JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

PREDICTION MODEL ON AIR QUALITY OF PUNE REGION USING SOFT COMPUTING TOOLS

Prof. Sneha Khedekar, Akshita Patil, Sharyu Patil, Viraj Salunke ,Samir Shanware
Department of civil engineering, All India Shri Shivaji Memoria Society's college of Engineering Pune, Savitribai Phlue Pune University.

Abstract: Recently, global pollution has increased and the increase in the content of harmful substances ($PM_{2.5}$ and PM_{10}) in the air can cause adverse effects such as heart disease of insects and pose a threat to the country and people. Monitoring and predicting air quality have become essentially important in this era, especially in developing countries like India. Air quality models simulate the atmospheric environment systems and provide increased domain knowledge and reliable forecasting. They provide early warnings to the population and reduce the number of measuring stations. Due to the complexity and non-linear behavior associated with air quality data, soft computing models became popular in air quality modeling (AQM) .$PM_{2.5}$ and PM_{10} are a deadly pollutant, a mixture of solids and coarse liquids with a diameter of 2.5 microns and 10 microns respectively .It is necessary to accurately estimate the $PM_{2.5}$ & PM_{10} concentration in order to prevent the public from being affected by the weather. Changes in $PM_{2.5}$ & PM_{10} depend on factors such as the weather and the amount of other pollutants in the city.We implemented soft computing tool to predict the daily forecast $PM_{2.5}$ and PM_{10} concentration in Pune, India, based on SARIMA model with a spatial-temporal feature by combining historical data of pollutants, meteorological data, and $PM_{2.5}$ PM_{10} concentration in the selected station.

Keywords: Global pollution, SARIMA model, soft computing tool, historical data ,meteorological data

I. INTRODUCTION

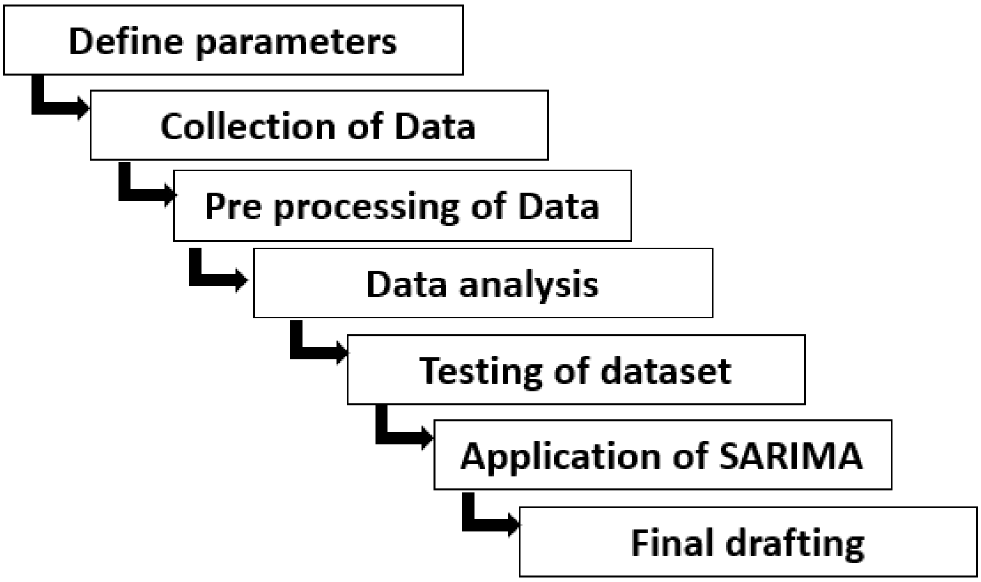
Located in the state of Maharashtra, Pune has a population of around 7 million and thecity has been growing rapidly in recent years. Air quality in cities remains an ongoing problem: fine particulate matter levels (PM2.5) consistently exceed Indian standards and World Health Organization guidelines. The city uses local experts to play an important role in solving public health problems caused by air pollution. Air pollution affects many countries around the world, causing serious illness or death. The increased dependence on fossil fuels over the past century is responsible for the worsening of our climate. The exhaust fumes emitted by many vehicles also cause serious pollution.
To reduce human health concerns about air pollution, the government should make recommendations to better control pollution using data that tracks pollution over time. Air quality in Pune, India has declined over the past five years. Over the past two decades, research on the epidemiology of air pollution has increasingly focused on the effects of exposure to small particles. An analysis based on international transportation standards and satellite monitoring shows that annual PM2.5 air pollution in Pune is consistently exceeding the health limit.
The main causes of outdoor pollution are materials and liquids called aerosols and gases from vehicle emissions, constructions, factories, burning straw and fossil fuels, and forest fires. The main causes of indoor air pollution are air pollution from cooking (such as wood, crop waste, coal and manure), humidity, mold fumes, chemicals in fabric cleaning papers, etc. Currently, Pune's 5th concentration is 14.1 times the World Health Organization's current annual guidelines. In general, air quality in Pune starts to deteriorate at the end of October. In terms of bad weather, winter is the worst season. Traffic congestion not only harms people but is also a major source of pollution. It can also affect a person's respiratory system. A survey conducted by the city's agency to measure public awareness of air pollution revealed that almost half of respondents did not know that Pune is an unsustainable city – a city that does not meet national air standards.

Air pollution in the city has been increasing since 1 November, and particulate matter both PM10 and PM2.5 exceeds the maximum limit. Climate forecasting is an investment worth making at many levels individual, community, national and global. Forecasts help people plan ahead, reducing health impacts and associated costs. Moving on, SAFAR's air pollution report says the city's air quality could deteriorate by PM2.5 if the temperature is low then the value will be higher than 100 grm. It is said Pune's Air Quality index (AQI) showed air quality was on the "weak end" on Tuesday and could be "poor" or "on the high end of average" for the next two days.

Of the four cities currently monitored by SAFAR (Mumbai and Ahmedabad), Pune and Delhi are the only two to experience a decline in air quality. Air quality in Pune, which has been mostly mode rate since Diwali deteriorated on Tuesday mainly due to the high temperature. " This is one of the reason for the bad weather in Delhi and Pune because both have a night temperature of 8-9oC," an IITM-SAFAR researcher told TOI. SAFAR data shows that air quality has dropped to "poor" in places like Shivajinagar and Alandi , while areas like Kothrud, Hadapsar, Bhosari and Nigdi have poor air quality. According to Indian Meteorological Department (IMD), morning thunderstorms were also observed in Pune mainly due to pollution associated with high temperatures. With rapid development and urbanization, many countries are suffering from serious air pollution. Exposure to high levels of fine particles (PM2.5) can cause serious health problems due to PM2.5. There are small wastes or small liquids to be drawn deep into the human lungs. Early climate monitoring and forecasting is critical to human health and government decision-making.

II. METHODOLOGY

Figure 1: Methodology flowchart



Various work done is specified as follows:

PARAMETERS:

- There are many pollutants in our air, but we only watch/ observe those that are not important to us.
- We consider these two important as they exceed PM2.5 and PM10 limits.
- Particulate matter (also known as particulates or PM) consists of airborne particles, including dirt, dust and smoke and small amounts of liquid. There are three sizes of particulate matter in the air: PM10 and PM2.5.

COLLECTION OF DATA:

- Data collected are PM2.5 and PM10 in Shivajinagar region from 2016 to 2021.
- Measured data collected for 2021.
- Source: IITM (Indian Institute of Tropical Metrology)

PRE-PROCESSING OF DATA:

- Data representation and performance are important before run and analysis.
- Data prefixes to check for nulls and related values to get correct data.

EXPLORATORY DATA ANALYSIS:

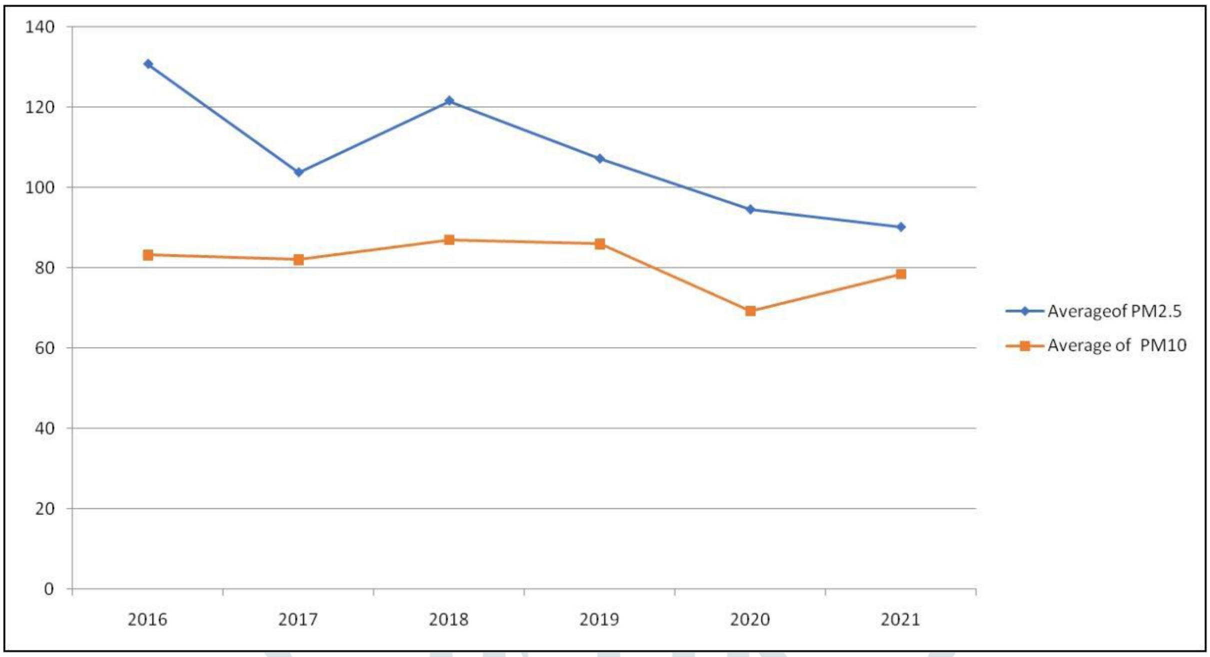


Fig. No. 2.1 Graph of average yearly readings of PM2.5 and PM10 pollutants

The graph is for PM2.5 and PM10 average yearly readings.
On X axis : Year
On Y axis : Average yearly value of PM2.5 & PM10.
Above fig shows that concentration of PM2.5 is more than PM10 .Concentration PM10 of has been increasing gradually from year 2018 while concentration of PM2.5 was maximum in 2016.

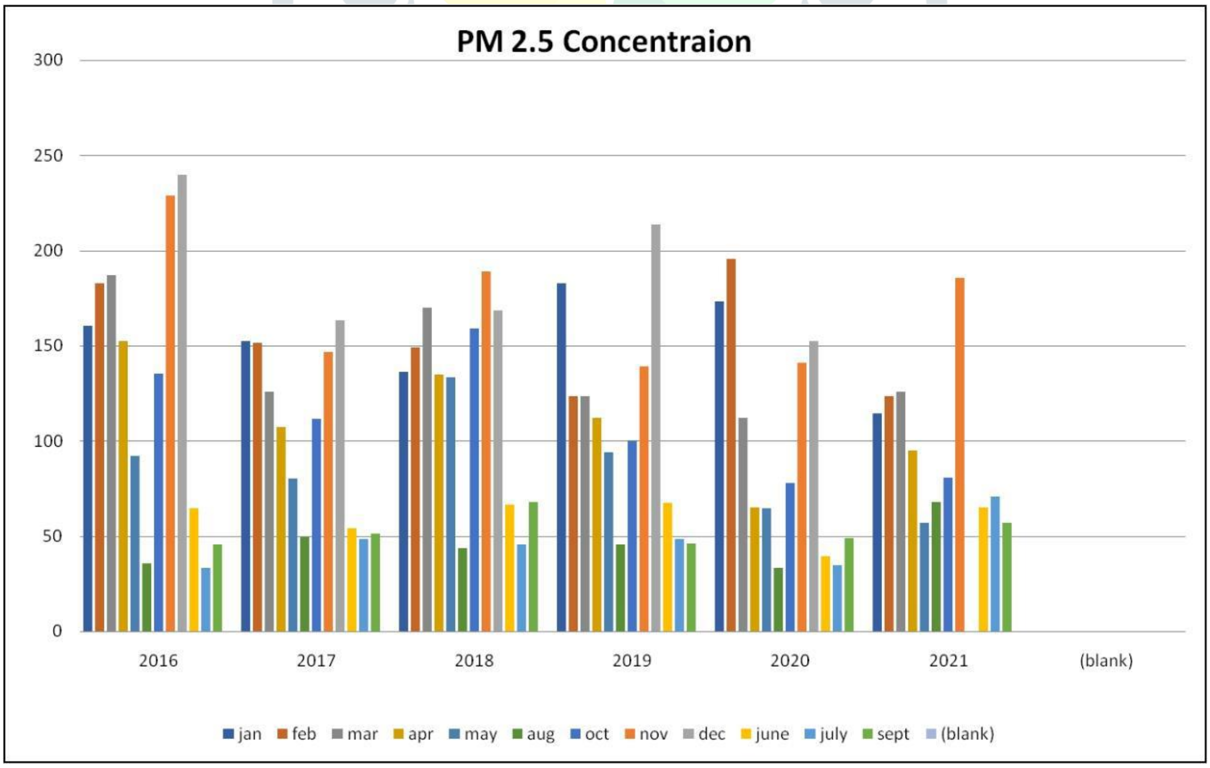


Fig. No. 2.2. Graph of monthly average readings of PM2.5 concentration
This graph depicts PM2.5 concentration from year 2016-2021(monthly average).
Above fig shows month-wise concentration of PM2.5. During month of November and December the concentration is more. As colder and drier air traps more pollution.

PEARSON CORRELATION COEFFICIENT

• Pearson relationship coefficient could be a degree of straight relationship between two sets of data.

• It is the proportion between the covariance of two factors and the item of their standard deviations; in this way, it is basically a normalized estimation of the covariance, such that the result continuously incorporates a esteem between −1 and 1.

• A key numerical property of the Pearson correlation coefficient is that it is invariant beneath partitioned changes in area and scale within the two variables.

• The relationship coefficient ranges from −1 to 1.

• A esteem of +1 infers that all information focuses lie on a line for which Y increments as X increments, and bad habit versa for −1.

• A esteem of infers that there's no straight reliance between the factors.

- Month-wise correlations has been done for year 2021 for the Shivajinagar station .
- All months from Jan to Dec (2021) has been considered to find out the relationships between pollutants month-wise.
-

Table No 2.1: Month wise correlation

Month	$PM_{2.5}$ & PM_{10}
Jan	High
Feb	High
March	High
April	Very High
May	Moderate
June	Moderate
July	High
Aug	High
Sept	High
Oct	High
Nov	High
Dec	NA

III SOFTWARE DESCRIPTION

Introduction to SARIMA

• The Seasonal Autoregressive Integrated Moving Average, SARIMA, or Seasonal ARIMA is an extension of ARIMA that supports non-uniform data on a seasonal basis.

• Adds three new hyperparameters and one more seasonal parameter to show the autoregressive (AR), variance (I) and moving average (MA) seasonal characteristics of the series.

• We can do one of two things. To explain the remaining (unexplained) variable, we can add another factor to our SARIMA model or create a seasonal correction, i.e. a new Y variable.

• Autoregressive Integrated Moving Average, or ARIMA, is one of the most widely used methods of time estimation. Although this method can process the data with the model, it does not support the season and time to support direct modeling of the seasonality of the series, an extension of ARIMA is called SARIMA.

• Seasonal ARIMA models were created by incorporating additional seasonal elements into ARIMA.

• Seasonal ARIMA models use equal lagged variances between seasons to eliminate additional seasonal effects. Like 1-differences delayed to eliminate variance, lagged s-differences represent the mean time. Seasonal Arima models include lagged autoregressive and moving average points.

• Seasonal ARIMA models can have multiple variables and time combinations. Therefore, while fitting the data, it is necessary to try several models and select the most suitable model using appropriate criteria.

CONFIFURATION OF SARIMA

Configuring a SARIMA requires selecting hyperparameters for both the trend and seasonal elements of the series.

TREND ELEMENTS

There are three trend elements that require configuration. They are the same as the ARIMA model; specifically:

• p: Trend autoregression order.

• d: Trend difference order.

• q: Trend moving average order.

SEASONAL ELEMENTS

There are four seasonal elements that are not part of ARIMA that must be configured; they are:

• P: Seasonal autoregressive order.

• D: Seasonal difference order.

• Q: Seasonal moving average order.

• m: The number of time steps for a single seasonal period.

The SARIMA time series forecasting method is supported in Python via the stats model library. To use SARIMA there are three steps, they are:

1. Define the model.
2. Fit the defined model
3. Make a prediction with the fit model.

IV **APPLICATION OF SARIMA MODEL**

SARIMA for PM2.5

- In all 2 years (730 days) wer considered – 1st Jan, 2020 to 31st Dec, 2021.
- Training data – First 23½ months
- Predicted period – Last 15 days (From 15th Dec to 31st Dec, 2021).
- non-seasonal (trend): AR(3), I(2), MA(3)
- seasonal: AR(1) I(1) MA(4)
- Residual: 8.54
- Model Used for prediction: SARIMAX (ts_train, order =(3,2, 3) , seasonal_order =(1,1,4,275)

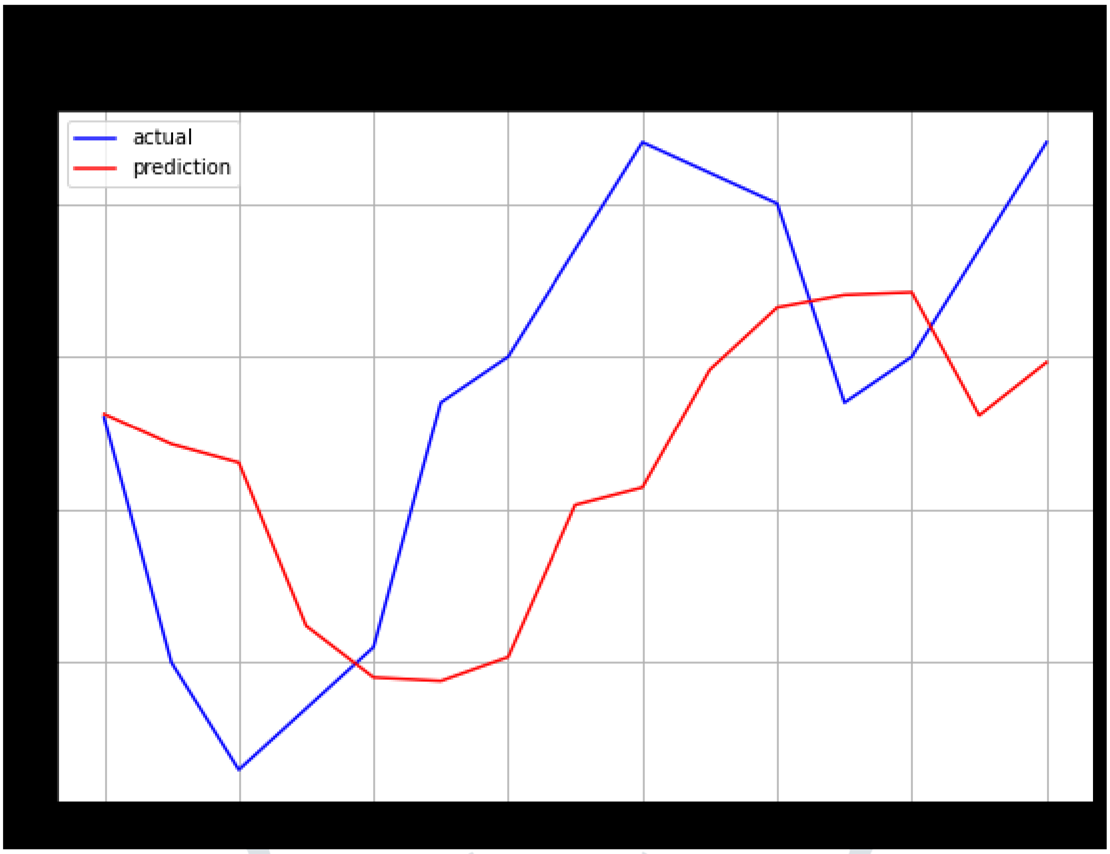


Fig. No. 4.1 SARIMA Forecasting For PM2.5

The above graph shows that there is less difference between actual values and predicted values. The prediction is for 15 days while actual values are of 715 days.

SARIMA for PM10

- In all 2 years (730 days) were considered – 1st Jan, 2020 to 31st Dec, 2021.
- Training data – First 23½ months
- Predicted period – Last 15 days (From 15th Dec to 31st Dec, 2021).
- non-seasonal (trend): AR(1), I(2), MA(4)
- seasonal: AR(3) I(1) MA(4)
- Residual: -16
- Model Used for prediction: SARIMAX (ts_train, order=(1,2,4) , seasonal_order= (3,1,4,275)

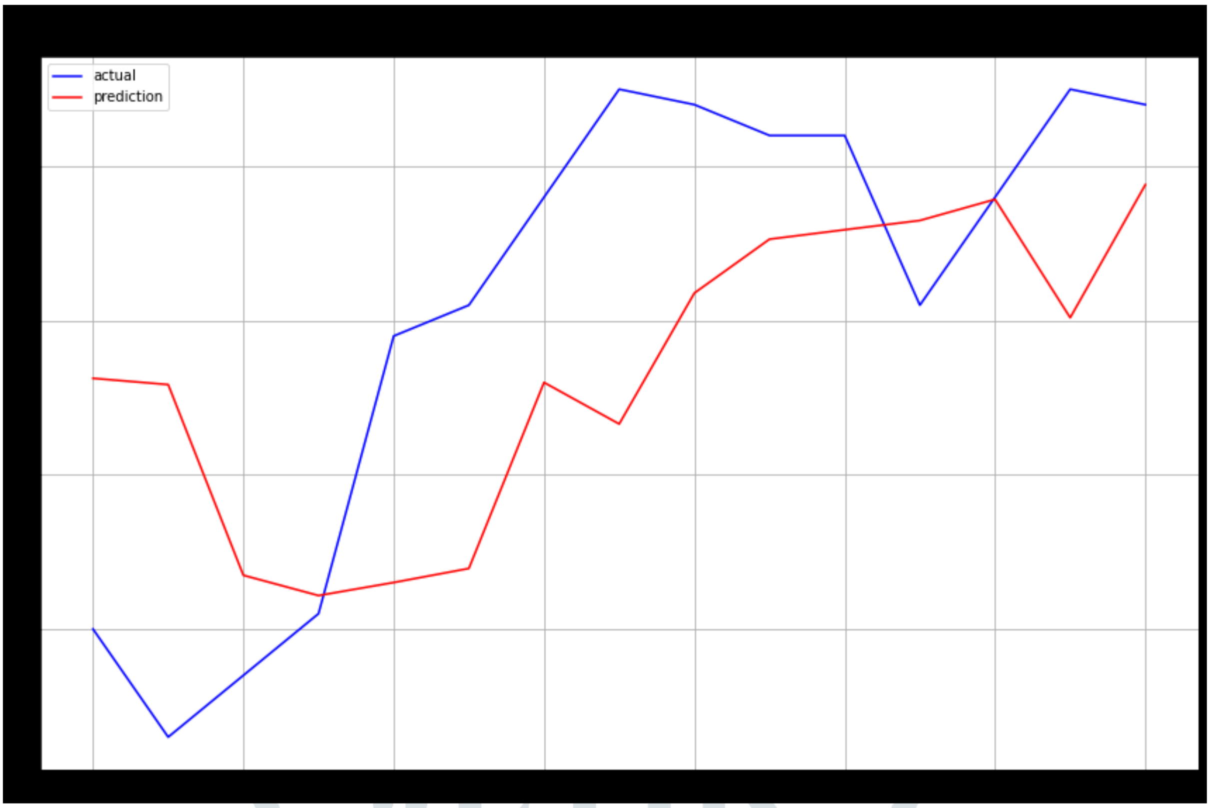


Fig. No. 4.2 SARIMA Forecasting For PM10

The above graph shows that there is less difference between actual values and predicted values.
The prediction is for 15 days while actual values are of 715 days.

V CONCLUSION

SARIMA model provided insights into:
o Seasonality: Expected values of a pollutant based on annual weather seasons.
o Trend: Expected values of a pollutant due to multi-year trend (population, global warming).
o Residuals: Expected noise/variation in daily recorded values based on external metrological factors and unpredictable variations.
•It allows to create model for each of the above separately and allowing them to be analyzed and optimized separately.
•Once optimized, it provides accurate coefficients which are human interpretable

Drawbacks of SARIMA Model
• Globally Fit – a single model cannot fit all variations: one-size fit all, fit no one.
SARIMA prediction error:
• SARIMA cannot fit best for entire year.
• Individual SARIMAs fit each season, each month, each week, each day.
• But SARIMAs are difficult to optimize, multiple models = more difficult.

EXPECTED OUTCOMES
As discussed above, the chosen SARIMA model can be used to predict future profits as forecast accuracy is achieved and also takes into account seasonal factors. To solve this weather problem, accurate forecasting model will be prepared from PM2.5 and PM10 emission data in Shivajinagar region. Air quality will be determined according to this model.

VI REFERENCES

• Kazi Erum ,Study of Air quality in different areas of Pune City. International Journal of creative Research Thoughts (IJCRT) Volume 8, Issue 3 March 2020
• Rohan S Towar, Keerthi Gurani,U.G Student Civil Department, Renuka Binayake Professor Civil Department,Air Pollution Monitoring in Pune City. IJIRT | Volume 8 Issue 11
• Khaiwal Ravindra et.al. , Air Pollution in India: Bridging the Gap between Science and Policy. Journal of Hazardous, Toxic, and Radioactive WastePublished by American Society of Civil Engineers.
• Suman Yadav IIT Bombay. The effect of climate and meteorological changes on particulate matter in Pune, India. Enviornmental Monitoring and Assessment

• Sandeep Sunori ,Developing Soft Computing based Models for Prediction of Pollutant PM10 of Air. IEEE (INSTITUTE OF ELECTICAL AND ELECTRONIC ENGINEERS)
• Surendra Roy ,Prediction of Particulate Matter Concentrations Using Artificial Neural Network. Resources and Environment · March 2012 .
• Doreswamy∗, K. S. Harish Kumar, and Ibrahim Gad ,Time series analysis for prediction of PM 2.5 using SARIMA model on TAQMN data . Journal of Computational and Theoretical Nanoscience.
• PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network. Scientific Reports
• Erdinç Aladağ ,Forecasting of particulate matter with a hybrid ARIMA model based on wavelet transformation and seasonal adjustment. Science Direct
• Nur Haizum et.al., 2013 ,Forecasting of Air Pollution Index with Artificial Neural Network . Jurnal Teknologi
• Sandeep Sunori ,Developing Soft Computing based Models for Prediction of Pollutant PM10 of Air. IEEE (INSTITUTE OF ELECTICAL AND ELECTRONIC ENGINEERS)
• Surendra Roy ,Prediction of Particulate Matter Concentrations Using Artificial Neural Network. Resources and Environment · March 2012 .
• Doreswamy∗, K. S. Harish Kumar, and Ibrahim Gad ,Time series analysis for prediction of PM 2.5 using SARIMA model on TAQMN data . Journal of Computational and Theoretical Nanoscience.
• PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network. Scientific Reports
• Erdinç Aladağ ,Forecasting of particulate matter with a hybrid ARIMA model based on wavelet transformation and seasonal adjustment. Science Direct
• Nur Haizum et.al., 2013 ,Forecasting of Air Pollution Index with Artificial Neural Network .