Analysis Of Air Pollutants Affecting The Air Quality Using ARIMA

Akanksha Bhosale[1], Karuna Chaudhari[2], Komal Andhe[3], Farhat Naik[4], Sonal Chaudhari[5]

[1]Student, Computer Engineering, Datta Meghe College Of Engineering, Navi Mumbai, India [2]Student, Computer Engineering, Datta Meghe College Of Engineering, Navi Mumbai, India [3]Student, Computer Engineering, Datta Meghe College Of Engineering, Navi Mumbai, India [4]Student, Computer Engineering, Datta Meghe College Of Engineering, Navi Mumbai, India
[5]Assistant Professor, Computer Engineering, Datta Meghe College Of Engineering, Navi Mumbai, India

---

Abstract - *The reason for human life or any other living organism is air. However, this air is getting polluted because of a variety of factors, including traffic situations, industrial development and various construction businesses. Because of these factors, the quality of air is decreasing day by day, and all the life forms who depend on this air are gettingheavilyaffected. There are various factors who influence the quality of air, and it is measured by the Air Quality Index (AQI). NO2, CO, C6H6, SO2, and CO2 are some of the components which affect the quality of our air. The aim of our paper is to predict and forecast the AQI by using time series generalized models such as the Auto-Regressive Integrated Moving Average (ARIMA) model. The time series data collected has a lot of missing and corrupt values, and hence, it is subjected to cleaning, modification and aggregation as the requirement arose. The data is then checked for its stationarity, by performing various tests, and then the model is deployed. Prediction was performed on the aggregated data using ARIMA.*

***Key Words*: ARIMA, time-series data forecasting, Moving Average model, Auto-regressive**

1.INTRODUCTION

Air is one of the most crucial natural resources for all life on this planet's survival. Every life form depends on air for their existence and hence all the living beings require good air quality which is free from harmful gases for their existence. According to the Blacksmith Institute, two major pollution problems in the world are outdoor city air quality and indoor air pollution[1].

Air quality forecasting is conducted to obtain advanced knowledge of the air environment and to take preventative measures to avoid health problems. Pollution, which is found both indoors and outdoors, is causing the quality of air to deteriorate in emerging and even developed countries all over the world. Air pollution causes short-term and long-term health issues, mostly affecting the elderly and young children[2]. Short-term issues can include throat irritation, headaches, upper respiratory infections, and other short term but dangerous issues. Lung cancer, kidney damage, respiratory disease, heart disease, and brain damage are some of the long-term effects on health due to air pollution. Air pollution also causes depletion of the ozone layer, which is a major issue as it protects everyone from the sun's harmful UV rays [3]. Another harmful result of air pollution is acid rain, which affects rivers, trees, wildlife, and soils. Some of the environmental repercussions of air pollution include eutrophication, global warming, and haze. With the advent in technology, we aim to predict and forecast the Air Quality Index (AQI) by using unsupervised machine learning techniques. Primarily, Auto - Regressive Integrated Moving Average (ARIMA) model is used for air quality analysis.

The ARIMA model known as the auto-regressive integrated moving average is a model that includes two processes that are MA and AR. For developing the ARIMA(p,d,q) model both PACF and ACF auto- correlation functions are very useful.

1.1 Motivation

Air quality management has been acknowledged as a key issue at both national and local levels. In the past few years, research has been undertaken to identify the significant challenges because it is critical to ensure the health and cleanliness of the local surrounding and the community. It has been ascertained that different topic such as technological advancement, detail procedure regarding the measurement of air quality, determination of pollutant variable and its interconnection, identifying the causes and effects of air pollution and lastly forecasting of periodic and geographical variations in atmospheric levels are some of the topics that has been covered in regard to the study of air quality. Although, it has been investigated that developing countries are facing these air quality issues due to insufficient funds, and technological support.

Air pollution in a city requires immediate attention as a large number of people live in the city, and hence, air pollution may affect more people. There is an immediate need of strict laws and constant monitoring of air pollutants for air pollution management[4]. For this reason, our paper focuses on a prediction model that will help us with prediction of the air pollution.

2. LITERATURE REVIEW

A paper presented a neural calibration for benzene concentration prediction utilizing a gas multi-sensor system (solid-state) developed to measure urban pollution. The results are evaluated and analyzed using prediction error characterization during a 13-month period. The relationship between training duration and efficiency is also being

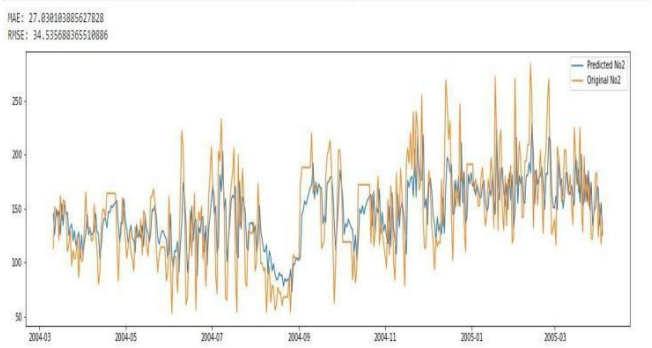

Fig-8: Predicted and Original values of NO2

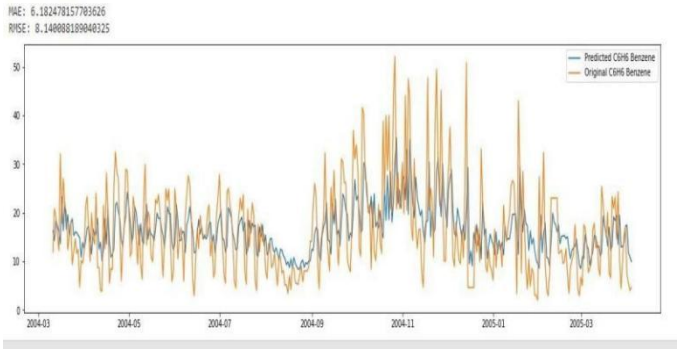

Fig-9: Predicted and Original values of C6H6

We can see that the accuracy of the model has improved drastically, once we use the p, d and q values, which are the hyperparameters of the ARIMA model. We can raise the accuracy by changing the parameters p, d, q accordingly as the data requires.

ARIMA Future Forecast :

On the basis of the train data, we can forecast the future values using ARIMA model. Using the parameter values and the accuracy metrics we can forecast for a whole year and use this forecasting to understand how this can impact the environment in the upcoming years. Currently, we have forecasted for only one year as we have only yearly amount of data.

Fig-10: Forecasting for NO2 concentration

```
2005-04-04 18:00:00    153.477264
2005-04-05 18:00:00    148.090201
2005-04-06 18:00:00    149.974908
2005-04-07 18:00:00    148.738107
2005-04-08 18:00:00    145.199518
2005-04-09 18:00:00    145.191795
2005-04-10 18:00:00    145.500353
2005-04-11 18:00:00    145.638852
2005-04-12 18:00:00    146.161360
2005-04-13 18:00:00    146.280792
2005-04-14 18:00:00    146.136716
2005-04-15 18:00:00    146.030059
2005-04-16 18:00:00    145.912900
2005-04-17 18:00:00    145.842850
2005-04-18 18:00:00    145.836065
2005-04-19 18:00:00    145.835529
Freq: D, dtype: float64
<Figure size 720x360 with 0 Axes>
```

Fig-11: Forecasting for C6H6 concentration

Final prediction for forecast and original values is done using ARIMA model with the predicted values. The p=5, d=0, q=1 and the parameters are set the same. To get more accuracy and exact forecasting of the data, the accuracy metrics of MAE and RMSE would have to be optimized using differentiation. We can further

optimize the forecast by increasing the differentiation and accuracy metrics.

7.FUTURE SCOPE

The prediction model can be improved by strengthening the methods to forecast the concentration of air quality factors, majorly for O3, as O3 does not come from direct sources but due to multiple sources of emission and their reaction to each other. There are multiple time series models which can be used for this. The time series data can be collected for two or three years, or more than that, and we can work on that data in order to make more accurate predictions.

8.CONCLUSIONS

ARIMA model is suitable for short-term predictions because with the help of stationary data, accurate predictions can be made. Time series model used in forecasting is an important tool which helps us to control, analyse and monitor the air quality condition. It is useful to take quick action before the situation worsens in the long run [10]. For that reason, we need our model performance to be as accurate as possible so that good air quality forecasting can be achieved. Moreover, the pollutants must be considered in analysis of air pollution data.

REFERENCES

[1] Blacksmith Institute Press Release'. (October 21, 2008). [Online]. Available: http://www.blacksmithinstitute.org/the-2008-top- ten-list-of-world-s-worst-pollution-problems.html

[2] Huang, W., Li, T., Liu, J., Xie, P., Du, S. and Teng, F. (2021) An overview of air quality analysis by big data techniques: Monitoring, forecasting, and traceability. Information Fusion 75, 28–40.

[3] Baralis, E., Cerquitelli, T., Chiusano, S., Garza, P. and Kavoosifar, M.R. (2016) Analyzing air pollution on the urban environment. 1464 1469https: //ieeexplore.ieee.org/abstract/document/7 522370 Accessed 14 May 2022.

[4] Abhilash, M.S.K., Thakur, A., Gupta, D., Sreevidya, B. (2018). Time Series Analysis of Air Pollution in Bengaluru Using ARIMA Model. In: Perez, G., Tiwari, S., Trivedi, M., Mishra, K.(eds) Ambient Communications and Computer Systems. Advances in Intelligent Systems and Computing, vol 696. Springer, Singapore. https://doi.org/10.1007/978-981-10-7386-1_36

[5] De Vito, S., Massera, E., Piga, M., Martinotto, L. and Di Francia, G. (2008) On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. Sensors and Actuators B: Chemical 129 (2), 750–757.

[6] Nandini, K. and Fathima, G. (2019) Urban Air Quality Analysis and Prediction Using Machine Learning. 98–102

https://ieeexplore.ieee.org/document/9063845 Accessed 14 May 2022.

[7] Cheung, Y., and Xu, L. (2001). Independent component ordering in ICA time series analysis. Neurocomputing, 41, 145–152

[8].https://en.wikipedia.org/wiki/Python_(programming_language)

[9].https://research.google.com/colaboratory/faq.html [10]. Lee, M. H., Rahman, N. H. A., Suhartono, Latif, M. T., Nor, M. E. & Kamisan, N. A. B. (2012). Seasonal ARIMA for Forecasting Air Pollution Index: A Case Study. *American Journal of Applied Sciences*, *9*(4), 570-578. https://doi.org/10.3844/ajassp.2012.570.578.