE value of approximately 21, which is quite good and we can approximately judge the scale of error by comparing with the mean values of AQI which is 177 and hence the error is approximately 1/9 of the actual values.

Next, we will proceed with forecasting model to predict into the unknown data, i.e., year 2021 and 2022. This poses a problem, as if we predict including 2020 data, we are bound to get an inaccurate prediction for next year simply due to the fact that 2020 is an outlier. But if we remove 2020 from our dataset and predict the values for 2021 and 2022. we are left with wrong predictions for sure and considering that covid-19 could have further lasting effects we will predict poorly. We will choose to include 2020 as well for this prediction. We could compare the values next year.

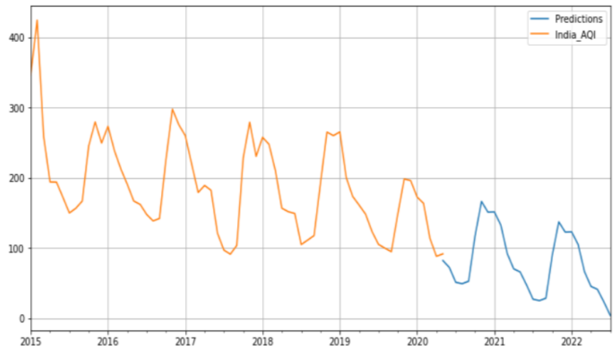

Fig 10: Forecasting model

Fig 10 shows the final model which displays the future prediction based on the previous year's data. The orange line in the graph represents the previous 5 years AQI values of India. The blue line represents the predicted values of AQI for the year 2021 and 2022. Here we can observe the seasonal pattern that repeats itself year after another. This signifies that there is a strong correlation between seasons and AQI and which is repeated every year with slight decline in trend. Here we can see decrease in AQI values form the year 2020 due to Covid-19 pandemic, which resulted in decrease in pollution. The decrease in pollution is due to Covid-19 restrictions which forced less movement of unnecessary vehicles and major pollution causing factories were closed.

4. CONCLUSION

Using Tracking of global air quality using azure maps and the proposed system, it is very much possible for the common civilians to access the AQI data and gain insights about their region and any other geographical location. We observed that the predicted values are close to our actual values using SARIMA and hence is quite interesting how looking at previous values gives us so much deeper insight into future air pollution. However, there is a discrepancy at the peak of the graph where our model has not been able to predict with a high accuracy. To obtain the value of error we will be using root mean square error(RMSE) for comparison between the models. This project will create an awareness among the common people and thus, people can take necessary measures to reduce the health-related issues caused by air pollution and identify the deteriorating environment

REFERENCES

[1] Dong-Her Shih, Ting-Wei Wu, Wen-Xuan Liu, Po-Yuan Shih, (2019), "An Azure ACES Early Warning System for Air Quality Index Deteriorating" , *MDPI*

[2] Huan Li, Hong Fan, Feiyue Mao, "A Visualization Approach to Air Pollution Data Exploration" ,(2016) , *RESERCH GATE*

[3] Appasani Geetha, Pasumarthi Sai Ramya, Chenikala Sravani, M.Ramesh,"Real Time Air Quality Index from Various Locations" ,(2020), *IJERT*

[4] Vineeta ,Ajit Bhat, Asha S Manek, Pranay Mishra, "Machine Learning based Prediction System for Detecting Air Pollution", (2020), *IJERT*

[5] Pranav Shrirama, Srinivas Malladi, " A Study and Analysis of Air Quality Index and Related Health Impact on Public Health" , (2020), *ICICNIS.*

[6] N. Vallileka, S.Jai Ganesh, Nimmathi Satheesh, R.SivaKumar, N.Muthu Lakshmi, "An Optimized Technique for Sharing Air Quality Index Using Cloud Technology" ,(2020) ,IJITEE.

[7] Microsoft docs: https://docs.microsoft.com/en-us/learn/modules/azure-maps

[8] An Azure ACES Early Warning System for Air Quality Index Deteriorating: http://dx.doi.org/10.3390/ijerph16234679

[9] A Visualization Approach to Air Pollution Data Exploration—A Case Study of Air Quality Index (PM2.5) in Beijing, China

[10] World Health Organization Regional Office for Europe. Monitoring Ambient Air Quality for Health Impact Assessment; WHO Regional Office for Europe: Copenhagen, Denmark, 1999.

[11] Air Quality index From Wikipedia, the free encyclopedia

[12] Accuweather reference API