International Journal of Early Childhood Special Education (INT-JECSE)
DOI: 10.48047/intjecse/V14I7.368 ISSN: 1308-5581 Vol 14, Issue 07 2022

# Predicting Air Quality Index Values Using ARIMA Model

**[1]Dr.S.Meenakshi Sundaram [2]Dr. P. Sivakumar, [3]Dr. M. Parthiban, [4]Dr.U.Revathy**

[1] *Professor, Department of Computer Science and Engineering*
*Mahalakshmi Engineering College, Trichy,Tamil Nadu,India,bosemeena@gmail.com*

[2] *Associate Professor, Department of Computer Science and Engineering*
Sasi Institute of Technology and Engineering, Tadepalligudem, Andhra Pradesh, India,sivakumarperumal@sasi.ac.in

[3] *Associate Professor, Department of Computer Science and Engineering*
Sasi Institute of Technology and Engineering, Tadepalligudem, Andhra Pradesh, India,parthimitit@sasi.ac.in
[4]*Guest Lecturer, Department of Computer Science Govt Arts College for Women, Sivagangai, Tamil Nadu,India, urevathymeena@gmail.com*

***Abstract-***Air pollution is a serious issue that has negative consequences for human health and living situations. As a result, pollution levels must be monitored in order to keep people informed about the air quality. This is accomplished through the use of an index known as the Air Quality Index AQI, which converts the concentrations of multiple contaminants into a single number. Government offices utilize an air quality list AQI to illuminate the general population about how defiled the air is currently or will become from here on out. The AQI is determined by averaging readings from an air quality sensor, which can rise attributable to traffic, woodland fires, or different elements that add to air contamination. As the AQI rises, general wellbeing concerns ascend too, hurting youngsters, the old, and those with respiratory or cardiovascular issues. An air contamination fixation during a characterized averaging period, gathered from an air screen or model, is expected for AQI computation. The work used to switch from air poison focus over completely to AQI fluctuates by poison, while the capacity used to change from air toxin focus over completely to AQI changes by toxin. The upsides of its air quality file are normally ordered into ranges
***Keywords-****Machine Learning, ARIMA,K-Means, AQI, RMSE, AR.*

## I. INTRODUCTION

People have begun to pay plenty of attention to the impact of the setting on health in recent years, and information concerning air quality has become attention of people's everyday lives. fundamental measure air quality observance data is also obtained exploitation current air quality observance sensors, stations, and satellite meteoric data. However, this will be lightweight, and it is important to forecast the long-haul trend of pollution. Currently, prognosis data encompasses a high level of accuracy and responsibleness. supported this, we have a tendency to suggest fusing precognitive data, i.e., prognosis data, with historical air quality data and meteoric data, aided by machine learning ways in which, to analysis data correlation and mining build a ARIMA model for predicting the air quality index values. This model permits Associate in Nursing economical declare predicting air quality index values with the improved performance. The aim of projected methodology is to Predict the Air Quality Index worth exploitation ARIMA methodology. The target of the projected methodology is

- To increase the accuracy by decreasing the worth of RMSE worth.
- To predict the AQI worth in conjunction with required parameter values.

The Air Quality Index (AQI) might be a data point that the govt. uses to talk the quality of air to the general population. With a rise in pollution concentrations, air quality deteriorates. as a result of the AQI grows, thus do thethreats to public health. Air quality indices varies by country and correlate to distinct national air quality standards.The Air Quality Index measures but unhealthy pollution is for traditional people. Indian AQI varies [1] from 0-500,with zero being good and 5 hundred being severe, in line with the Indian government (CPCB). material (PM 10 andPM 2.5), oxide (CO), ozone(O3), gas (NO2), oxide (SO2),

ammonia (NH3), and lead unit of measurement the eight primary pollutants that has to be taken into thought whereas scheming the AQI(Pb). To cipher the AQI, there ought to be data for a minimum of three pollutants, one in every of that ought to be PM10 or PM2.5. each waste material encompasses a definite concentration in Associate in Nursing AQI varyfrom0-500,thathasvariedhealthimpacts.

*A. Particulate Matter(PM2.5 and PM10)*

PM2.5 refers to material at intervals the atmosphere with a diameter of however 2.5 microns, or roughly three-dimensional of the diameter of a person's hair. PM2.5 particles unit of measurement thus minute compared to PM10that they're going to exclusively be known with Associate in Nursing electronic scientific instrument. PM10 particles have a diameter of 10 microns and unit of measurement classified as fine particles. Respirable material is another name for PM10. Soot, smoke, metal, nitrates, dust, water, and rubber, among different things, structure material. it's mainly created by smoke from vehicles, industry, and thus the burning of agricultural waste, among different sources.

*B. Carbonmonoxide (CO)*

Carbon monoxide is Associate in Nursing part waste material created once carbon-containing fuels like gas, gas,oil,coal,and wood unit of measurement in completely burned. High amounts of CO unit of measurement dangerous to people, nonetheless it is not attainable for humans to sight since it lacks color and scent. CO levels at intervals the air unit of measurement naturally concerning zero.2 parts per million, which may not be harmful to people. Volcanoes and bushfires unit of measurement two natural sources of oxide.

*C. Nitrogen Dioxide(NO2)*

The gaseous air pollutants NO2 unit of measurement created as a results of traffic and different fuel combustion processes. Overexposure to NO2 causes metabolism problems and inflammation of the lungs' lining. the simplestallowedamountofNO2iseightyg/m3, and humans unit of measurement unaffected.

*D. Ozone(O3)*

Although gas is present at intervals the atmosphere, concentration maxima is additionally found at two levels:thelayer(15-50km)andthusthelayer(0-15km).Thetransmissionofultraviolet ultra violet illumination light weight to the Earth's surface is regulated by stratospheric O3. once stratospheric air is mixed with it, a nature average back ground of roughly10-20 parts per billion (ppb) is formed, however the concentration is debatable.

## II. RELATEDWORK

Upadhyaya et al, (2010) [1] anticipated the Air Quality Index esteems that are basic and wide proportion of air quality in China, Hong Kong, Malaysia, and presently India. The Indian Air Quality Index (IND-AQI) is essentially a wellbeing record, containing descriptor words, for example, "Great (0-100), "Moderate (101-200), "Poor (201-300), "Exceptionally Poor (301-400), and Severe (401-500)." China's State Environment Protection Agency (SEPA) is responsible for estimating air contamination levels.

Zadeh et al, (2017) [1] proposed fluffy rationale as a device for naturally suspecting in a manner that is more human-like. Fluffy rationale has been generally utilized to display and control complex frameworks over the course of the past ten years. That can involve a person's current information and experience as contribution to build a model using this strategy. This property of fluffy rationale is one of its essential benefits that different frameworks don't consider. Different models in view of fluffy rationale have been made.

R S Kumar et al, (2017) [13] proposed the dark model GM (1,1), which involves a first-request subordinate and a solitary free factor, is utilized to figure results with more accuracy. Because of wrong discoveries of higher request subordinates, settling the dark model GM (n, m) for expectation is as yet a trouble, where n is the subsidiary request and m is the quantity of free factors.

Essentially, Kai Zhang et al, (2019) [21] proposed Gray affiliation range, dark co-range, dim gather range, dim stage range, dim slack time length, and different markers were found when the level of connection in dim social examination was traded for the relationship number in range investigation [16, 17].

Jiao, S. et al, (2010) [15] investigated on help vector relapse (SVR) move toward which is an exceptional sort of learning machine calculation that is seldom used to develop metropolitan air quality models that include various different financial perspectives. For the expectation of metropolitan residue fall levels, this paper proposes four SVR models utilizing direct, spiral premise, spline, and polynomial capacities as bits,

separately.

Essentially, P J Garcia Nieto et al, (2013) [7] proposed in view of a help vector machine (SVM) strategy to gauge day to day air contamination in the Oviedo metropolitan district (northern Spain) at a nearby scale. Any particle that might cause or add to an expansion in mortality or difficult disease, or that might offer a present or possible mischief to human wellbeing, was alluded to as an unsafe air poison or harmful air pollutant.

Additionally, Bing LIU et al, (2021) [2] further developed the forecast exactness of the six sorts of contaminations in the air, an air quality expectation model melding head part relapse (PCR), support vector relapse (SVR) machine, and autoregressive moving normal (ARMA) model was proposed to remedied the observing information of the smaller than usual air quality locator.

A Challoner et al, (2015) [7] proposed this study's displaying means to give a system to epidemiological examinations by consolidating openly available information from fixed external checking stations to better definitively conjecture inside air quality. The Personal-openness Activity Location Model (PALM), which predicts outside air quality at a particular structure, and Artificial Neural Networks (ANN), which model the structure's indoor/outside communication, are utilized to make forecasts.

Ziyu Wang et al, (2019) [1] proposed in view of the broad observing of air contaminations has brought about the assortment of enormous measures of data. Due to our restricted comprehension of the collaboration of human activities and dubious natural components, customary quality control frameworks in view of existing information might be incapable.

Fatemeh Gholamzadeh et al, (2020) [6] examined the air contamination guaging in Tehran city.In this work, they will use a Vector Auto Regression (VAR) model to expect day to day air contamination fixations in Tehran for the following day. Since there are sure connections between air toxins, apparently utilizing these relationships can further develop anticipating results.

Jibo Chen et al, (2020) [30] utilized a versatile Kalman separating approach for detecting and anticipating air quality qualities utilizing Auto regression model. Albeit the air quality file (AQI) has for quite some time been utilized to address the power of dimness and other air poisons, it is wasteful and inconsistent with ongoing discernment and conjecture. An auto-backward (AR) expectation model in view of detected AQI values was proposed in this exploration.

## III. ARCHITECTURE

A calculated model of a framework's construction, conduct, and different viewpoints is known as framework engineering. An engineering portrayal is a proper depiction and portrayal of a framework that is coordinated such that makes it simpler to reason about the framework's designs and ways of behaving. Framework engineering can be comprised of a framework and the subsystems that will be worked to carry out the general framework.

The block diagram in the Figure a for showing the steps in prediction model is given as follows. This procedure is for data preparation, training to get the desired predicted output. The above diagram represents the system architecture for Predicting Air Quality Index values using ARIMA Model.

## IV. METHODOLOGY

The prediction is made based on the user’s input. If the input is in corrector in appropriate data then it shows error and exits the system. The dataset considered is used for training and testing. In this model, based on the user’s input the prediction will work. The user input is taken in the form of number of days that required for prediction and dates forprediction. The out come of the prediction values are CO, NO, $PM_{2.5}$, $PM_{10}$, O3, $NO_2$and AQI value for the givendates.

The proposed work is executed in Python 3.6.4 with libraries Pandas, statsmodel. programming interface, stats model.tsa.stat tools, NumPy, adfuller and other compulsory libraries. For this venture a redid dataset containing a few significant boundaries like CO, NO, PM2.5, PM10, O3, NO2 are thought of. These boundaries are pre-handled and afterward Machine learning calculations are applied. At last, the outcomes are assessed.

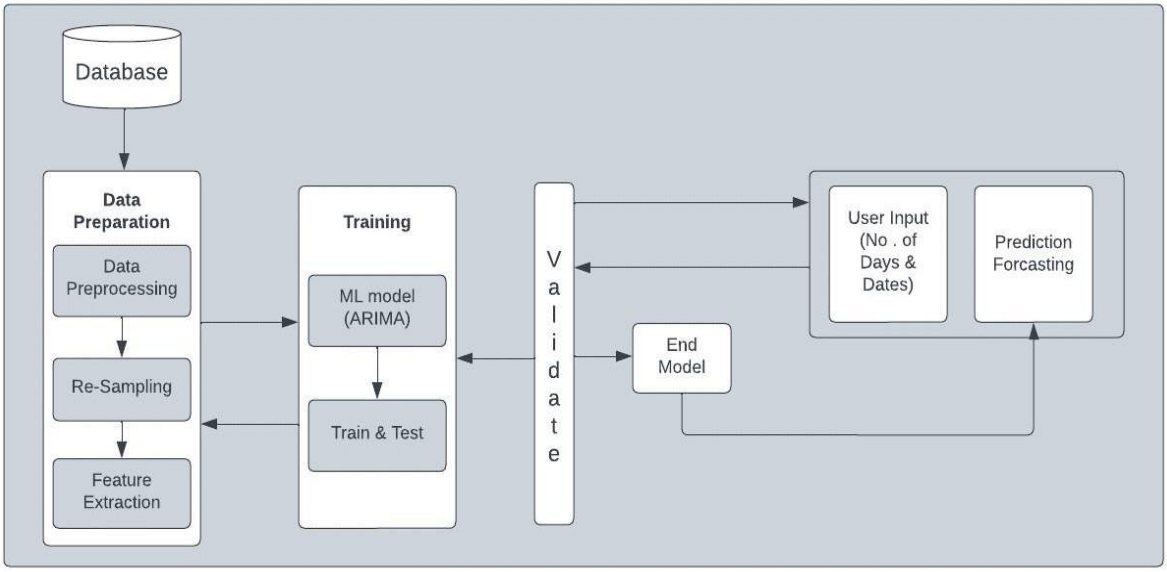


Figure1:Block Diagram for prediction of AQI Value.

Therearefourstepswithintheproposedmethodology:Datapreparation,training,validationandpredictionforecasting.

*A. Data Preprocessing*

Data preprocessing is that the process of reworking data into a lucid format. One can't work with data, thus this can be a key stage in data processing. Before using machine learning or data processing methods, the standard of the info should be checked. the aim of knowledge preprocessing is to make sure that the info is of fine quality. The standard of the info is check edby following factors.

**Exactness:** Pre-handling of data can once in a while end in startling improvement inside the model exactness. How much information precisely depicts a true situation and is affirmed by an unquestionable source is thought as information exactness. Information precision guarantees that this present reality elements included may partake as expected. It checks regardless of whether the data entered is right.

**Fulfillment:** Depending on the substance, this aspect can cover an enormous style of properties. The fulfillment of the data decides if framing valuable deductions and decisions is adequate.

**Consistency:** This aspect shows in the event that the indistinguishable information is recorded and utilized in different spots. Communicated as a level of values match across different records. Information consistency ensures that investigation gather and use the value of data precisely. Information consistency is normally connected to information exactness, in this way any informational collection that scores well on both are viewed as excellent.

**Legitimacy:** This aspect signifies the stockpile significant qualities for arrangement with a particular space or interest. Any invalid information will have an influence on the information's fulfillment. to affirm fulfillment, rules is characterized to overlook or determine invalid information.

**Uniqueness: This** aspect shows in the event that the data set used contains a solitary reported occurrence. the principal significant ponder staying away from duplication and cross-over is uniqueness. the peculiarity of information is established by contrasting all standards in an extremely information assortment or across informational collections. A high uniqueness score guarantees that copies and covers are limited, prompting information and examination that might be relied upon. Recognizing covers can assist with keeping records one of a kind, while information purifying and deduplication gets hinder copies. Information uniqueness likewise assists with information administration and consistence.

**Uprightness:** Integrity shows that the traits are kept up with accurately, when information gets put away and used in different frameworks. Information trustworthiness guarantees that each one endeavor information is followed and associated. Information change across frameworks can influence its quality connections.

**Idealness:** The value of data lies in its application. In the event that information isn't given as quickly as possibly, it's useless. The term practicality alludes as to if or not information is available when it's required.

***B. Data cleaning***

Information cleaning alludes to procedures to 'clean' information by eliminating exceptions, supplanting missing qualities, smoothing uproarious information, and remedying conflicting information. Every one of those exercises is administrated utilizing a sort of ways, every one of which is instant to the client's inclinations or issue set. Various methodologies are there to oblige missing qualities. Eliminating the preparation model, filling in missing worth physically, utilizing a standard worth to trade the missing worth, utilizing focal propensity like mean, middle and mode for trait to trade the missing worth, involving focal inclination measures for property having a place with same class to switch the missing worth, utilizing the preeminent plausible worth to fill inside the missing worth these methodologies are acclimated handle numerous information. Commotion is characterized as an irregular difference in an extremely estimated variable. To pander to these atypical qualities, information smoothing procedures are applied like binning, relapse and exception investigation.

***C. Dataintegration***

Data integration has become an important aspect of the method because data is acquired from numeroussources.

thiscouldendinduplicateandinconsistentdata,leadingtopoordatamodelaccuracyandspeed.Approacheslike tuple duplication detection and data conflict detection are being sorted to address these difficulties and maintaindata integrity. the foremost frequent data integration approaches are data consolidation, data propagation and datavirtualization.

***D. Feature extraction***

New features were created using the date component of the Central Pollution instrument panel data and therefore the date-time component of the US embassy data. A field called seasons was created using the date component. there have been four seasons in total (Summer, Fall, Winter, Spring). The cyclic structure of the time component was exploited to provide two fields, s in (2hour/24), and cos (2hour/24)respectively. Fields for day, month, and year were also acquired using the date component.

**E. ARIMA Model**

The abbreviation ARIMA represents Autoregressive Integrated Moving Average model, which consolidates Autoregressive (AR), Integrated, and Moving Average (MA), ARIMA models are a sort of complicated straight model that might deal with both fixed and non-fixed measurement. These models haven't any free factors in their structure and assemble gauges utilizing information from measurement. AR and MA models are OK for fixed measurement information, though I model are fundamental for non-fixed measurement information.

As its very own straight combination verifiable qualities, past blunders, and present qualities, an ARIMA model expects a cost during a dormancy series. It utilizes an intelligent way to deal with search out a fitting model from a larger than usual class of models, which is then contrasted with verifiable information to decide whether it precisely depicts the measurement. Occasional ARIMA models will be made by stretching out ARIMA models to include occasional parts of your time series information (SARIMA). ARIMA and SARIMA are expansions of the ARMA model that consolidate more exact elements, as non-stationarity in mean and contrasts because of the time. The ARIMA model is fitted to each boundary as follows.

```
[ ]  stepwise_fit = auto_arima(df['PM2.5'],
                               suppress_warnings=True)
     stepwise_fit1 = auto_arima(df['PM10'],
                                suppress_warnings=True)
     stepwise_fit2 = auto_arima(df['NO'],
                                suppress_warnings=True)
     stepwise_fit3 = auto_arima(df['CO'],
                                suppress_warnings=True)
     stepwise_fit4 = auto_arima(df['SO2'],
                                suppress_warnings=True)
     stepwise_fit5 = auto_arima(df['O3'],
                                suppress_warnings=True)
     stepwise_fit6 = auto_arima(df['AQI'],
                                suppress_warnings=True)

     stepwise_fit.summary()
     stepwise_fit1.summary()
     stepwise_fit2.summary()
     stepwise_fit3.summary()
     stepwise_fit4.summary()
     stepwise_fit5.summary()
     stepwise_fit6.summary()
```

Figure b: Stepwise suited each parameter

**F. Model Building**

The objective of this study is to utilize a measurement method to take a gander at the air quality list utilizing Autoregressive Integrated Moving Average (ARIMA) model. Various scholastics utilize this model in numerous areas of examination on the grounds that to its viable prompts the anticipating area. The AQI recipe is utilized to see the air quality list as follows:

$$I_p = \left[\frac{(I_{Hi} - I_{Lo})}{(BP_{Hi} - BP_{Lo})}\right][C_p - BP_{Lo}] + [I_{Lo}]$$

Where,

$I_p$ Individual AQI for poison ‘P’

$C_p$ Everyday mean convergence of contamination 'P'

$I_{Hi}$ Sub list worth or individual AQI esteem comparing to$BP_{Hi}$

$I_{Lo}$ Sub record worth or individual AQI esteem comparing to $BP_{Lo}$

$BP_{Hi}$ Breakpoint or close by higher worth more prominent than or equivalent to$C_p$

$BP_{Lo}$ Breakpoint or close by lower esteem not exactly or equivalent to$C_p$

On the possibility of the recorded poison focus at the checking station, the sub file values and breakpoint values could likewise be determined right away. The ARIMA model is made of three sorts of models: autoregressive (AR), coordinated (I), and moving normal (MA). On the off chance that 'Y' could be a variable and 't' might be a date, the value of variable 'Y' at time 't' is 'Yt'. A distinction condition is an articulation that interfaces the upsides of a variable 'Yt' to their earlier qualities. Condition communicates 'Yt' as a straight capacity of 'Yt 'and 'Wt'.

Anticipating inside the ARIMA model for autoregressive (AR)models is predicated on noticed values from the past major amount. inside the instance of moving normal (MA) models, nonetheless, the conjecture is predicated on past upsides of the blunders rather than the actual factors. The AR model, besides on the grounds that the pth request autoregressive model, are appropriate for fixed measurement information designs. The request for the autoregressive piece is meant by the letter 'p.'

Likewise, the qth request moving normal model. The Auto backward moving normal (ARMA) model is likewise made by joining autoregressive (AR) and moving normal (MA) parts. ARMA (p, q), where 'p' and 'q' are the request for the autoregressive and moving typical parts, separately.

```
[ ]  # from statsmodels.tsa.ARIMA import ARIMA
     # from statsmodels.tsa.arima_model import ARIMA
     import statsmodels.api as sm
     model=sm.tsa.ARIMA(train['PM2.5'],order=(1,1,1))
     model1 = sm.tsa.ARIMA(train['PM10'],order=(1,1,1))
     model2=sm.tsa.ARIMA(train['NO'],order=(1,1,1))
     model3=sm.tsa.ARIMA(train['CO'],order=(1,1,1))
     model4=sm.tsa.ARIMA(train['SO2'],order=(1,1,1))
     model5=sm.tsa.ARIMA(train['O3'],order=(1,1,1))
     model6=sm.tsa.ARIMA(train['AQI'],order=(1,1,1))
     model=model.fit()
     model1 = model1.fit()
     model2 = model2.fit()
     model3 = model3.fit()
     model4 = model4.fit()
     model5 = model5.fit()
     model6 = model6.fit()
     model.summary()
     model1.summary()
     model2.summary()
     model3.summary()
     model4.summary()
     model5.summary()
     model6.summary()
```

Figure c: Model Building for AQI

The ARIMA PC record series should be fixed, which proposes it should have a normalized mean, fluctuation, and autocorrelation over the course of time. Non-fixed information ought to be separated to make the data fixed. At the point when it includes functional purposes, the thing that matters is much of the time a couple (d<=2). A coordinated (I) model is shaped by differencing non-fixed series at least multiple times to acknowledge fixed. Accordingly, following differencing, the information series 'Y t' is affirmed to follow a coordinated autoregressive moving typical model and is addressed as ARIMA (p, d, q).

Where,

- 'p' is the autoregressive part's structure.
- The request for the moving typical piece is alluded to as 'q.'
- The times the information should be differenced to shape a fixed series is meant by the letter 'd.'

The Box-Jenkins method might be utilized to assess occasional time series information. The model is known as the Seasonal Autoregressive Integrated Moving Average (SARIMA) Model when occasional parts are incorporated. This occasional ARIMA model might be composed as SARIMA (p, d, q) (P, D, Q)s, with the non-occasional part in lowercase and the occasional part in capitalized.

Where,

- 'p' denotes the auto regressive part's order.
- The order of the moving average portion is referred to as 'q.'
- The difference order is denoted by the letter'd.'
- 'P' denotes the seasonal auto regressive part's order.
- The order of the seasonal moving average portion is referred to as 'Q.'
- The order of seasonal difference is denoted by the letter 'D.'
- The rear shift operator is referred to as 'B.'
- The duration of the seasonal period is denoted by the letter 'S.'

Sulfur dioxide (SO 2), Nitrogen dioxide (NO 2), Respirable suspended particulate matter (RSPM), and suspended particulate matter (SPM) air quality information from the India Pollution Control Board (IPCB) were used in this review for the years 2015-2020.

**Root Mean Squared Error (RMSE)**

The square base of the mean of squared mistakes is RMSE. The root mean square blunder (RMSE) measures how close to the expected qualities are to the real qualities. Subsequently, a lower RMSE number shows that the model is getting along admirably. It is utilized to decide in the RMSE beneath condition and

computation of RMSE calculation as displayed in the figure d.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}}$$

```
[ ] test['PM2.5'].mean()
    test['PM10'].mean()
    test['NO'].mean()
    test['CO'].mean()
    test['SO2'].mean()
    test['O3'].mean()
    test['AQI'].mean()
    from sklearn.metrics import mean_squared_error
    from math import sqrt
    rmse=sqrt(mean_squared_error(pred,test['PM2.5']))
    rmse1=sqrt(mean_squared_error(pred,test['PM10']))
    rmse2=sqrt(mean_squared_error(pred,test['CO']))
    rmse3=sqrt(mean_squared_error(pred,test['NO']))
    rmse4=sqrt(mean_squared_error(pred,test['SO2']))
    rmse5=sqrt(mean_squared_error(pred,test['O3']))
    rmse6=sqrt(mean_squared_error(pred,test['AQI']))
    # print(rmse)
    # print(rmse1)
    # print(rmse2)
    # print(rmse3)
    # print(rmse4)
    # print(rmse5)
    # print(rmse6)
    model2=sm.tsa.arima.ARIMA(df['PM2.5'],order=(1,0,5))
    model2=model2.fit()
    model3=sm.tsa.arima.ARIMA(df['PM10'],order=(1,0,5))
    model3=model3.fit()
    model4=sm.tsa.arima.ARIMA(df['NO'],order=(1,0,5))
    model4=model4.fit()
    model5=sm.tsa.arima.ARIMA(df['CO'],order=(1,0,5))
    model5=model5.fit()
    model6=sm.tsa.arima.ARIMA(df['SO2'],order=(1,0,5))
    model6=model6.fit()
    model7=sm.tsa.arima.ARIMA(df['O3'],order=(1,0,5))
    model7=model7.fit()
    model8=sm.tsa.arima.ARIMA(df['AQI'],order=(1,0,5))
    model8=model8.fit()
    df.tail()
```

Figured: RMSE values for AQIparameters

When comparing the proposed ARIMA (Auto Regression Integrated Moving Average) model with the other AR (Auto Regression) model. In this project, ARIMA model is used for the Feature Extraction and get a testing accuracy of 90%. By comparing the proposed algorithm with AR, seeing the existing algorithm doesnot give the higher accuracy. So, the proposed system gives higher accuracy with existing ones.
Here the accuracy can be calculated by RMSE value in which lesser the RMSE will get more accuracy. The existed system got the RMSC value more then the15 but by using the ARIMA method the RMSE value is below 8.

## V. RESULTS AND DISCUSSIONS:

This proposed project is applicable to predict the Air Quality Index Values. This prediction of values would help how the pollutants ranges in the air and how the values will be changes from day to day. In predicting the AQI value and the parameters of air by using the ARIMA model is used for the classification in our proposed system. We have taken the data set and extract features for the output as shown in below figure f.

```
start=len(train)
end=len(train)+len(test)-1
#if the predicted values dont have date values as index, you will have to uncomment
#index_future_dates=pd.date_range(start='2018-12-01',end='2018-12-30')
pred=model.predict(start=start,end=end,typ='levels').rename('ARIMA predictions')
#pred.index=index_future_dates
pred.plot(legend=True)
test['PM2.5'].plot(legend=True)
test['PM10'].plot(legend=True)
test['NO'].plot(legend=True)
test['CO'].plot(legend=True)
test['SO2'].plot(legend=True)
test['O3'].plot(legend=True)
test['AQI'].plot(legend=True)
```

```
usr/local/lib/python3.7/dist-packages/statsmodels/tsa/base/tsa_model.py:843: ValueWa
 data=self.data,
matplotlib.axes._subplots.AxesSubplot at 0x7f82ff8823d0>
```

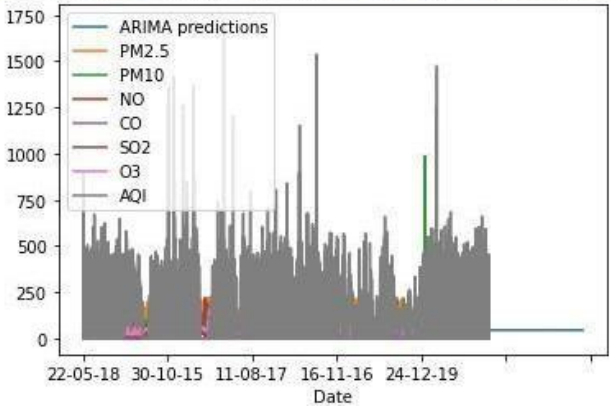


Figure f:code for training and testing datasets

The features like Particulate Matter ($PM_{2.5}$ & $PM_{10}$), Carbon Monoxide (CO), Nitrogen Dioxide ($NO_2$), Sulphur Dioxide ($SO_2$), Ozone($O_3$) and AQI values in different regions of India are considered for predicting the output. The parameters that has compared to know the relationship between any two parameters as following. The graph plotted between NO and CO to know the relationship between them as shown in the following figure i.

```
pred2.plot(figsize=(12,5),legend=True)
pred3.plot(figsize=(12,5),legend=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f82fed93d90>
```

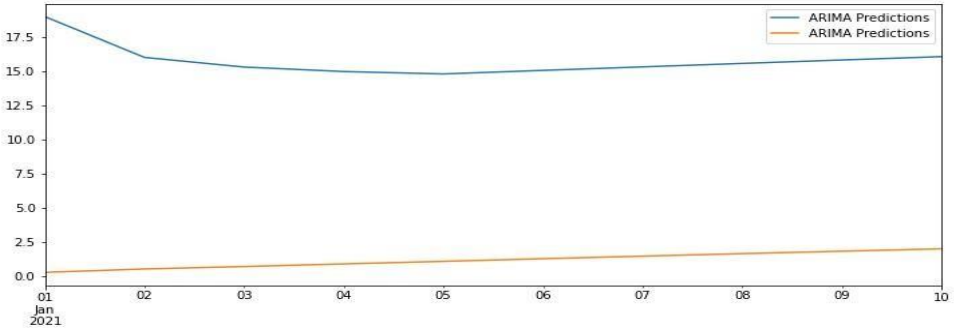


Figure i: Graph plotting to NO and CO parameters.

The graph plotted between NO2 and Ozone (O3) to know the relationship between them as shown in the following figure j.

```
[ ]  pred4.plot(figsize=(12,5),legend=True)
     pred5.plot(figsize=(12,5),legend=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f82feac2d10>
```

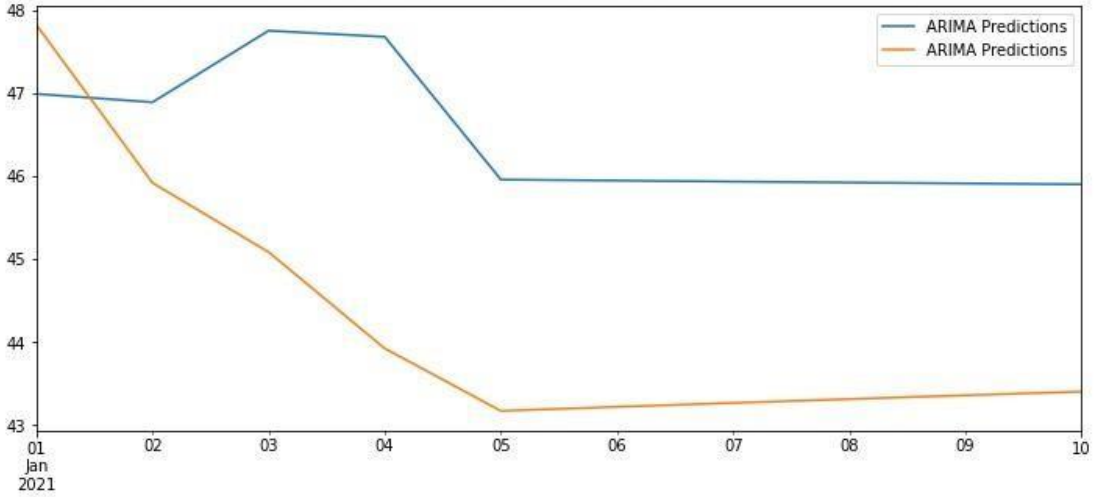


Figure j: Graph plotting between Nitrogen Dioxide(NO2)and O3parameters.

After training the dataset and applying ARIMA method get the classification accuracy of 90% by getting the RMSC value below 8. The results are shown in below figure g.

```
pred.plot(figsize=(12,5),legend=True)
pred1.plot(figsize=(12,5),legend=True)
pred2.plot(figsize=(12,5),legend=True)
pred3.plot(figsize=(12,5),legend=True)
pred4.plot(figsize=(12,5),legend=True)
pred5.plot(figsize=(12,5),legend=True)
pred6.plot(figsize=(12,5),legend=True)
# test.plot(figsize=(12,5),legend=True)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f82fec03550>
```

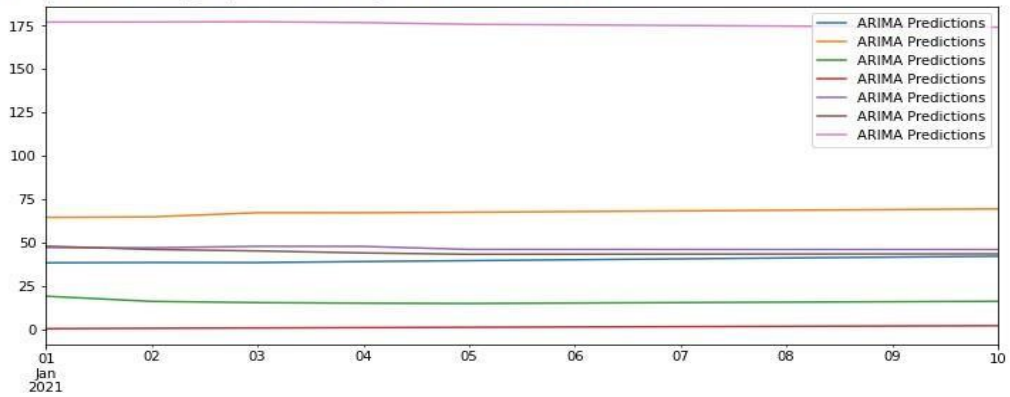


Figure g: Output of the predicted values of AQI

The inputs that has taken in the format of number of days ,starting date and ending date. The final graph plotted between all the days that in the range of user given number of days as shown in the below figure h.

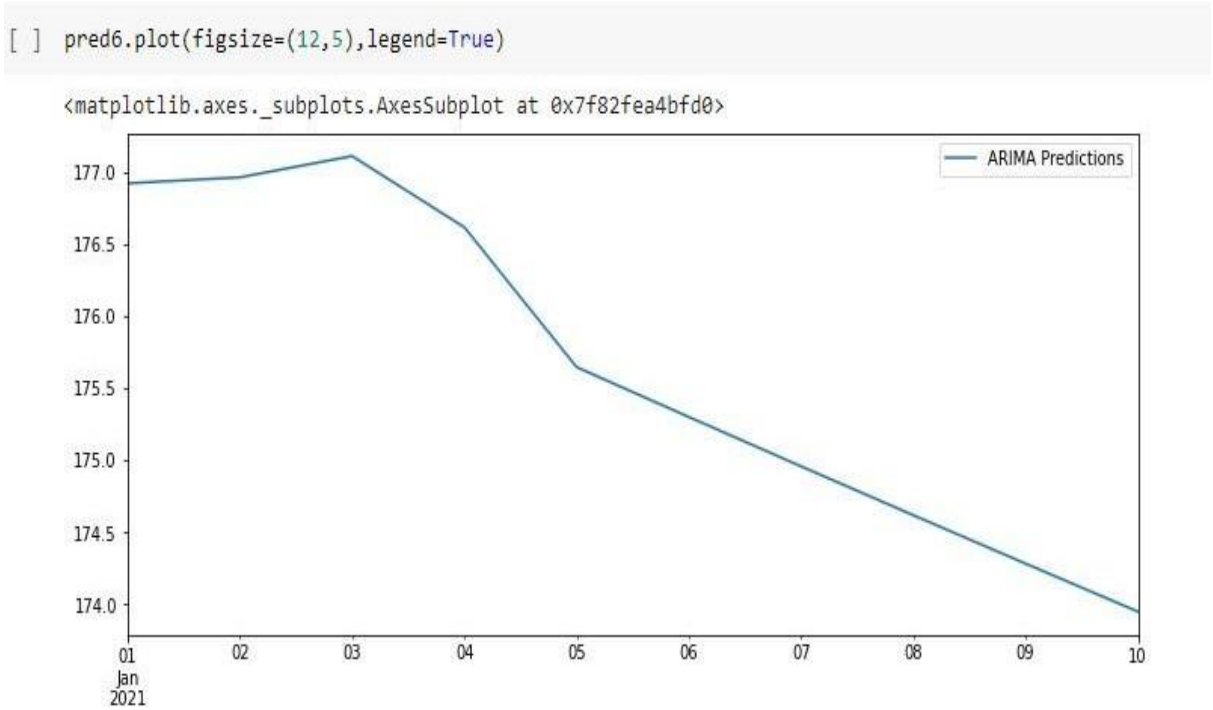


Figure h: Graph plot for AQI values in ranged input day.

**CONCLUSIONANDFUTUREENHANCEMENT**

This task "Foreseeing Air quality Index Value Using ARIMA Method" is savvy, eco-accommodating, down to earth and the most secure method for anticipating the Air Quality Index esteems precisely. Government offices utilize an air quality list (AQI) to illuminate people in general about how debased the air is presently or will become from here on out.

The AQI is determined by averaging readings from an air quality sensor, which can rise attributable to traffic, woodland fires, or different elements that add to air contamination. Ozone, nitrogen dioxide, and Sulfur dioxide, among different poisons, were. As the AQI rises, general wellbeing concerns ascend also, hurting kids, the old, and those with respiratory or cardiovascular issues. During these times, government offices frequently prescribe individuals to restrict their open air actual work or conceivably try not to go outside by any means. tried. We have looked at and tracked down a classifier that gives high exactness as per the prepared model. We have looked at changed arrangement calculations, for example, support vector machines, Artificial brain organizations, Fuzzy model, Auto relapse and Auto relapse coordinated moving normal techniques. This proposed ARIMA model increments precision by diminishing the worth of RMSE. Tests were led utilizing the Kaggel dataset, which contains five boundaries named, PM (2.5,10), CO, NO2, O3, SO2 and it is shown that the coordinating execution with the proposed insurance plot with an exactness score 90%. Matching execution, precision, and format security are improved by utilizing the proposed secure multimodal framework when contrasted with the current system. This project principally centers around anticipating the Air Quality list esteems and further develops the expectation exactness. The most recent superior calculations are being utilized to foster the exactness rate. Later on work of this exploration, more proficient techniques can be utilized and include more boundary elements to future improve the framework in distinguishing the AQI values. For future improvements the Air Quality Values can be anticipated in strange circumstances.

**REFERENCES**

[1] L.Miller and X.Xu,"Ambient PM2.5humanhealtheffects—FindingsinChinaandresearch directions",

Atmosphere,vol.9,pp.424,Oct.2018

[2] J. Chen, Y. Song, G. Wang and Q. Liu, "WSN-aided haze pollution governance: Modelling public willingness based on structuralequations",Int.J.SensorNetw.,vol.29,no.2, pp.111-120,2019.

[3] P. E. Saide, G. R. Carmichael, S. N. Spak, L. Gallardo, A. E. Osses, M. A. Mena-Carrasco, et al., "Forecasting urban PM10 and PM 2.5pollutionepisodesinverystablenocturnalconditionsandcomplexterrainusingWRF–ChemCOtracer model ",Atmos.Environ.,vol.45,no.16,pp.2769-2780,May2011.v

[4] [4]. X. Yang, Z. Zhang, Z. Zhang, L. Sun, C. Xu and L. Yu, "A long–term prediction model of beijing haze episodes using time seriesanalysis", Comput.Intell.Neurosci.,vol.2016,pp.1-7,Jul.2016.

[5] J.J.Carbajal-Hernández,L.P.Sánchez-Fernández,J.A.Carrasco-OchoaandJ.F.Martínez-Trinidad,

[6] "Assessment and prediction of airquality using fuzzy logic and autoregressive models",Atmos. Environ.,vol.60,pp. 37 -50, Dec. 2012.[6]. P.J.G.Nietoo,E.Combarro, J.DelCoz DíazandE.Montañés,"ASVM-basedregressionmodeltostudytheairqualityatlocal scalein

[7] Oviedourbanarea(NorthernSpain):Acasestudy",Appl.Math.Comput.,vol.219,no.17,pp.8923-8937,May2013.

[8] A.Challoner, F.Pilla and L.Gill, "Prediction of indoor air exposure from outdoor air quality using an artificial neural network model for inner city commercial buildings",Int. J. Environ. Res. PublicHealth, vol.12,no. 12,pp. 1523315253,Dec.2015.

[9] Y. Bai, Y. Li, X. Wang, J. Xie and C. Li, "Air pollutants concentrations forecasting using back propagation neural network based onwavelet decomposition with meteorological conditions", Atmos.Pollution Res., vol.7, no.3, pp.557-566,May2016.

[10] Q. Wu and H. Lin, "A novel optimal-hybrid model for daily air quality index prediction considering air pollutantfactors",Sci.TotalEnviron.,vol.683,pp.808-821,Sep.2019.

[11] R.E.Kalman, "Anewapproachtolinearfilteringandpredictionproblems",J.BasicEng.,vol.82, no.1,pp.35-45,Mar.1960.

[12] J. Guo, W. Huang and B.M. Williams,"Adaptive Kalman filter approach forstochastic short-term trafficflow rate prediction and uncertainty quantification",Transp. Res.CEmerg.Technol., vol.43,pp.50-64,Jun.2014.

[13] S.Hua,S.Wang,S.Jin,S.FengandB.Wang,"Wind speed optimization method of numerical prediction for wind farm based on Kalman filter method",J.Eng.,vol.2017,no.13,pp.1146-1149,Jan.2017.

[14] K. Chen and J. Yu, "Short-term wind speed prediction using an unscented Kalman filter based state-spacesupportvectorregressionapproach",Appl.Energy,vol.113,pp.690-705,Jan.2014.

[15] G.Galanis,P.Louka,P.Katsafados,I.PytharoulisandG.Kallos,"ApplicationsofKalmanfiltersbasedonnon-linearfunctionstonumericalweatherpredictions",Ann. Geophys.,vol. 24,no. 10,pp.2451-2460,Jun.2010.

[16] M.ChaabeneandM.BenAmmar,"Neuro-fuzzy dynamic model with Kalman filter to forecast irradiance and temperature for solar energy systems",Renew.Energy,vol.33,no.7,pp.1435-1443,Jul.2008.

[17] X.Lai, T.Yang,Z.Wang and P. Chen,"IoTimplementation of Kalman Filterto improve accuracy of airquality monitoring andprediction", Appl.Sci.,vol.9,no.9,pp.1831,May2019.

[18] M. Mastali, J. Vazquez-Arenas, R. Fraser, M. Fowler, S. Afshar and M. Stevens, "Battery state of the charge estimation using Kalman filtering",J.PowerSour.,vol. 239,pp.294-307,Oct. 2013.

[19] T.Soubdhan,J. Ndong,H.Ould-BabaandM.-T.Do, "Arobust forecasting frame work based on the

[20] Kalman filtering approach with a twofold parameter tuning procedure: Application to solar and photovoltaic prediction", SolarEnergy,vol.131,pp.246-259,Jun.2016.

[21] G.Rigatos and P.Siano,"Forecasting of power corporations' default probability with nonlinear Kalman filtering", IEEE Syst.J.,vol.no.2,pp.1099-1107,Jun.2018.

[22] Z.Yang,L.CeandL.Lian,"ElectricitypriceforecastingbyahybridmodelcombiningwaveletransformARMA andkernel-basedextremelearning machinemethods",Appl.Energy,vol.190,pp.291-305,Mar.2017.

[23] Y.Zhang,G.Jia,N.LiandM.Bai,"A novel adaptiveKalman filter with colored measurement noise", IEEE Access,vol.6,pp.74569-74578,2018.

[24] R.Ferrero,F.GandinoandM.Hemmatpour,"EstimationofdisplacementforInternetofThingsapplicationswithKalmanfilter",Electronics,vol.8,no.9,pp.985,Sep.2019

[25] Y.Liu,T.Zhi,H.Xi,X.DuanandH.Zhang,"AnovelcontentpopularitypredictionalgorithmbasedonautoregressivemodelininformationвЂ“centricIoT",IEEEAccess,vol.7,pp.27555-27564,2019.
[26] R.Ferrero,F.GandinoandM.Hemmatpour,"EstimationofdisplacementforInternetofThingsapplicationswithKalmanfilter",Electronics, vol.8,no.9,pp.985,Sep.2019.
[27] AshutoshRath,Kirtisudhabhoi,SandeepSamantary,Prof.PrakashChandraSwain,"FlowForecastingofHirakudReservoirwithARIMAModel",IEEEAccess,vol.9,pp.2952,ICECDS-2017.
[28] RinaldiRusli,AchmadNizarHidayanto,YovaRuldeviyani,"ConsumptionPredictiononSteamPowerPlantUsingAutoRegressiveIntegratedMoving Average",IWBIS,vol.9,no. 10,pp.15,june2019.
[29] AdebiyiA.Ariyo,AdewumiO.Adewumi,CharlesK.Ayo,"StockPricePredictionUsingtheARIMAModel",IEEE,vol.11,no.12,pp.253,Feb.2015.
[30] A.Meyler,G.KennyandT.Quinn,"ForecastingIrishInflationusingARIMAModels",CentralBankofIreland ResearchDepartmentTechnicalPaper,1998.
[31] M.khasel,M.BijariandG.A.RArdali,ImprovementofAuto-RegressiveIntegratedMovingAverageModelsusingFuzzylogic",pp.956-967,2009.
[32] C.Lee,C.Ho,"Short-TermloadforecastingusingliftingschemeandARIMAmodel",ExpertSystemwithApplications",vol.38,no.5,pp.5902-5911,2011.
[33] [31].C.Wang,"A comparison study of between fuzzy time series model and ARIMA model for forecasting Taiwan Export",ExpertSystemwithApplications,vol.38,no.8,pp.9296-9304,2011.
[34] P.Pai,C.Lin,"Ahybrid ARIMA and Suppor tVector machines model in stock price prediction", Omega,vol.33, pp.758 -766, 2012.[33]. Jan Mochnac, Stanislav Marchevsky, Pavol Kocan, "Bayes ian Filtering Techniques: Kalman and Extended Kalman Filter Basics", 19th International Conference Radio elektronika,2009.
[35] Jibo Chen, Keyao Chen,Chen Ding,Guizhi Wang,Qi Liu and Xiaodong Liu," An Adaptive Kalman Filtering Approach to Sensing and PredictingAir Quality IndexValues",IEEEAccess,Vol.8, 2020.