Prediction and forecasting of air quality index in Chennai using regression and ARIMA time series models

Geetha Mani*, Joshi Kumar Viswanadhapalli* and Albert Alexander Stonier**

* *School of Electrical* Engineering*, Vellore Institute of Technology, Vellore.*
* *Department of Electrical and Electronics Engineering, Kongu Engineering College, India.*
* *Corresponding Author: geetha.mani@vit.ac.in*

ABSTRACT

Air is one of the most fundamental constituents for the sustenance of life on earth. The meteorological, traffic factors, consumption of nonrenewable energy sources, and industrial parameters are steadily increasing air pollution. These factors affect the welfare and prosperity of life on earth; therefore, the nature of air quality in our environment needs to be monitored continuously. The Air Quality Index (AQI), which indicates air quality, is influenced by several individual factors such as the accumulation of NO_2, CO, O_3, PM2.5, SO_2, and PM10. This research paper aims to predict and forecast the AQI with Machine Learning (ML) techniques, namely linear regression and time series analysis. Primarily, Multilinear Regression (MLR) model, supervised machine learning, is developed to predict AQI. NO_2, Ozone(O_3), PM 2.5, and SO_2 sensor output collected from Central Pollution Control Board (CPCB), Chennai region, India, fed as input features and optimized AQI calculated from sensor's output set as a target to train the regression model. The obtained model parameters are validated with new and unseen sensor output. The Key Performance Indices (KPI) like coefficient of determination, root mean square error, and mean absolute error were calculated to validate the model accuracy. The K-cross-fold validation for testing data of MLR was obtained as around 92%. Secondly, the Auto-Regressive Integrated Moving Average (ARIMA) time series model is applied to forecast the AQI. The obtained model parameters were validated with unseen data with a timestamp. The forecasted AQI value of the next 15 days lies in a 95 % confidence interval zone. The model accuracy of test data was obtained as more than 80%.

Keywords: Air quality index; Machine learning; Prediction; Forecasting; Multi linear regression; ARIMA time series model.

INTRODUCTION

Air is one of the essential elements to sustain life on earth. The nature of air increases the life span of every species on our earth. Breathing polluted air causes harmful diseases such as coronary heart disease (CHD), Chronic Obstructive Pulmonary Diseases (COPD), and lung cancer, exceptionally high in youngsters and newborns. The long-term effect of poor air quality leads to worldwide increases in temperature and changes in climatic patterns (Sankar & Arulmozhivarman, 2017). Air pollution is one of the risk factors in densely populated regions in the world.

New ailments are being analyzed each day, and more findings are being accounted for a consistent increase in air pollution from industrialization and consumption of nonrenewable energy. Various surveillance stations have been set up worldwide to check the nature of air quality (Sankar et al., 2018).

According to the World Health Organization (WHO), there are about 2.4 million deaths worldwide because of poor air quality compared to other causes (Ganesh et al., 2018; World Health Organization, 2016). The AQI has been proposed to address the measure of air quality in a region. The primary goal of an AQI is to rapidly publicize real-time information on air quality, especially pollutants having short-term impacts. Previously, various Machine Learning & Artificial Intelligence (ML-AI) methodologies have been proposed for air quality forecasting. Individually, computational strategies like regression models and time-series analysis contribute much precision in the prediction of AQI. These strategies have confirmed their efficiency and robustness under several circumstances.

Literature Survey

Some of the notable works regarding AQI monitoring are done by YYang et al. who implemented a mobile AQI monitoring system using the Gaussian plume model based on the neural network (Yang et al., 2018). Y Yang et al. developed an ImgSensingNet, a vision-based aerial-ground sensing system for AQI monitoring and forecasting by the fusion of images taken from the Unmanned Aerial-Vehicle (UAVs) (Yang et al., 2019). Y Yang et al. presented an aerial-ground wireless sensor (WSN) network for monitoring of real-time PM2.5 using unmanned-aerial-vehicle (UAVs) in an urban city (Yang et al., 2018). Z Zheng et al. designed a 3-dimensional (3D) real-time AQI monitoring using the Adaptive Gaussian Plume (AGPM) model with the help of Unmanned Aerial-Vehicle (UAVs) (Yang et al., 2017). Z hu et al. developed real-time, fine-grained, power-efficient air quality sensing for the smart city and compares the ground sensing data and aerial sensing data to improve the data collected (Z Hu et al., 2019). M Khashei et al. presented the performance analysis and comparison of both the ARIMA model and the Artificial Neural Network (ANN) model for various data set like a sunspot, Canadian lynx, and dollar exchange to forecast future values (Mehdi Khashei et al., 2011). Neural networks are challenging to deal with because of a complex nature. Also, it is not suitable for real-time data change for a short period as per the literature survey (Mehdi Khashei et al., 2011). Ligang et al. proposed time-based competition, which facilitates the fast movement to study a multijoint replenishment problem (Ligang Cui et al., 2020). Jie Deng et al. presented a stochastic lead-time and demand that optimize ordering cost, holding cost, lost-scale cost, transportation cost, and analysis two stochastic factors on the total cost are performed for joint replenishment distribution problem (Ligang Cui et al., 2020). Ligang et al. presented a new algorithm, namely the bare-bones differential evolutionary algorithm, to reduce the uncertainty in the joint replenishment problem (Ligang Cui et al., 2020). Liga Cui et al. presented the best ordering strategies for stakeholders with RFID (Ligang Cui et al., 2020). In Thailand a unified method which combines IoT and data analytics for prediction of particulate matter pollution (John Joseph 2019). Hourly data of weather concentrations and PM2.5 was collected and predicted future values using support vector regression (John Joseph 2019). A unique weather monitoring system was implemented using IOT and raspberry pi which reduced the power consumption and increased the robustness (John Joseph, Ferdin Joe. (2019). A detailed review on IoT system in environment was provided in which various sub domains of IoT and research challenges are listed out (John Joseph 2019).

The algorithm should be very lightweight and should show the results immediately without any delay. The data dealt with is linear. Hence, multilinear regression is suitable. The other methods, like k-nearest neighbors/random forests, are more suited for nonlinear data.

The annual exposure to PM2.5 in India during 2017 was 89.9 μg/m³, which was one of the plant's extreme conditions. The highest yearly PM2.5 in India in 2017 was in Delhi-209.0 μg/m³. In India's entire populace during 2017, 42.6% of occupants were exposed to PM2.5 with a concentration level of more than 80 μg/m³. The extent of the people utilizing fuels in India is one factor for an increase in air pollution during 2017. It was reported as 55.5%

in Jharkhand, Bihar, and 76.7-81.5% in Odisha. Ozone's annual exposure (O_3) during 2017 in India was 60.1 parts per billion (ppb). In India, during 2017, there were 12 lakh deaths due to air pollution. Out of total mortality, 12.5% are due to air pollution. This extent was greater in Jharkhand, Chhattisgarh, and Bihar than in other states. The deaths due to air pollution in India during 2017 were 3.1 times more when compared to the previous year (the impact of air pollution on deaths, 2017).

This research paper is prepared and presented as follows: Section 2 includes the structure and calculation of AQI. Section 3 addresses the data preparation and preprocessing data approaches for air pollution monitoring and forecasting. The machine learning techniques, such as MLR and ARIMA, and their performance indices are presented in Section 4. The results and discussion are shown in Section 5 to validate the effectiveness of the present research work. Finally, Section 6 gathers the conclusion of the work.

AIR QUALITY INDEX (AQI)

AQI can be well defined as a strategic methodology that convert search parameter's weighted values (for instance, pollutant concentrations) relating to air pollution into a single value or set of values. It can be noted that actual concentrations cannot be accepted as an index if they are observed and reported in μg/m3 or ppm (parts per million) along with standards (CPCB, 2014).

Structure of an Index

In the formulation of AQI, there are two primary steps:
Forming subindices for each pollutant
Accumulation of subindices to get the AQI.

Subindices (I_1, I_2...., I_n) for n pollutant variables (X_1, X_2...., X_n) are formed using subindex functions that are dependent on air quality standards. The mathematical representation is given in equation (1)

$$I_i = f(X_i), \quad i = 1,2, \dots \dots n \tag{1}$$

Moreover, the subindex signifies the strong relationship between pollutant concentrations and health effects. This functional relationship between pollutant variables (X_i) and subindex (I_i) value is gingerly well-explained in subsection 2.2 of this paper. The subindices I_i are then aggregated using the function (F) in equation (2) to get the overall Index (I), and it is termed as AQI.

$$I = F(I_1, I_2 \dots \dots I_N) \tag{2}$$

Calculation of Subindices

The subindex function signifies the strong relationship between pollutant variable X_i and corresponding subindex I_i. It is an attempt to show the relationship between environmental consequences as the concentration of specific pollutants. Also, it can be linear, nonlinear, or segmented linear. Equation (3) represents the relationship between pollutant concentration X_i and corresponding subindex I_i.

$$I = \alpha X + \beta \tag{3}$$

where α =slope of the equation (3) of a line,

β = y-intercept value when X is zero.

The equation for subindex (I_i) for a particular pollutant concentration (C_p) depends on the linear segmented principle, and it is defined using equation (4)

$$I_i = \left[\left\{\frac{I_{HI} - I_{LO}}{B_{HI} - B_{LO}}\right\} * C_P - B_{LO}\right] + I_{LO} \tag{4}$$

where

B_{HI}= Breakpoint concentration greater or equal to a given concentration
B_{LO}= Breakpoint concentration smaller or equal to a given concentration
I_{HI} =AQI value corresponding to BHI
I_{LO} = AQI value corresponding to BLO
I_i = Subindex of the pollutant
C_p = Pollutant concentration

Aggregation of Subindices

Once the subindices are formed, they are combined using equation (5), the maximum value of all subindices (CPCB, 2014).

$$\boldsymbol{I} = \text{Aggregated Index} = Maxium(I_1, I_2, I_3, I_4, \dots\dots I_n) \tag{5}$$

where
I_n= Subindices for n pollutants,
n = Number of pollutant variables

The recommended measures of AQI in India are provided by the Central Pollution Control Board (CPCB) (CPCB, 2014) and are given in Table 1.

Table 1. AQI Standards by CPCB, India.

Range	Air Quality Index (AQI)	Associated Health Impacts
0-50	Good	Minimal Impact
51-100	Satisfactory	May cause minor breathing discomfort to sensitive people
101-200	Moderately Polluted	May cause breathing discomfort to children and Older adults
201-300	Poor	May cause breathing discomfort to people
301-400	Very poor	May cause respiratory illness to the people
401-500	Severe	May cause respiratory impact even on healthy people, and severe health impacts on people

MATERIALS

Data Preparation

In this research, air pollution monitoring and forecasting are carried using data obtained from Chennai, India. Chennai is one of the most important cultural, economic, and educational centers in South India. It is one of the most crowded cities in India, with a populace of around 10 million. The primary air pollutants in and around Chennai include NO_2, CO, O_3, PM2.5, PM10, and SO_2. The air pollutants used in the present investigation are shown in Table 2. The plot for air pollutants is shown in Figure 1. The data was collected from the CPCB website from September 2018 to April 2020 (http://www.cpcb.nic.in/).

Table 2. Statistical measure of Chennai data samples.

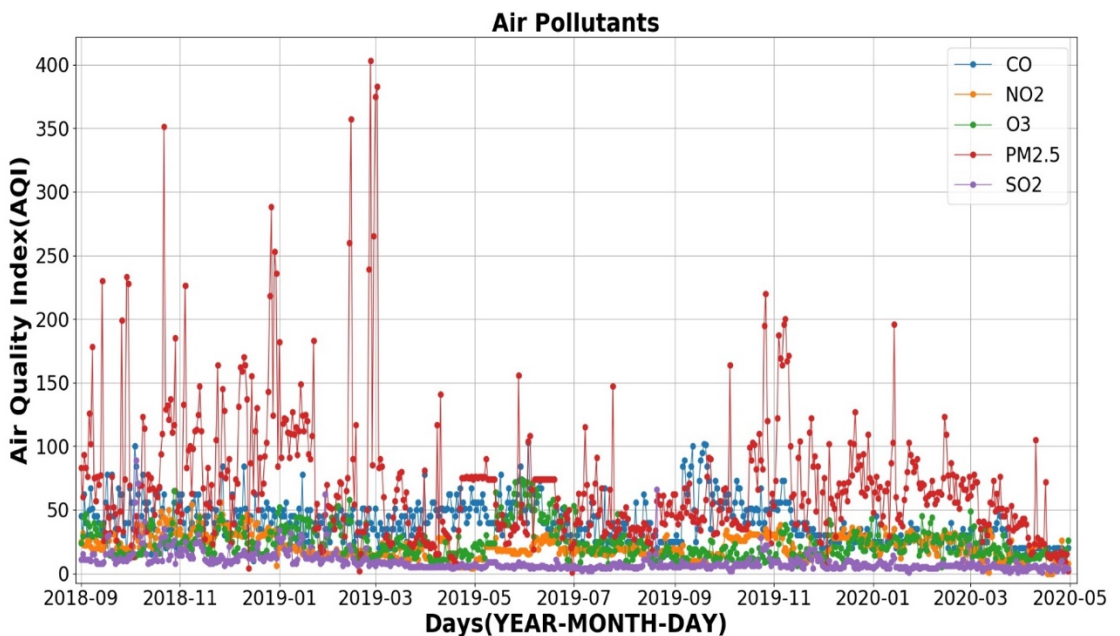| Air Pollutant	Units	Range	Mean	Standard Deviation	Variance
CO	mg/m^3	0-2	39.48	16.87	284.56
NO_2	µg/m^3	0-80	21.07	9.20	84.64
Ozone(O_3)	µg/m^3	0-180	23.84	12.42	154.25
PM2.5	µg/m^3	0-60	70.10	52.12	2716.49
SO_2	µg/m^3	0-80	9.26	8.05	64.80

Figure 1. Air Pollutants Graph.

The data set contains 607 values that are partitioned into 486 (80%) values for training and 122 (20%) values for testing of prescient models. The proposed approach uses the Multilinear Regression (MLR) model to check the linearity between air pollutants and the optimized AQI. The prediction of AQI is made with the help of the MLR

model. The time series analysis-ARIMA model is used for forecasting the AQI. Each model's performance is assessed using the error-index such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and cross-validation test.

Data Preprocessing

Each air pollutant has its own units and subindices, according to CPCB (CPCB, 2014), India. The data that is obtained from different ground stations should be preprocessed. Hence, the raw data was filtered before applying machine learning techniques (ML) for the prediction. The air pollutant data collected from CPCB and their standard ranges are given in Table 2. The individual air pollutants converted to corresponding AQI with the help of guidelines provided by the CPCB website. The data set contains 607 air pollutant values considering one reading per day. 8 % of the data set having a NAN value. The statistical measure of Mean value imputes the missing values in the data set. The modified data set is then used for both the MLR model (prediction) and the ARIMA model (forecasting). The block diagram for data processing and optimal AQI calculation is shown in Figure 2.
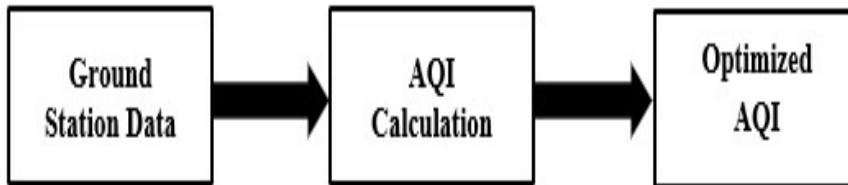

Figure 2. Block Diagram AQI calculations.

METHODOLOGY

Multilinear Regression

MLR is a supervised learning model that attempts to establish the relationship between two or more explanatory variables and a response variable by fitting a linear equation of the data. The MLR block diagram is shown in Figure 3. It is an extension to an ordinary least square regression that involves over one explanatory variable. The regression line for p explanatory variable x_1, x_2……. x_p is defined as μ. This regression line mentions the mean response of μ_y, which is going to change with explanatory variables. The observed values of y differed about their means μ_y and were assumed to have the standard deviation σ (Juile et al., 2004). The formula for MLR is given in eq. (6)

$$y_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \cdots + \alpha_p x_{ip} + \varepsilon \tag{6}$$

where

    i=nth reading
    x_i= explanatory variable
    y_i=dependent variable
    α_0=y-intercept
    α_p=slope value of each explanatory variable
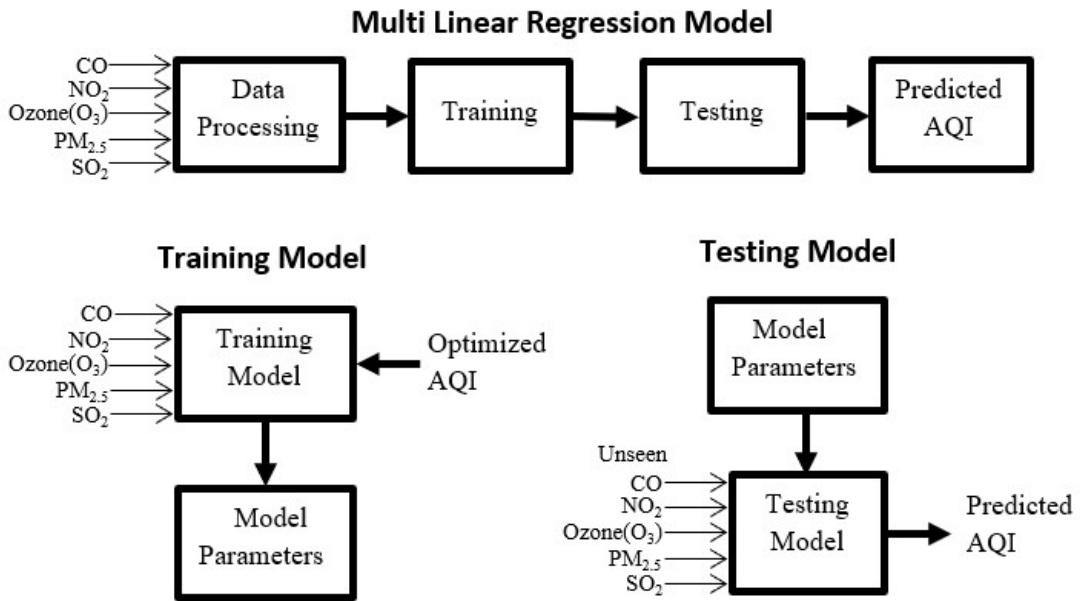 ε= error term

Figure 3. Block Diagram of Multilinear Regression.

Time-Series Analysis

Time-series Analysis is data points indeed in a timely order. It is a sequence taken at successive intervals in time. It is a method of forecasting to predict the future based on previous data. ARIMA is defined as Auto Regressive Integrating Moving Average. ARIMA model combines three different models: the Auto-Regressive model, integrated model, and moving average model. ARIMA model can be applied to data, which is of nonstationary type. Nonstationary information is the data that does not have continuous successive intervals in the series (Wang et al., 2015; S Gocheva-Ilieva et al., 2019).

ARIMA model are generally denoted with p, d,q which are nonnegative integers

Where

p is the number of time lags in the Auto-Regressive (AR)Model
d is the degree of differencing(I) Model
q is the order of the Moving Average (MA) Model

Data transformation has been performed in the ARIMA model during data identification to make the nonstationary data to stationary data. A stationary is a necessary condition for ARIMA Model. The stationary of the data is characterized by mean, standard deviation, and autocorrelation structure. If the data present any trend, then applying the differencing and power transformation trend will be removed. Once the ARIMA model is identified, model parameters are estimated, and the final selected model is used for prediction purposes (S Gocheva-Ilieva et al., 2019). The block diagram for the time-series Analysis-ARIMA model is shown in Figure 4.
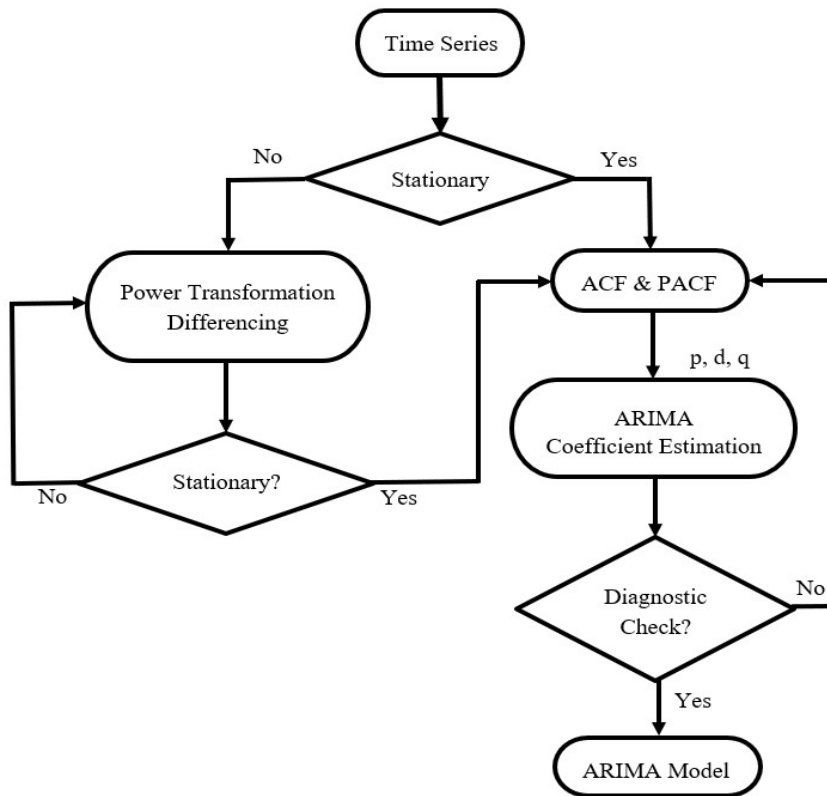
Figure 4. Block Diagram of ARIMA Model.

Performance Indices

The statistical criteria such as R-squared, RMSE, and MAE are used (M Swamynathan, 2017) to evaluate each developed model's performance measure.

R-Squared for Goodness of Fit

The R-squared metric is the most popular practice of evaluating how well the model fits the data. It is a value between 0 and 1; the value toward 1 indicates a better model fit. It is calculated using eq. (7).

$$R^2 = \frac{\sum(\widehat{x_i}-\bar{x})^2}{\sum(x_i-\bar{x})^2} \tag{7}$$

where

x = dependent variable
$\hat{x}$ = predicted variable
$\bar{x}$ = mean of the dependent variable
x_i = i[th] value of the dependent variable column
$\widehat{x_i}$ = i[th] value of predicted dependent variable column

4.3.2 Root Mean Squared Error (RMSE)

RMSE is the square root of the mean of the squared errors. RMSE indicates how close the predicted values are to the actual values. Hence, the lower RMSE value signifies that the model performance is good. RMSE is calculated using eq. (8).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2} \qquad (8)$$

Mean Absolute Error (MAE)

MAE is the mean or average of the absolute value of the errors, the Predicted – Actual. It is calculated using eq. (9).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - \hat{x}_i| \qquad (9)$$

k-Fold Cross-Validation Method

K-fold Cross Validation Method is a resampling method to test the ML models using limited data samples. Single parameter 'k' is referred to as the number of the group splits for the data sample, such as k=10, which means splitting the data into 10-fold cross-validation to evaluate the ML model. The results are much more accurate and less biased.

RESULTS AND DISCUSSION

Multilinear Regression

MLR is used to predict the AQI. However, before that, a correlation is obtained between the air pollutants mentioned in Table2. The correlation coefficients of air pollutants are shown in Table 3. From correlation coefficients, this inferred that they are no strong correlation between any two air pollutants. With the help of Training data, the MLR model is trained using python. The coefficients of the MLR equation are given in Table 4. The equation for MLR is given in equation (10).

$$x_i = -0.5570 + 0.3647y_{i1} - 0.1792y_{i2} + 0.0092y_{i3} + 0.9138y_{i4} + 0.1199y_{i5} \qquad (10)$$

Table 3. Correlation coefficients of air pollutants.

	CO	NO_2	O_3	$PM_{2.5}$	SO_2
CO	1.0000	0.09249	0.1186	0.1232	0.2073
NO_2	0.0924	1.0000	0.1495	0.2831	0.2980
O_3	0.1186	0.1495	1.0000	0.1887	0.0660
$PM_{2.5}$	0.1232	0.2831	0.1887	1.000	0.2350
SO_2	0.2073	0.2980	0.0660	0.2350	1.0000

Table 4. Coefficients of MLR equation.

Model Parameters	Values
y-intercept (α_0)	-0.5570
α_1	0.3647
α_2	-0.1792
α_3	0.0092
α_4	0.9138
α_5	0.1199

The prediction of AQI using training data is shown in Figure 5. The MLR was tested using unseen test data using the obtained model parameters of the MLR.
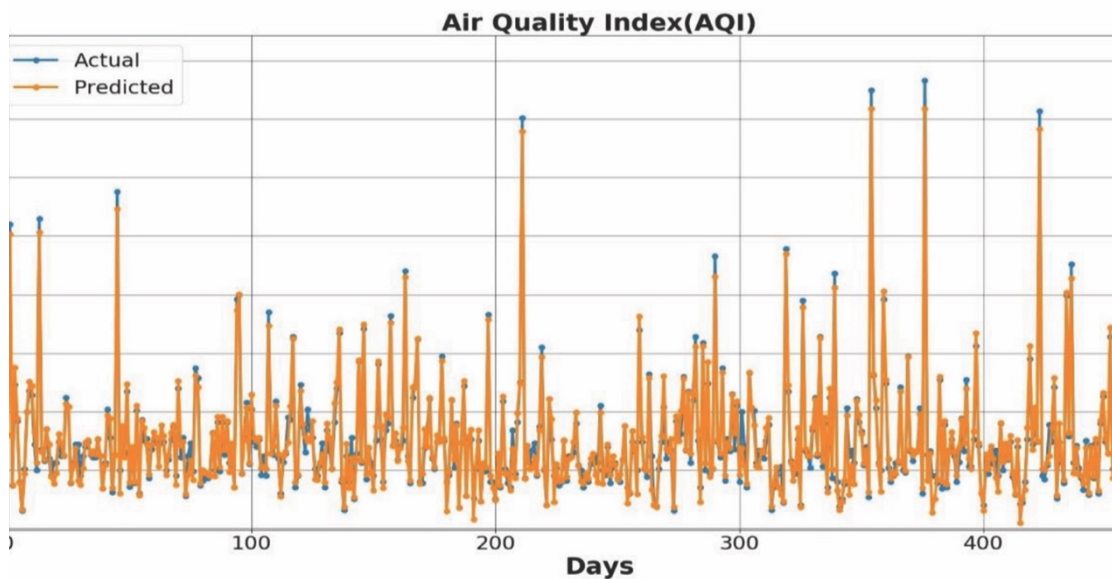

Figure 5. Prediction of AQI using training data.

Figure 5 shows the graph for the prediction of AQI using the training data. The training data consist of 80% (486 samples) of the aggregate data, and the data sample is taken randomly for training the MLR algorithm. From the above Figure 5, it was inferred that the MLR algorithm is very accurate for predicting AQI. The unseen data is used for validating the obtained model parameters of the same algorithm. The validation results for the AQI prediction are shown in Figure 6.
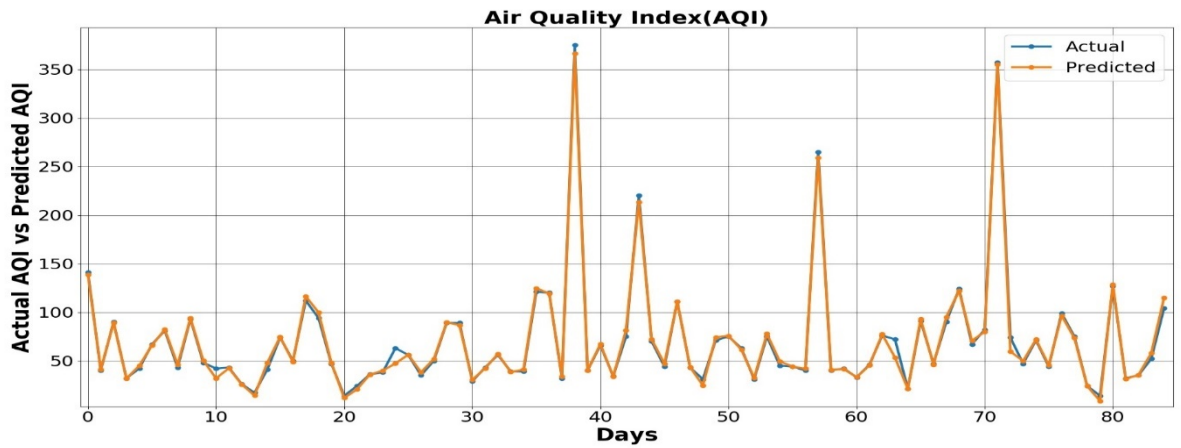
Figure 6. Prediction of AQI using Test data.

From Figure 6, it was inferred that the MLR algorithm has more accuracy for the prediction of AQI. The test data set consists of 20% (122 data samples) of the whole data set. The test data is given separately to the algorithm, and these data are not used for training the algorithm to test the accuracy of the MLR algorithm. The calculated quantitative performance indices are given in Table 5 and Table 6.

Table 5. Error variance for Each Fold in K-fold Cross-Validation.

Number of Folds in K-fold Cross-Validation	Training Data	Test Data
1-Fold	0.000006	0.000041
2-Fold	0.000063	0.000616
3-Fold	0.000025	0.000673
4-Fold	0.000050	0.000984
5-Fold	0.000161	0.003448
6-Fold	0.000094	0.009386
7-Fold	0.000232	0.001509
8-Fold	0.000695	0.003720
9-Fold	0.000136	0.015213
10-Fold	0.000887	0.003795

Table 6. Performance indices of Multilinear Regression model.

Performance Indices	Training Data	Testing Data
K-fold cross validation	96.62 %	91.74 %
R^2	0.99	0.97
MAE	2.7922	2.8041
RMSE	4.5159	4.4455

Time-Series Analysis

Time-series analysis-ARIMA is used to forecast the AQI. As discussed in section 3.2 ARIMA model is the combination of three different individual models known as the Auto-Regressive (AR) model denoted by p, differencing (I) model indicated by d, moving average (MA) model denoted by q. The coefficients AR model and MR model are calculated with the help of Partial Auto-Correlation Function (PACF) and Auto-Correlation Function (ACF). The coefficient of the Differencing model depends on the number of times the data is differentiated. Differentiation relies on the stationarity of the data. The dickey-fuller test is performed to find whether the given data is stationary or not. The results of the dickey fuller test confirmed that the dataset is nonstationary. Hence, the data is differentiated by two times to make it stationary, and the coefficient of the differencing model (d) is calculated as 2. The p and q coefficients were obtained from PACF and ACF graphs. Table 7 shows the obtained ARIMA model coefficients.

Table 7. ARIMA model coefficients.

Model Coefficients	Values
Auto Regressive Model (p)	8
Differencing Model (d)	2
Moving Average Model (q)	1

ARIMA model is constructed with the help of training data and the coefficient values given in Table 7. AQI is forecasted up to April 2020 using the obtained ARIMA model with a 95 %confidence interval. The 95% interval indicates that future AQI will present within this interval range, shown in Figure7. The obtained ARIMA model is validated using unseen test data to ensure the developed model's performance, represented in Figure 8.
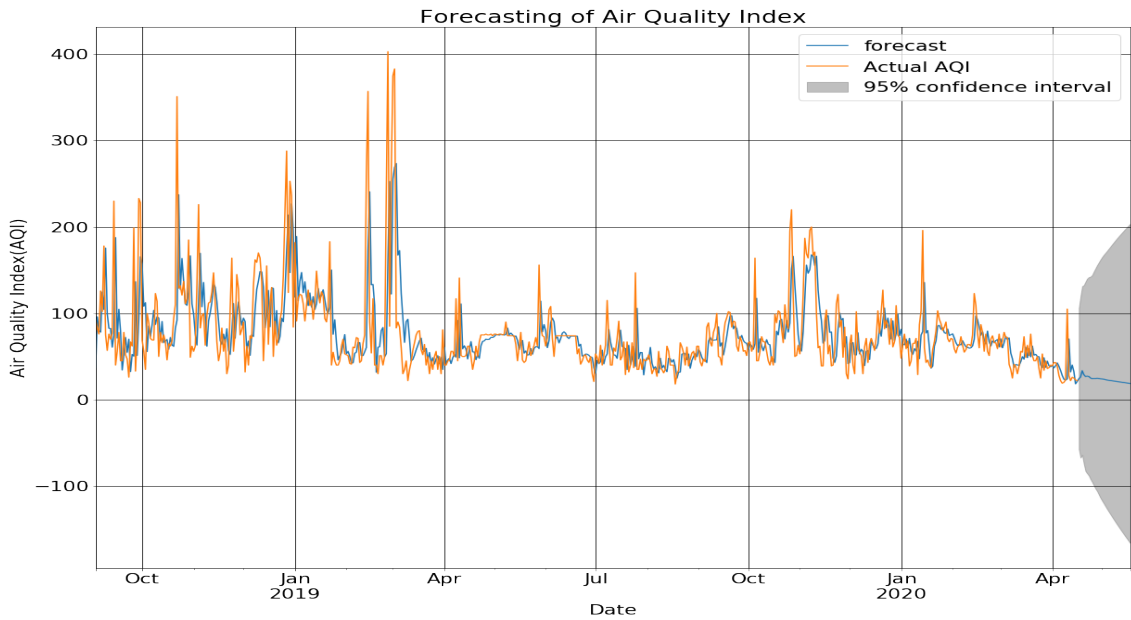
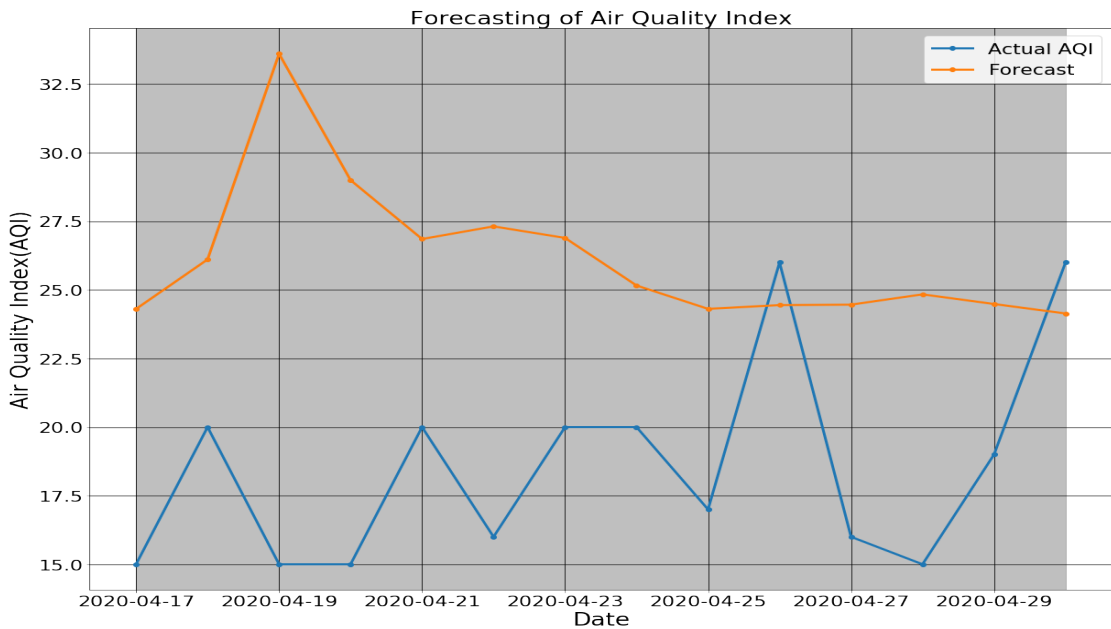**Figure 7.** Forecasting of AQI using ARIMA model using Training data.

Figure 8. Validation of Forecasting of AQI using ARIMA model with test data.

Figure 8 presents the comparison of the actual value and the forecasted value obtained by the ARIMA model with unseen test data. The obtained forecasted values present in a 95 % confidence interval zone. The quantitative performance evaluation indices of the ARIMA model for the test data set are calculated and presented in Table 8.

Table 8. Quantitative Performance indices of ARIMA model.

Performance Indices	Test Data
R^2	0.90
MAE	8.05404
RMSE	9.15379
Average Accuracy of Test Data	82%

CONCLUSION

The prediction and forecasting of AQI comprehend people and the environment from adverse health conditions because of poor AQI. Chennai region air pollution data is collected from September 2018 to March 2020 to assess and validate the developed machine learning models. In this research paper, the MLR model is used to predict the AQI. Firstly, the correlation between air pollutants was found, and then the MLR model was trained using the training data set and validated with the unseen test data. The performance indices such ask-fold cross-validation R^2, MAE, and RMSE for both training and test data were found to be satisfactory. Secondly, the ARIMA model is formulated and used for the forecasting of AQI. In the first place, the ARIMA model is trained with the help of training data and assessed with test data.

The performance indices for the ARIMA model with test data were found to be acceptable. Both linear and nonlinear machine learning methods were tested for AQI prediction. The dataset is linear. Hence, the linear ML technique provided the best fit. On the other hand, nonlinear ML techniques give a poor fit because the dataset is linear and not complex. Neural networks are challenging to deal with because of a complex nature. Also, it is not suitable for real-time data change for a short period. The algorithm should show the forecasting results quickly also feasible to implement with low computational cost in the hardware platform. ARIMA model used the data set up to April 2020 and forecasting the future values of the next 15 days within a 95% confidence interval. Therefore, the proposed methodology is recommended to predict AQI and the forecasting of AQI suits for real-time implementation in the future.

REFERENCES

CPCB (2014). National Air Quality Index. Control of urban pollution series CUPS 82/2014–15.; https://tinyurl.com/y78slqnq.

Ganesh, S.S., Arulmozhivarman, P., Tatavarti, &V.S.N.R. (2018). Prediction of $PM_{2.5}$ using an ensemble of artificial neural networks and regression models. *Journal of Ambient Intelligence and Human Computing.* https://doi.org/10.1007/s12652-018-0801-8. http://www.cpcb.nic.in/.

Joseph, F. J. J. (2019). A Review of IoT Implementations in Environment and Agriculture. *JARES*, *6*(1&2), 1-5

John Joseph, Ferdin Joe. (2019). Empirical Dominance of Features for Predictive Analytics of Particulate Matter Pollution in Thailand.

Joseph, F. J. J. (2019). IoT based weather monitoring system for effective analytics. *International Journal of Engineering and Advanced Technology*, *8*(4), 311-315.

John Joseph, Ferdin Joe. (2019). IoT Based Unified Approach To Predict Particulate Matter Pollution In Thailand.

Juile, Barber.,&Simon, Thompson. (2004). Multiple regression of cost data: use of generalised linear models. *Journal of Health Services Research & Policy*, 9(4), 197–204.https://doi.org/10.1258/1355819042250249.

Ligang, Cui., Jie, Deng., Fan, Liu., Yajun, Zhang.,& Maozeng, Xu. (2017). Investigation of RFID investment in a single retailer two-supplier supply chain with random demand to decrease inventory inaccuracy. *Journal of Cleaner Production,*142(4), 2028-2044.https://doi.org/10.1016/j.jclepro.2016.11.081.

Ligang, Cui., Jie, Deng., Rui, Liu., Dongyang, Xu., Yajun, Zhang., &Maozeng, Xu. (2020). A stochastic multi-item replenishment and delivery problem with lead-time reduction initiatives and the solving methodologies. *Applied Mathematics and Computation,*374. https://doi.org/10.1016/j.amc.2020.125055.

Ligang, Cu.i, Jie, Deng., Yajun, Zhang., Guofeng, Tang., &Maozeng, Xu. (2020). Hybrid differential artificial bee colony algorithm for multi-item replenishment-distribution problem with stochastic lead-time and demands. *Journal of Cleaner Production*, 254. https://doi.org/10.1016/j.jclepro.2019.119873.

Ligang, Cui., Jie, Deng., Yajun, Zhang., Zijian, Zhang.,& Maozeng, Xu. (2020). The bare-bones differential evolutionary for stochastic joint replenishment with random number of imperfect items. *Knowledge-Based Systems,* 193. https://doi.org/10.1016/j.knosys.2019.105416.

M, Swamynathan. (2017). Machine Learning with Python in Six Steps: A Practical Implementation Guide to Predictive Data Analytics Using Python, Berkeley CA: Apress.https://doi.org/10.1007/978-1-4842-2866-1

Mehdi, Khashei.,& Mehdi, Bijari. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*, 11(2), 2664–2675. https://doi.org/10.1016/j.asoc.2010.10.015.

S, Gocheva-llieva.,& A, Ivanov. (2019). Assaying SARIMA and generalised regularised regression for particulate matter PM10 modelling and forecasting. *International Journal of Environment and Pollution*, 66 (1-3), 41-62. https://doi.org/10.1504/IJEP.2019.104520.

Sankar, Ganesh, S., Arulmozhivarman,& P., Tatavarti, R. (2017). Forecasting Air Quality Index Using an Ensemble of Artificial Neural Networks and Regression Models. *Journal of Intelligent Systems*, 28(5), 893-903.https://doi.org/10.1515/jisys-2017-0277.

Sankar, Ganesh, S., Arulmozhivarman, P., &Tatavarti, R. (2018). Air quality index forecasting using artificial neural networks-a case study on Delhi. *International Journal of Environment and Waste Management,* 22(1-4), 4-23. https://doi.org/10.1504/IJEWM.2018.094105.

The Global Burden of Disease Study (2017). The impact of air pollution on deaths, disease burden, and life expectancy across the states of India. *The Lancet Planetary Health,* https://doi.org/10. 1016/S2542-5196(18)30261-4.

Wang, W., Chau, K., Xu, D., &Chen, X. (2015). Improving Forecasting Accuracy of Annual Runoff Time Series Using ARIMA Based on EEMD Decomposition. *Water Resour Manage,* 29, 2655-2675.

World Health Organization, (2016) . Ambient air pollution: a global assessment of exposure and burden of disease. *World Health Organization*. https://apps.who.int/iris/handle/10665/250141.

Y, Yang., Z, Bai., Z, Hu., Z, Zheng., K, Bian.,& L, Song. (2018). AQNet: Fine-grained 3D spatio-temporal air quality monitoring by aerial-ground WSN. *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS),* 1-2.https://doi.org/10.1109/INFCOMW.2018.8406985.

Y, Yang., Z, Hu., K, Bian., L, &Song. (2019). ImgSensingNet: UAV Vision Guided Aerial-Ground Air Quality Sensing System. *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications,* 1207-1215.https://doi.org/10.1109/INFOCOM.2019.8737374.

Y, Yang., Z, Zheng., K, Bian., L, Song., &Z, Han. (2018). Real-Time Profiling of Fine-Grained Air Quality Index Distribution Using UAV Sensing. *IEEE Internet of Things Journal,* 5(1), 186-198. . https://doi.org/10.1109/JIOT.2017.2777820.

Y, Yang., Z, Zheng., K, Bian., Y, Jiang., L, Song., Z,& Han. (2017). Arms: A Fine-Grained 3D AQI Realtime Monitoring System by UAV. *GLOBECOM 2017 - 2017 IEEE Global Communications Conference,* 1-6.https://doi.org/10.1109/GLOCOM.2017.8253968.

Z, Hu., Z, Bai., Y, Yang., Z, Zheng., K, Bian.,& L, Song. (2019). UAV Aided Aerial-Ground IoT for Air Quality Sensing in Smart City: Architecture, Technologies, and Implementation. *IEEE Network,* 33(2), 14-22.https://doi.org/10.1109/MNET.2019.1800214.