TIME SERIES ANALYSIS(MTH517)

Prediction of Time Series Data Using RNN and ARIMA Models

-Kshitij Yeotikar (160345)

-Namit Garg (160429)

-Madhurish Gupta (160374)

-Vishal Rathore (160803)

-Akashdeep Singh (160072)

Mentor-Prof. AMIT MITRA

Introduction

The project consists of two models based on ARIMA process for predicting air quality and the other on Recurrent Neural Network approach for prediction of stock prices .
This model uses MA and ARIMA processes to predict the air quality based on time series data for 18 years from 2000-2018. We took 24 stations from the city of Madrid and used the data for different gases for 18 years. Our AIM is to predict the quality of air and amount of pollutant present in next two years.

Dataset

1. We used a publicly available dataset which comprises of 24 stations from the city of Madrid. Data set consist of amount of pollutants present in air each hour for a time span of 18 years for each station. We considered the amount of gases present in air for 10 gases. We considered the data which consist of the concentration of various gases for every hour and for every station.

ARIMA MODEL

ARIMA is a very popular statistical method for time series forecasting. ARIMA stands for **Auto-Regressive Integrated Moving Averages**. ARIMA models are applied in some cases where data show evidence of non-stationarity, where an initial differencing step can be applied one or more times to eliminate the non-stationarity.

ARIMA has three components – AR (autoregressive term), I (differencing term) and MA (moving average term).

1) AR term refers to the past values used for forecasting the next value. The AR term is defined by the parameter 'p' in arima. The value of 'p' is determined using the PACF plot.

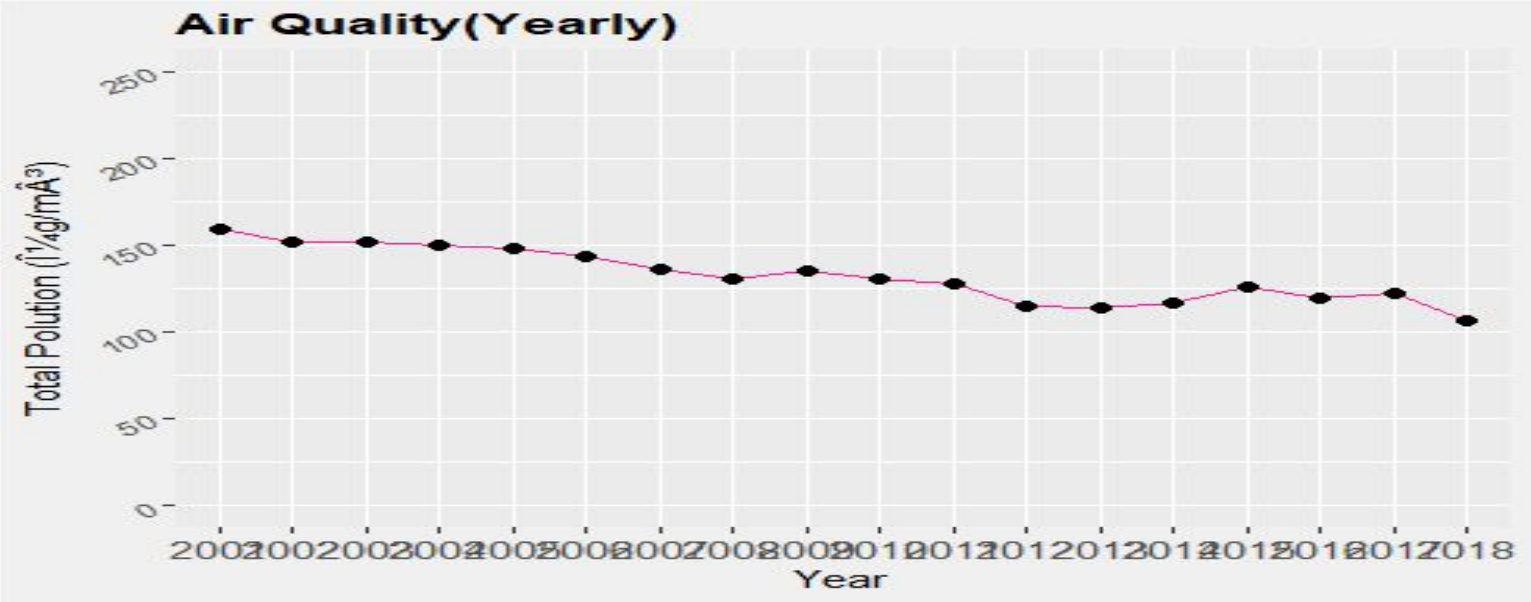
2) MA term is used to defines number of past forecast errors used to predict the future values. The parameter 'q' in arima represents the MA term. ACF plot is used to identify the correct 'q' value.

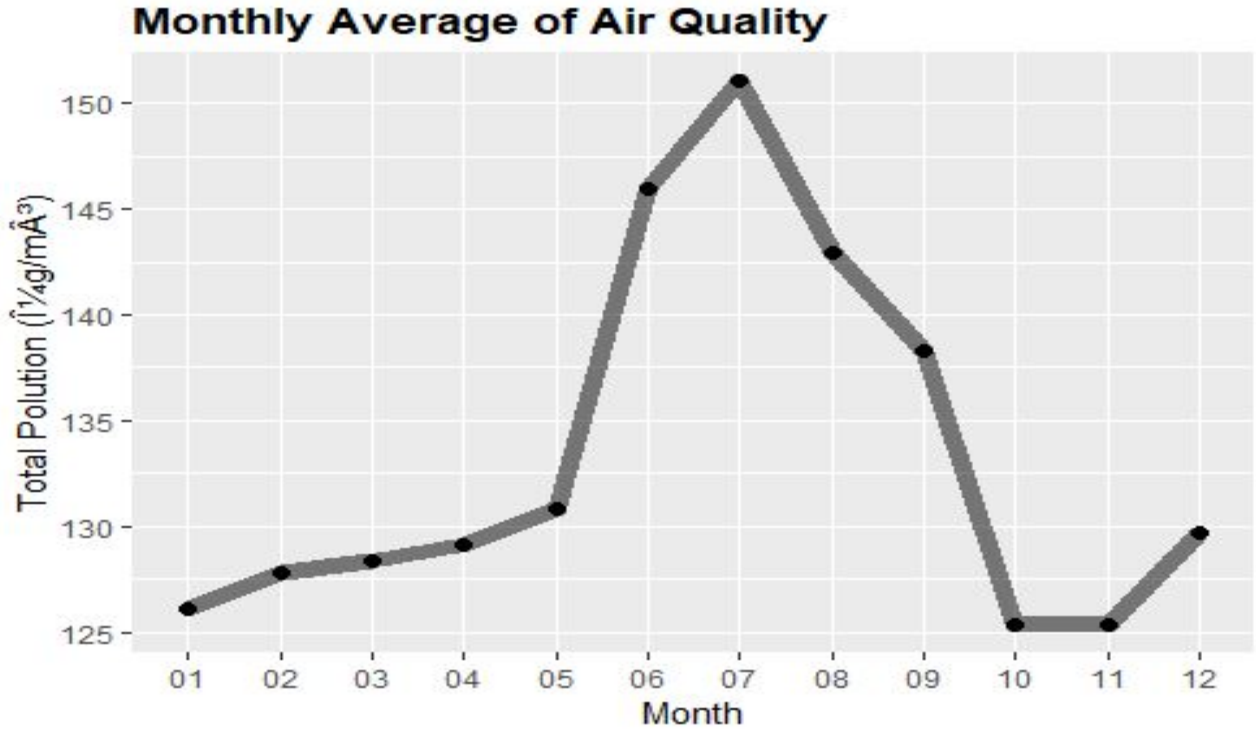
3) Order of differencing specifies the number of times the differencing operation is performed on series to make it stationary. Test like ADF and KPSS can be used to determine whether the series is stationary and help in identifying the d value.

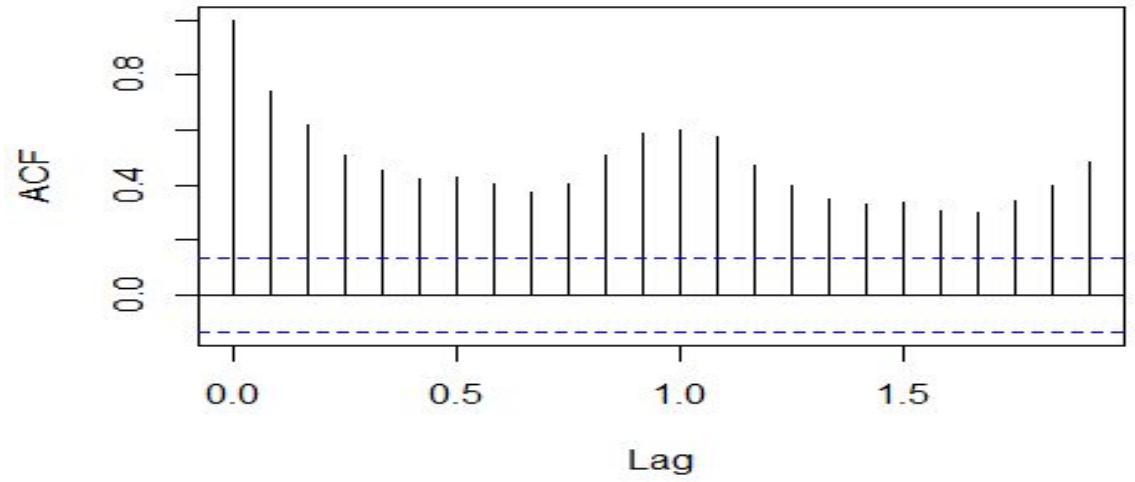
The general steps to implement an ARIMA model are –

1. **Load the data:** The first step for model building is of course to load the dataset
2. **Preprocessing:** Depending on the dataset, the steps of preprocessing will be defined. This will include creating timestamps, converting the dtype of date/time column, making the series univariate, etc.
3. **Make series stationary:** In order to satisfy the assumption, it is necessary to make the series stationary. This would include checking the stationarity of the series and performing required transformations
4. **Determine d value:** For making the series stationary, the number of times the difference operation was performed will be taken as the d value
5. **Create ACF and PACF plots:** This is the most important step in ARIMA implementation. ACF PACF plots are used to determine the input parameters for our ARIMA model
6. **Determine the p and q values:** Read the values of p and q from the plots in the previous step
7. **Fit ARIMA model:** Using the processed data and parameter values we calculated from the previous steps, fit the ARIMA model
8. **Predict values on validation set:** Predict the future values.

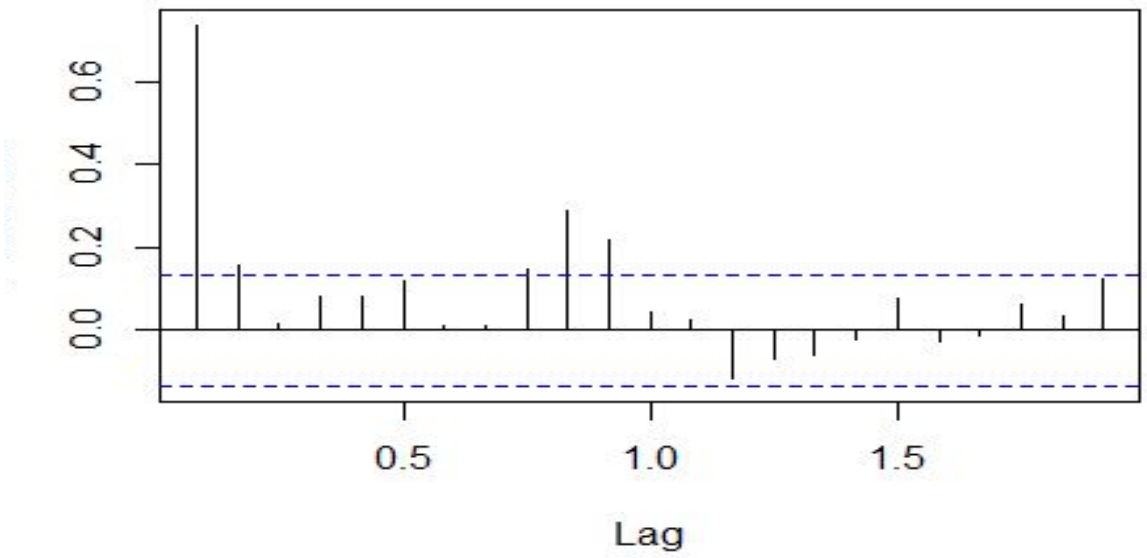
GRAPHS


Air Quality(Yearly)
Total Polution (Î¼g/mÂ³)
250
200
150
100
50
0
Year

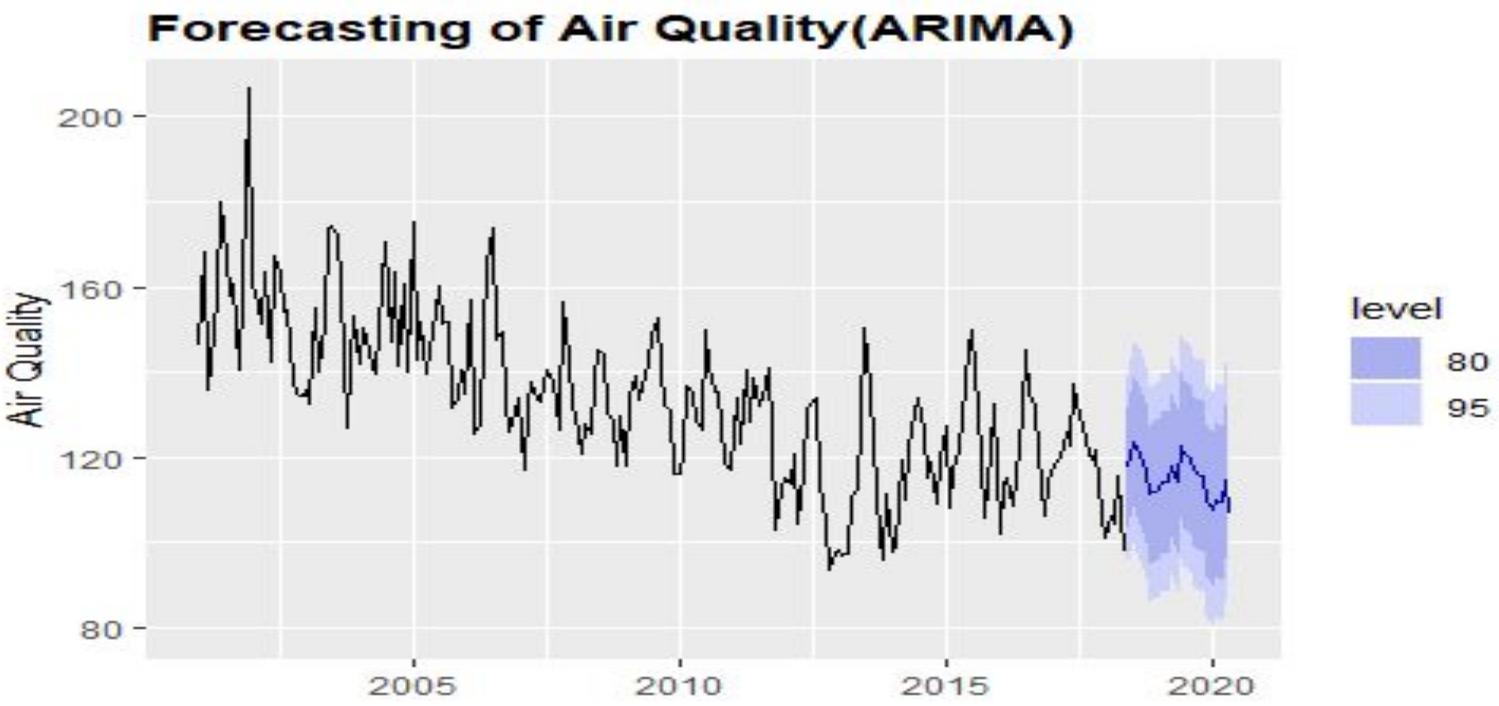

Monthly Average of Air Quality
Total Polution (Î¼g/mÂ³)
150
145
140
135
130
125
01
02
03
04
05
06
07
08
09
10
11
12
Month



ACF Plot to initialise the value of parameter p.



PACF Plot to initialise the value of parameter q.



We finally predicted the Air Quality for the next 2 years and found that Air quality decreases in madrid.