2023-05

Time Series Analysis and Forecasting of Air Quality Index of Dhaka City of Bangladesh

Sakib, Sheikh Rahmatulla

Independent University, Bangladesh

Draft Version

Time Series Analysis and Forecasting of Air Quality Index of Dhaka City of Bangladesh

Sheikh Rahmatulla Sakib[1], Kamarun Nahar Sara[1], Md. Tahmid Hossain Rasel[1], Md. Masudul Islam Asif[1], Md. Aynul Hasan Nahid[2], Md. Saifur Rahman[1], M. F. Mridha[3], Ashraful Islam[4]
[1]Department of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh
[2]Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh
[3]Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh
[4]Center for Computational and Data Sciences, Independent University, Bangladesh, Dhaka, Bangladesh
Corresponding Author: Ashraful Islam (ashraful@iub.edu.bd)

***Abstract*—In Dhaka, the capital city of Bangladesh, various sources including vehicle emissions, industrial activities, brick kilns, building sites, and open rubbish burning contribute to the air pollution problem. To assess the air quality, the Air Quality Index (AQI) is utilized, which categorizes air quality based on pollutant concentration. In this study, we have built ARIMA, Auto-ARIMA, SARIMAX, and VAR models to predict the air quality of Dhaka. Unlike previous studies, we have utilized hourly air pollutants factors such as $PM_{2.5}$, PM_{10}, SO_2, CO, NO_2, and O_3 to forecast air quality. Our novel approach enables us to predict the monthly and weekly air quality of Dhaka city. Our analysis reveals that the SARIMAX model, which takes into account seasonal patterns, trends, and external factors, is the most accurate in predicting Dhaka city's air quality. The model's prediction performance is assessed using statistical indicators such as mean absolute percentage error and root mean square error. The study highlights that the SARIMAX model could aid policymakers in evaluating the efficacy of air pollution control measures.**

***Index Terms*—AQI, Machine Learning, Time Series Data, RMSE, MAPE, Forecasting**

I. Introduction

Time series analysis has emerged as a powerful tool for air quality prediction. Dhaka is known for its poor air quality, which has a negative impact on human health and the environment. With an air quality index (AQI) of 297 as of 8:50 am on March 2, 2023, Dhaka ranked first in the list of the most polluted cities [1]. In Bangladesh, the AQI is based on 5 pollution criteria; Particulate matter ($PM_{2.5}$ and PM_{10}), NO_2, CO, SO_2 and Ozone (O_3) [2].

According to the World Health Organization (WHO), air contamination kills around 7,000,000 individuals overall every year, essentially because of expanded mortality from stroke, coronary illness, persistent obstructive aspiratory sickness, cellular breakdown in the lungs, and intense respiratory diseases [3]. The pollutants with the strongest evidence of adverse health effects are particulate matter (PM), ozone (O_3), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2). In particular, PM is an air pollutant produced by coal burning and vehicle emissions. The chemical composition, distribution of particle sizes, and physical and chemical properties all undergo significant changes when PM enters the atmosphere, leading to several biological effects on human health. Therefore, all countries and major cities must carefully monitor and regulate atmospheric $PM_{2.5}$ levels to maintain a safe environment and promote everyone's health.

Time series analysis and forecasting using Seasonal Auto-Regressive Integrated Moving Average (SARIMAX), Vector autoregressive (VAR) is an exciting and powerful technique that has the potential to revolutionize how to understand and plan for the future of Dhaka city [4]. By analyzing and modeling key indicators over time, this research can uncover hidden patterns and trends that can be used to make accurate predictions about what the future holds. In recent years, researchers have employed various time series models including machine learning and deep learning [5] to predict air quality in Dhaka city.

This topic aims to develop a robust and accurate model for time series analysis and air quality prediction in Dhaka to help policymakers reduce air pollution and improve public health. The model should account for air pollution characteristics as traffic congestion, industrial activity, and weather. The model may help policymakers decrease air pollution and minimize its health and environmental effects by assessing previous data and projecting future trends.

The SARIMAX, VAR model is a sophisticated model that allows us to capture the complex relationships between multiple time series variables, including those that have a lagged effect on one another [6]. This research provided valuable insights into the future trajectory of important indicators such as population growth, economic trends, and traffic congestion. This information can be used to inform policy decisions and planning efforts that will shape the future of Dhaka city in a positive way.
The key contributions of our paper are:

- This study is based on statistical tests to check the effect of pollutants on AQI. checking outliers, correlation, and visual impacts of the data.
- Plotted ACF and PACF graphs to get the order of ARIMA model. Then fitted Auto-ARIMA model for getting the best order. The orders given by Auto-ARIMA are utilized to build a SARIMAX model.

- The methodology used in the study, Univariate analysis is done using seasonal Auto-Regressive Integrated Moving Average and Multivariate analysis is done using VAR.
- Finally, a comparative analysis for both Monthwise and Weekwise forecasting is done after fitting all the models.

II. Related Work

In a recent study by Islam et al. [7] predicted the levels of $PM_{2.5}$ and air quality in Dhaka, Bangladesh, using a number of machine learning models. Four models - RT, AR, REPTree, and RSS - were utilized to accomplish the exploration targets. In terms of R^2, MSE, and RMSE evaluation metrics, RSS outperformed all of the Gazipur and Darussalam sites. Their research demonstrated that the RSS model outperforms other models when it comes to predicting $PM_{2.5}$ concentrations in two locations.

Ahmed et al. [8] used the SARIMA model to predict air quality in the Dhaka and Sylhet regions of Bangladesh, this particular study clearly shows air pollution trends in two regions of the country and can predict air quality Mass seasonal differences are still healthy and unhealthy. Chowdhury et al. [9] employed LSTM and other machine learning models to estimate Dhaka's air quality. They also show how machine learning and deep learning models can categorize and forecast AQI values using specified domains. Best-case Random Forest [10] outperforms other models.

Atik et al. [11] employed machine learning to construct a model that can forecast future pollution levels, detect air pollution variables, and assess local air pollution. The review used straight relapse, Facebook Prophet, RNN, and ARIMA calculations. They employ RNNs and an LSTM model to forecast using particular and standard entities [12]. They use numerous computations to get the best results and fewer errors. The ARIMA model consistently outperformed the others. Emam et al. [13] implemented a Dhaka-Chittagong air quality index prediction model. GRU was the hidden model's initial layer, followed by LSTM and two dense layers. After data collection and processing, the model is trained on 80% of the data and verified on the rest. MSE, RMSE, and MAE indicate the model's error. Their model predicted AQI trends in both locations. They also tested their hybrid model against a GRU model and an LSTM model. Combining these two models improved the model's performance while shallow neural network [14] plays a vital role in terms of air quality.

Again, Afrin et al. [15] outlined a Dhaka, Bangladesh, PM model. The model predicts PM levels 72 hours ahead using meteorological data, land use patterns, and seasonal fluctuations. The model accurately predicted PM concentrations using real PM data, suggesting it might influence megacity public health initiatives and air quality management regulations. Shahriar et al. [16] explored Bangladeshi PM and air quality management using ML models. The research used five models. GPR has the best R_2, RMSE, and MAE scores. For Sylhet and Chattogram, $PM_{2.5}$ proposed ANN. PM10 utilized M-SVM and L-SVM. The research advises long-term data use. Time series analysis helps Dhaka air quality forecasters. ARIMA, SARIMA, and VAR time series models predicted Dhaka city air quality. The ARIMA and SARIMA models forecast PM2.5 and PM10 levels in Dhaka well.

Outside of Bangladesh, cities in India and China are hit hard by air pollution. Many studies are conducted by multiple researchers. Rambha et al. [17] proposed a VAR-LSTM model to enhance air quality prediction. The suggested strategy improves prediction due to VAR-LSTM model feature selection. The suggested VAR-LSTM model has 6.932 MSE, SARIMA 8.386 MSE, and ARIMA 15.46 MSE. This technology will be used to minimize air quality forecast errors using meteorological data.

Tomar et al. [18] modeled daily AQI prediction. They predict AQI using AR, ARIMA, and SARIMA autoregressive models. Use MAE, RMSE, and other metrics to assess model performance. The results show that the autoregressive model works very well. Nitrogen dioxide and particulate matter affected the city's air quality index, the study found. Yansui et al. [19] analyzed of VAR model in terms of meteorological factors and Beijing air quality. The VAR model can accurately forecast short-term air quality by capturing the dynamic interplay between meteorological conditions and air quality.

III. Dataset

A. Data Collection

The dataset used in this study is the Dhaka City (2016-2023) $PM_{2.5}$ (Hourly) Air Quality Index dataset available at airnow.gov[1]. This is a website administered by the United States of America. Environmental Protection Agency (EPA), provides information on air quality around the world. It provides real-time air quality data from thousands of monitoring stations across the city. Measurements of ozone, particulate matter, carbon monoxide, sulfur dioxide, and nitrogen dioxide are available on the website. This dataset contains information on AQI's performance in autonomous surveys from 2016 to 2023. The dataset consists of 58212 instances and 7 attributes, including categorical and numerical data. Missing data are imputed using the mean for each variable.

B. Dataset Analysis

$PM_{2.5}$ refers to atmospheric particles within a certain size range and has been listed as an important air pollutant by regulatory bodies around the world. However, the Air Quality Index (AQI) is calculated based on the concentrations of several air pollutants ozone (O_3), carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), particulate matter ($PM_{2.5}$ and PM_{10}). . from). The NowCast concentrations in this dataset are intended to provide estimates of current air

pollution levels, taking into account recent changes in air quality. NowCast PM2.5 concentration calculation formula is:

$$\mathrm{N}owCastConc = (L_1 \times 0.5) + (L_2 \times 0.25) + (L_3 \times 0.125) + (L_4 \times 0.6225) + (L_5 \times 0.031255) \tag{1}$$

[1] https://www.airnow.gov/,_(Airgov)

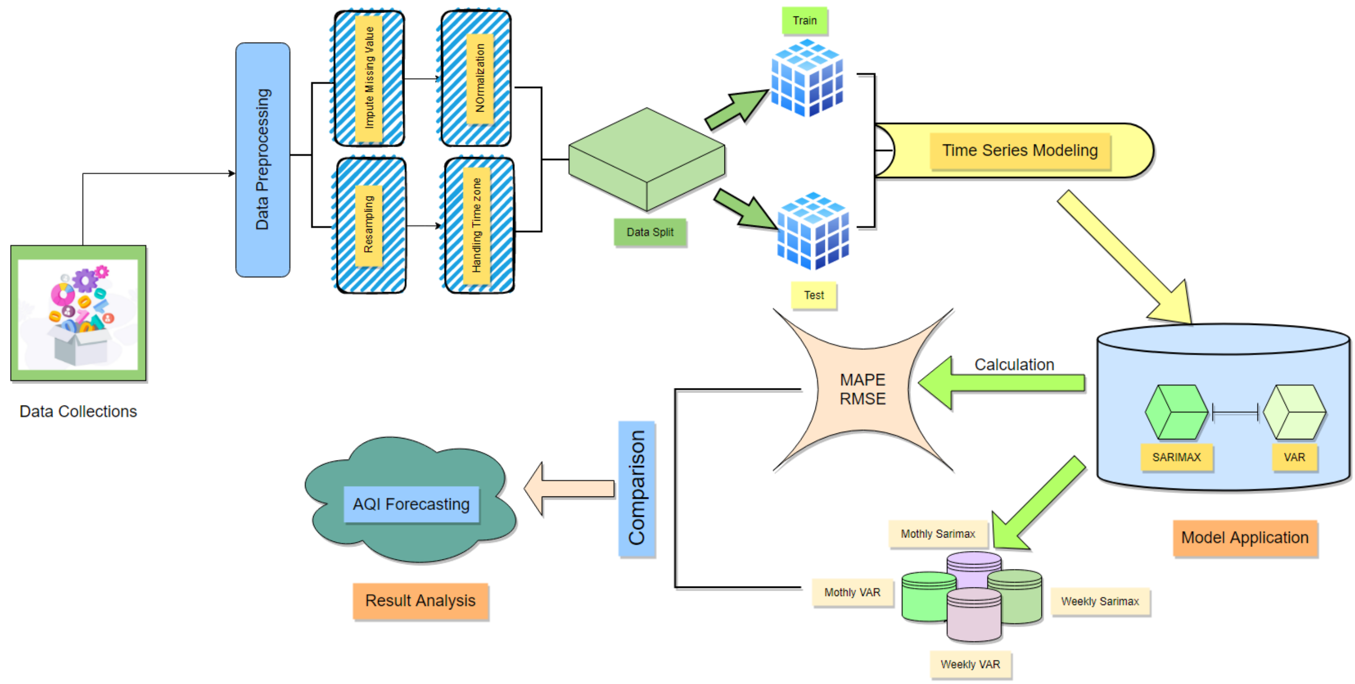

Fig. 1: Recommended design of SARIMAX & VAR model For Time-Series Forecasting

where L_1 is the current hourly $PM_{2.5}$ concentration and L_2 is the PM2 concentration. 5. One hour ago, L_3 was the $PM_{2.5}$ concentration two hours ago, L_4 is the $PM_{2.5}$ concentration three hours ago, and L_5 is the $PM_{2.5}$ concentration four hours ago. The Raw Concentration of pollutants refers to the actually measured concentration of pollutants in the air at a given place and time. Raw Concentration = (mass of pollutant) / (volume of air) To calculate the AQI, first calculate the concentration of each pollutant of interest (Raw or NowCast) over a period of time (eg, 24 hours). Once you have a concentration value, you can use the AQI formula and the cutoff point specific to that pollutant to calculate the AQI value. The AQI formula and breakpoints are:

$$\mathrm{AQI} = \frac{I_{Hi} - I_{LO}}{B_{Hi} - B_{LO} \times C - B_{LO}} + I_{LO} \tag{2}$$

I_{Hi} and I_{LO} are the AQI values for the high and low end of the interval breakpoint, respectively. B_{Hi} and B_{LO} are the upper and lower limits of the range, respectively. C is the $PM_{2.5}$ raw or NowCast concentration.

IV. Methodology

Research model is described here. Figure 1 shows how textual data is acquired, preprocessed, and normalized. After data processing, a time series modeling function using machine learning and statistical models receives training and testing data. Monthly and weekly projections were compared between these models. The final AQI forecast is created.

A. Data preprocessing

Prepare time series data for analysis and modeling. To measure missing values, we approximated time-series data missing values and showed them in a heat map. The dataset cannot remove missing rows due to missing timestamps. Data loss happens annually. Value is time-based. The data collection occasionally contains missing values for 3-4 days, therefore datetime and mean cannot be added. Fill gaps using one-day averages. Imputation neither forward nor backward. The mean and standard deviation were unaffected by missing data. Our study is shown in Figure 1.

B. Outlier Analysis

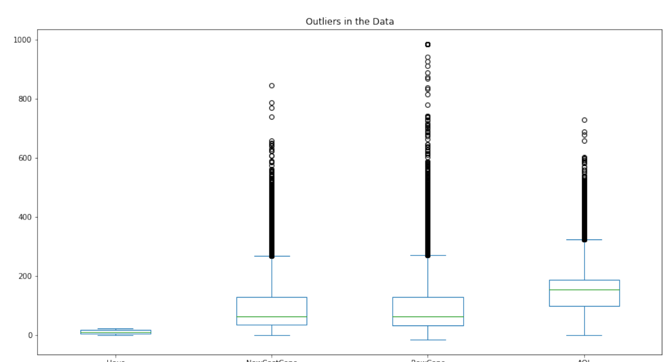

Fig. 2: Checking for outliers

Here from the boxplot that there are outliers in the data, but this is time series data. Outliers cannot be removed, if doing so, the data will not be continuous. Figure 2 shows the outliers of the dataset.

C. Skewness in Data

Time series skewness measures data point distribution asymmetry. Trends, seasonal patterns, and outliers may create time

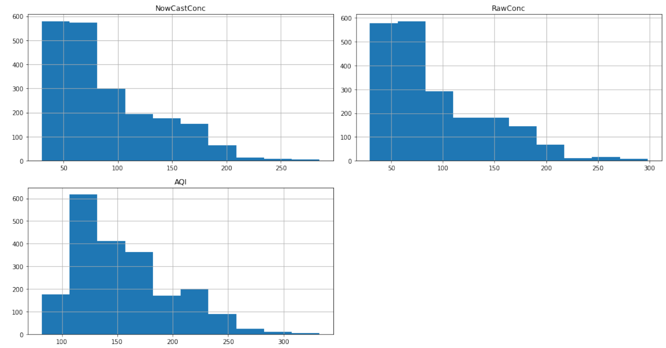

Fig. 3: Checking for outliers

series outliers. With AQI time series data with a positive skew, most data points are at the low end of the scale, but there are some extreme values at the high end. If there are outliers or high pollution levels on specific days, the AQI readings might drift right. Figure 3 shows the different skewness of outliers. From the plot and skewness, NowCastConc and RawConc can be said to be highly skewed and right-skewed.

D. *Correlation Analysis*

From the scatterplot, NowCastConc. RawConc is positively correlated with AQI. This means that as the pollutants increase, so does the AQI. Figure 4 and 5 depicts the graph of NowCastConc and RawConc. Using the heat map to find the correlation between the other variables. Heatmaps for correlation analysis are particularly useful for identifying which variables in a data set are highly correlated and may significantly affect the current analysis or modeling task. Once the matrix is calculated, it can be visualized using a heatmap function such as the Seaborn heatmap function in Python [20]. Figure 6 shows the heatmap. From the heat map, can tell that NowCastConc and RawConc have a positive and strong correlation.

E. *Time Series Decomposition*

Preprocessing of data in a time series dataset refers to the process of cleaning, transforming, and preparing the data for further analysis and modeling [21]. First, computed the missing values in the time-series data and visualized them in a heat map to assess the degree of missingness and decide on an appropriate attribution method. The technique of time series decomposition is used to break down a time series into its fundamental components, such as trends, seasonality, and residual volatility. Breaking down a time series into components helps identify underlying patterns and trends in the raw data that may not be immediately apparent. The three main components of time series decomposition are:

Trend: The trend component represents the long-term direction or pattern of data over time. It may increase, decrease or remain stable over time.

Seasonal: The seasonal component represents periodic data patterns or cycles that occur at regular intervals, such as daily, weekly, or monthly.

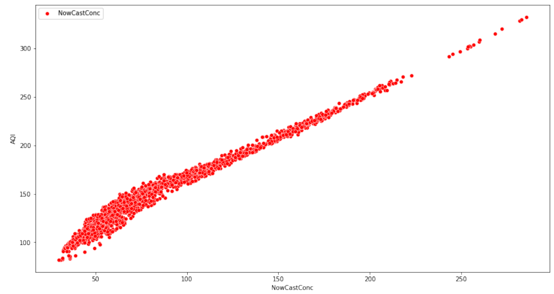

Fig. 4: impact of NowCastConc on AQI

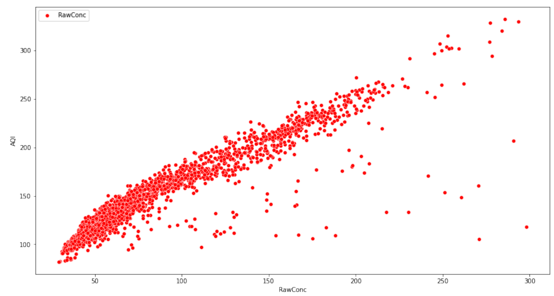

Fig. 5: impact of RawConc on AQI

Residuals: The residual component represents the residual volatility or noise in the data that cannot be explained by trend or seasonality.
The uppermost graph gives us Observed values in the data.

1) Trend: The next three graphs are Trend, Seasonality, and Residuals. By looking at the trend in the data, can see that the trend is decreasing gradually from 2019 to 2020 because of the Covid pandemic and there is a sudden increase after 2021.

2) Seasonality: Variations that occur at particularly regular intervals that are shorter than a year are referred to as seasonality. Seasonality is characterized by periodic, repetitive, and generally regular and predictable patterns in the levels of a time series. It can be caused by the weather. Here, in the mid-year, the AQI value is decreasing.

3) Residual: The data points which don't follow trends as well as seasonality, are plotted in the residual graph.

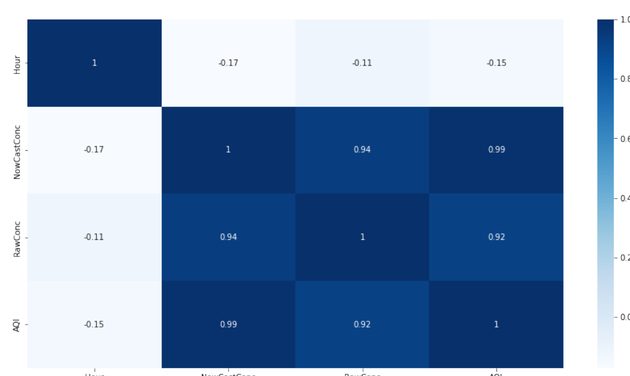

Fig. 6: Correlation of the numerical data with heatmap

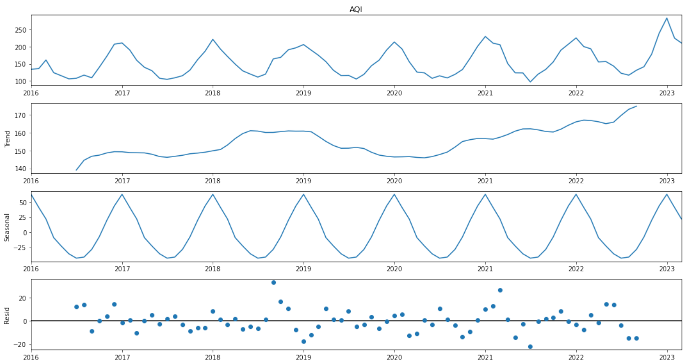

Fig. 7: Decomposition Plot

F. ACF and PACF plot

The terms “Autocorrelation Function” and “Partial Autocorrelation Function” are used interchangeably. Lag values for the AR and MA models are provided by the ACF and PACF plots. In Figure 7, the autocorrelation value for any series along with its lag can be found in the (complete) autocorrelation function known as ACF.

1) PACF: In Fig: 8 Partial autocorrelation calculates the “pure” correlation between xt and xt2 by removing the “transitive” correlation (that is, the amount of correlation explained by the first lag) and recalculating. For partial autocorrelation between xt and xt-3, remove the correlation with xt-1 and xt-2 and recalculate, and so on. The amount of correlation between a variable and its lag that cannot be explained by all lower-order lag is called partial autocorrelation.

G. Proposed Methodology

Air Quality Index (AQI) forecasting using time series analysis involves analyzing historical AQI data to develop models that predict future AQI values.

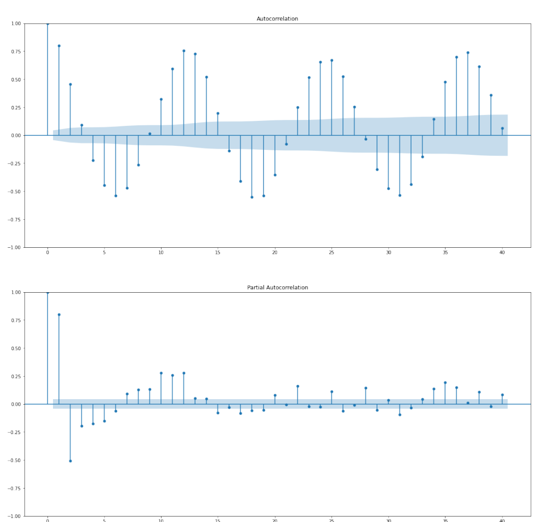

Fig. 8: ACF and PACF plot

1) ACF: This involves identifying patterns and trends in the data, such as seasonality and trends, and developing models that can capture these patterns. The model can then be used to predict AQI values for a specific period of time, such as a week or month.

There are several methods for time series analysis and AQI prediction, including autoregressive integrated moving average (ARIMA) base model, seasonally autoregressive integrated moving average (SARIMA), and vector autoregressive (VAR). These models can be used to capture the temporal relationship between AQI and its pollutants, as well as other variables such as weather, traffic volume, and industrial activity.

H. Model Description

1) ARIMA Model: ARIMA stands for Autoregressive Integrated Moving Average. It is a well-liked time series model that can be used to model and predict time series data [22]. A subset of linear models known as ARIMA models makes the assumption that a combination of autoregressive (AR), discriminant (I), and moving average (MA) terms can approximate the underlying process that generates the time series.

$$\text{AR} \;= \hat{A_t} = c + \Phi_1 A_{t-1} + \Phi_2 A_{t-2} + ... + \Phi_p A_{t-p} + \varepsilon_t \tag{3}$$

$$\text{MA} \;= \hat{A_t} = c + \Theta_1 \epsilon_{t-1} + \Theta_2 \epsilon_{t-2} + ... + \Theta_q \epsilon_{t-q} \tag{4}$$

Combining all three types of models mentioned above results in an ARIMA(p,d,q) model.

$$\hat{A'} = c + \Phi_1 A'_{t-1} + \Phi_2 A'_{t-2} + ... + \Phi_p A'_{t-p} + \Theta_1 \epsilon_{t-1} + \Theta_2 \epsilon_{t-2} + ... + \Theta_q \epsilon_{t-q} + \epsilon_t \tag{5}$$

2) Auto-ARIMA Model: Auto-ARIMA models are an extension of ARIMA models that automatically select optimal values for the p, d, and q parameters [23]. Auto-ARIMA models search for the optimal parameter combination that minimizes either the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) using a grid search algorithm.

3) SARIMAX Model: An extension of the ARIMA model with a seasonal component and exogenous variables is the SARIMAX model (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) [24]. Six parameters define the SARIMAX model: P, D, and Q are the seasonal autoregressive, seasonal difference, and seasonal mean, respectively, of the seasonal variable.

4) VAR Model: The statistical model known as a vector autoregression (VAR) model is used to investigate the connection that exists between various time series variables [25].

$$\text{A}_t = c + B_1 y_{t-1} + B_2 y_{t-2} + ... + B_p y_{t-p} + e_t \tag{6}$$

Each variable in the VAR model is assumed to be a linear function of its previous value as well as the past values of the other variables in the model. A VAR model is defined by the number of lag values included in the model and can be used to estimate the dynamic interaction and causality of the variables in the model. VAR models are commonly used in macroeconomic modeling and financial forecasting.

V. Result Analysis & Discussion

A. Define Accuracy Matrix

1) RMSE: The standard deviation of the residual (prediction error) is the root mean square error (RMSE) [26].

$$\text{RMSE } = \frac{1}{m}\sqrt{\sum_{k=1}^{m}\left(Z_o - Y_o\right)^2} \tag{7}$$

The regression residuals are used to calculate the slope of the regression line and the RMSE is used to calculate the distribution. In other words, it shows how close the dataset is to the line of best fit.

2) MAPE: Mean absolute percentage error (MAP), also known as mean absolute percentage deviation (MAPD), is a statistical forecasting method that can be used to evaluate trends [27]. It is also used as a loss function in machine regression problems.

$$\text{MAPE } = \frac{1}{m}\sum_{k=1}^{m}\left|\frac{Z_o - Y_o}{Z_o}\right| \times 100 \tag{8}$$

The basic model for the univariate analysis of the AQI column using ARIMA (1,1,1) gives good results, as indicated by p-values less than 0.05 for AR.L1 and MA.L1. The graphical representation of monthly forecasting and weekly forecasting is depicted in Figures 9 and 10 accordingly. However, using Auto ARIMA to find the best order, and the best model is ARIMA(3,0,1), which has a lower AIC value than the original ARIMA model. To consider the seasonality in the data, This research built a SARIMAX model on the basis of the ARIMA order. The SARIMAX model provided a forecast for the valid set and future values of AQI for 2023 month-wise Unvariate. where the forecast accuracy is 'mape': 10.88609557521429, 'rmse': 29.41702091207996. And after the statistical analysis of VAR model (month-wise - multivariate), individually forecast the accuracy of every column:

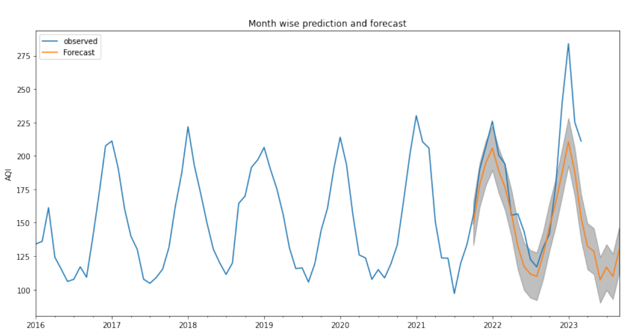

Fig. 9: Monthly-wise Forecasting

In the weekly analysis using Auto ARIMA, the best model was found to be ARIMA(5,0,1). Then built a SARIMAX model based on

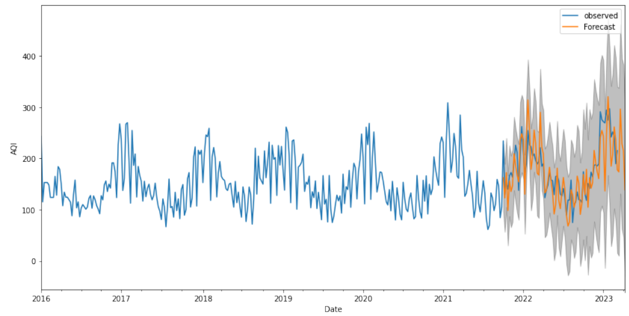

Fig. 10: Week-wise Forecasting

TABLE I: Individually forecast accuracy of every column

Column Name	MAPE	RMSE
NowCastConc	38.837414541791325	62.80987828981185
RawConc	39.14586865727524	60.2094452849837
AQI	23.04094161586495	57.95783252328217

this sequence and found that it predicts AQI very well, the best sequence is (1, 1, 1) and the RMSE and MAPE values are relatively high, where the forecast accuracy is 'mape': 15.322580803502529, 'rmse': 20.914404072797943 And after the statistical analysis of VAR model (week wise - multivariate), individually forecast the accuracy of every column: The actual predicted forecast accuracy is 'mape': 24.466374, 'rmse': 61.269930

B. Comparison

From the Table I, analyzed the SARIMAX model appears to be performing well for predicting AQI, and the results of our analysis suggest that there is some seasonality in the data that needs to be accounted for. Further analysis and modeling may be necessary to improve the accuracy of our predictions and gain a deeper understanding of the underlying patterns and trends in the data from both Table II and Table III.

VI. Conclusion

Dhaka's $PM_{2.5}$ pollution requires air quality index (AQI) calculations for a healthy and sustainable environment. Results-based SARIMAX and VAR models assessed AQI time series. SARIMAX predicted Dhaka AQI better than

TABLE II: Individually forecast accuracy of every column

Column Name	MAPE	RMSE
NowCastConc	47.44999415969162	66.83425010691313
RawConc	52.560670795138854	60.46391337660701
AQI	24.466373808858442	61.26992992089746

MAPE/RMSE models. This research utilized Bangladeshi data. Machine learning researchers may benefit from this study's AQI predictions of higher mortality from acute respiratory infections, heart disease, lung cancer, and stroke. These results corroborate the SARIMAX time series model. Machine

TABLE III: Comparison Between SARIMAX & VAR Model

Model Name	MAPE	RMSE
SARIMAX Month Wise	10.886096	29.417021
VAR Month Wise	23.040942	57.957833
SARIMAX Week Wise	16.718502	36.587722
VAR Week Wise	24.466374	61.269930

learning and sophisticated modeling may enhance Dhaka city AQI estimates. Deep learning and ensemble approaches may increase AQI prediction accuracy.

REFERENCES

[1] New AGE. Dhaka's air continues to be most polluted in the world, 2023. Accessed on March 3, 2023.

[2] Ministry of Environment and Forest. About air quality index (AQI) in Bangladesh.

[3] World Health Organization(WHO). Compendium of WHO and other un guidance on health and environment, 2022. Accessed on January 2, 2022.

[4] Brendan Artley. Time Series Forecasting with ARIMA, SARIMA, and SARIMAX, 2022. Accessed on April 27, 2022.

[5] Khan Md Hasib, Md Rafiqul Islam, Shadman Sakib, Md. Ali Akbar, Imran Razzak, and Mohammad Shafiul Alam. Depression detection from social networks data based on machine learning and deep learning techniques: An interrogative survey. *IEEE Transactions on Computational Social Systems*, pages 1–19, 2023.

[6] Sarit Maitra. Time-series Analysis with VAR VECM: Statistical approach, 2019. Accessed on November 13, 2019.

[7] Abu Reza Md Towfiqul Islam, Mohammed Al Awadh, Javed Mallick, Subodh Chandra Pal, Rabin Chakraborty, Md Abdul Fattah, Bonosri Ghose, Most Kulsuma Akther Kakoli, Md Aminul Islam, Hasan Raja Naqvi, et al. Estimating ground-level pm2. 5 using subset regression model and machine learning algorithms in asian megacity, dhaka, bangladesh. *Air Quality, Atmosphere & Health*, pages 1–23, 2023.

[8] MM Islam, M Sharmin, and F Ahmed. Predicting air quality of dhaka and sylhet divisions in bangladesh: a time series modeling approach. air qual atmos health 13: 607–615, 2020.

[9] Al-Sadman Chowdhury, Md Shihab Uddin, Md Rashad Tanjim, Fariha Noor, and Rashedur M Rahman. Application of data mining techniques on air pollution of dhaka city. In *2020 IEEE 10th International Conference on Intelligent Systems (IS)*, pages 562–567. IEEE, 2020.

[10] Khan Md Hasib, Md Iqbal, Faisal Muhammad Shah, Jubayer Al Mahmud, Mahmudul Hasan Popel, Md Showrov, Imran Hossain, Shakil Ahmed, Obaidur Rahman, et al. A survey of methods for managing the classification and solution of data imbalance problem. *arXiv preprint arXiv:2012.11870*, 2020.

[11] Sk Sihan, Atik Tajwar, Maisha Rabbani, Manish Agarwala, Sanjida Alam Maliha, et al. *Analyzing area-wise air pollution level using machine learning for a better future*. PhD thesis, Brac University, 2021.

[12] Khan Md Hasib, Nurul Akter Towhid, and Md Rafiqul Islam. Hsdlm: a hybrid sampling with deep learning method for imbalanced data classification. *International Journal of Cloud Applications and Computing (IJCAC)*, 11(4):1–13, 2021.

[13] Emam Hossain, Mohd Arafath Uddin Shariff, Mohammad Shahadat Hossain, and Karl Andersson. A novel deep learning approach to predict air quality index. In *Proceedings of International Conference on Trends in Computational and Cognitive Engineering: Proceedings of TCCE 2020*, pages 367–381. Springer, 2020.

[14] Khan Md Hasib, Anika Tanzim, Jungpil Shin, Kazi Omar Faruk, Jubayer Al Mahmud, and MF Mridha. Bmnet-5: a novel approach of neural network to classify the genre of bengali music based on audio features. *IEEE Access*, 10:108545–108563, 2022.

[15] Sadia Afrin, Mohammad Maksimul Islam, Tanvir Ahmed, et al. A meteorology based particulate matter prediction model for megacity dhaka. *Aerosol and Air Quality Research*, 21(4):200371, 2021.

[16] SA Shahriar, I Kayes, K Hasan, MA Salam, and S Chowdhury. Applicability of machine learning in modeling of atmospheric particle pollution in bangladesh. air qual atmos health 13: 1247–1256, 2020.

[17] Udaya Bharathi Rambha and Maruvada Seshashayee. Time series augmentation based on vector auto regression and long short term memory method for air quality prediction.

[18] Nimisha Tomar, Durga Patel, and Akshat Jain. Air quality index forecasting using auto-regression models. In *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pages 1–5. IEEE, 2020.

[19] Yansui Liu, Yang Zhou, and Jiaxin Lu. Exploring the relationship between air pollution and meteorological conditions in china under environmental governance. *Scientific reports*, 10(1):1–11, 2020.

[20] Khan Md. Hasib, Md. Ahsan Habib, Nurul Akter Towhid, and Md. Imran Hossain Showrov. A novel deep learning based sentiment analysis of twitter data for us airline service. In *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, pages 450–455, 2021.

[21] Zahra Hajirahimi and Mehdi Khashei. Hybridization of hybrid structures for time series forecasting: A review. *Artificial Intelligence Review*, 56(2):1201–1261, 2023.

[22] Hanif Bhuiyan, Jinat Ara, Khan Md Hasib, Md Imran Hossain Sourav, Faria Benta Karim, Cecilia Sik-Lanyi, Guido Governatori, Andry Rakotonirainy, and Shamsunnahar Yasmin. Crash severity analysis and risk factors identification based on an alternate data source: a case study of developing country. *Scientific reports*, 12(1):21243, 2022.

[23] Iqra Sardar, Muhammad Azeem Akbar, Víctor Leiva, Ahmed Alsanad, and Pradeep Mishra. Machine learning and automatic arima/prophet models-based forecasting of covid-19: Methodology, evaluation, and case study in saarc countries. *Stochastic Environmental Research and Risk Assessment*, 37(1):345–359, 2023.

[24] Kim Leone Souza da Silva, Javier Linkolk López-Gonzales, Josue E Turpo-Chaparro, Esteban Tocto-Cano, and Paulo Canas Rodrigues. Spatio-temporal visualization and forecasting of pm 10 in the brazilian state of minas gerais. *Scientific Reports*, 13(1):3269, 2023.

[25] Zixi Zhao, Jinran Wu, Fengjing Cai, Shaotong Zhang, and You-Gan Wang. A hybrid deep learning framework for air quality prediction with spatial autocorrelation during the covid-19 pandemic. *Scientific Reports*, 13(1):1015, 2023.

[26] Khan Md Hasib, Shadman Sakib, Jubayer Al Mahmud, Kamruzzaman Mithu, Md Saifur Rahman, and Mohammad Shafiul Alam. Covid-19 prediction based on infected cases and deaths of bangladesh using deep transfer learning. In *2022 IEEE World AI IoT Congress (AIIoT)*, pages 296–302. IEEE, 2022.

[27] Hong Yang, Yiting Zhang, and Guohui Li. Air quality index prediction using a new hybrid model considering multiple influencing factors: A case study in china. *Atmospheric Pollution Research*, 14(3):101677, 2023.