Vidyavardhini's College of Engineering and Technology

Department of Artificial Intelligence & Data Science

AY: 2023-24

Class:	TE	**Semester:**	VI
Course Code:	CSL601	**Course Name:**	Data Analytics and Visualization

Name of Student:	Ojasi Prashant Prabhu
Roll No.:	43
Experiment No.:	5
Title of the Experiment:	Implementation of ARIMA model in Python / R.
Date of Performance:	
Date of Submission:	

Evaluation

Performance Indicator	Max. Marks	Marks Obtained
Performance	5	
Understanding	5	
Journal work and timely submission	10	
Total	20	

Performance Indicator	Exceed Expectations (EE)	Meet Expectations (ME)	Below Expectations (BE)
Performance	4-5	2-3	1
Understanding	4-5	2-3	1
Journal work and timely submission	8-10	5-8	1-4

Checked by

Name of Faculty : Ms Bhavika Gharat

Signature :

Date :

Experiment No 5

Aim- Implementation of ARIMA model in **R** Programming.

Objective- To Understand use of Auto-Regression Integrated Moving Average Time Series Model

Theory-

In R programming , data analysis and visualization is so easy to learn the behavior of the data. Moreover, the **R** language is used mostly in the data science field after Python. Time series analysis is a type of analysis of data used to check the behavior of data over a period of time. The data is collected over time sequentially by the **ts()** function along with some parameters. It helps in analyzing the pattern of the data over a graph. There are many techniques used to forecast the time series object over the plot graph but the **ARIMA model** is the most widely used approach out of them.

Time Series Forecasting

Time series forecasting is a process of predicting future values with the help of some statistical tools and methods used on a data set with historical data. Some of the applications of time series forecasting are:

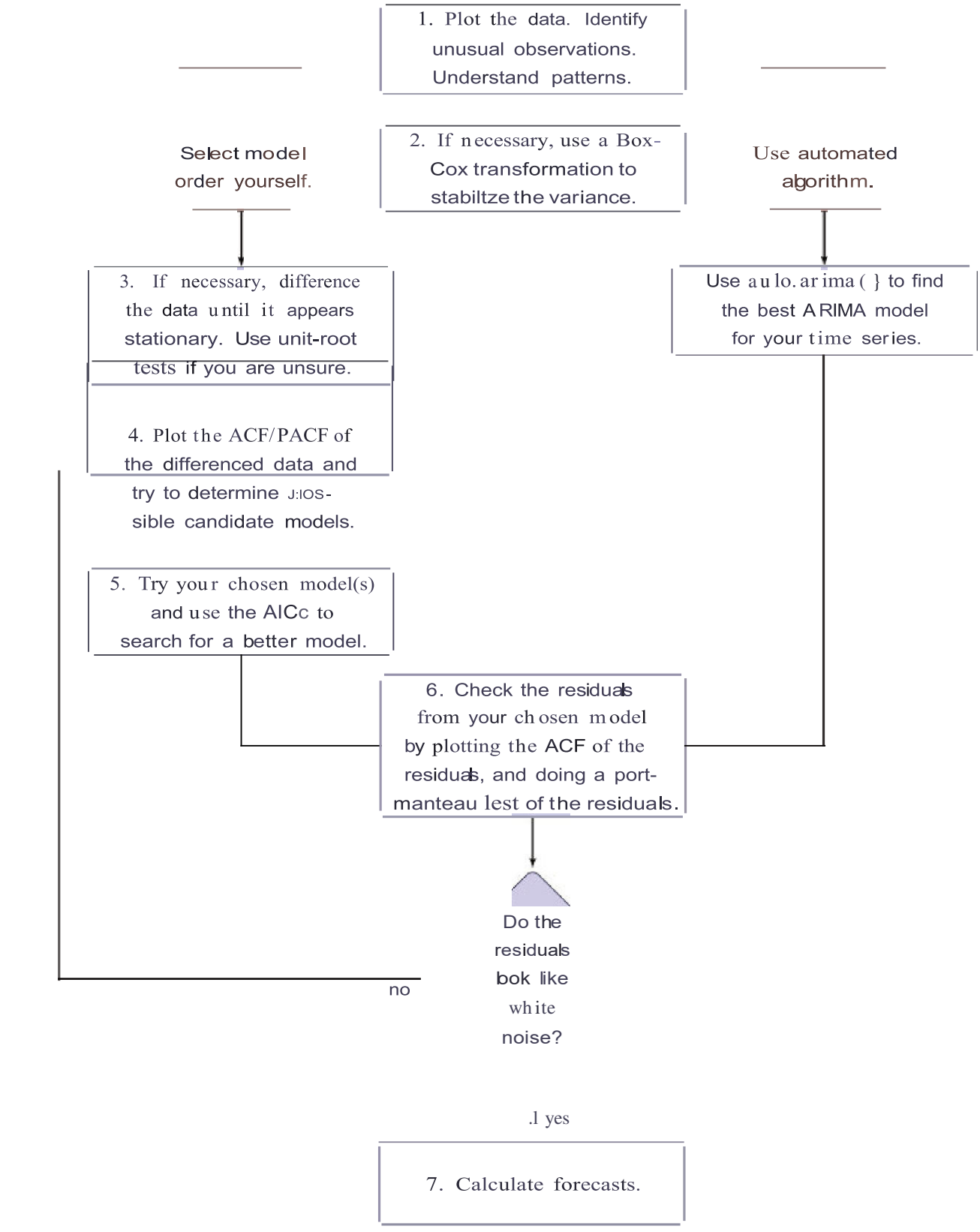
- Predicting stock prices
- Forecast weather
- Forecast the sales of a product

ARIMA model

ARIMA stands for AutoRegressive Integrated Moving Average and is specified by three order parameters: *(p, d, q).*

- **AR(p) Autoregression:** A regression model that utilizes the dependent relationship between a current observation and observations over a previous period.An auto regressive *(AR(p))* component refers to the use of past values in the regression equation for the time series.
- **I(d) Integration:** Uses differencing of observations (subtracting an observation from observation at the previous time step) in order to make the time series stationary. Differencing involves the subtraction of the current values of a series with its previous values d number of times.

- **MA(q) Moving Average:** A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations. A moving average component depicts the error of the model as a combination of previous error terms. The order *q* represents the number of terms to be included in the model.


1. Plot the data. Identify unusual observations. Understand patterns.
2. If necessary, use a Box-Cox transformation to stabiltze the variance.
Select model order yourself.
Use automated algorithm.
3. If necessary, difference the data until it appears stationary. Use unit-root tests if you are unsure.
Use au lo. ar ima ( ) to find the best ARIMA model for your time series.
4. Plot the ACF/PACF of the differenced data and try to determine J:IOS-sible candidate models.
5. Try your chosen model(s) and use the AICc to search for a better model.
6. Check the residuals from your chosen model by plotting the ACF of the residuals, and doing a port-manteau lest of the residuals.
Do the residuals bok like white noise?
no
.l yes
7. Calculate forecasts.

1. Load the data set after installing the package forecast.

2. The Steps of Pre-processing are done, which creates a separate time-series or timestamp.

3. Making Time-series stationary and check the required transformations.

4. The difference value 'd' will be performed.

5. The core important step in ARIMA is plotting ACF and PACF.

6. Determine the two parameters p and q from the plots.

7. The previously created value fits the Aroma model and predicts the future values. The Fitting Process is also named as Box-Jenkins Method.

8. Doing Validation.

auto. Arima() function is used for automatic prediction and ARIMA Models. This function uses unit root tests, minimization of the AIC and MLE to obtain an ARIMA model. To make the series stationary, we need to differentiate a previous value from the current value.

d= pval-cval, if the value is already stationary the d=O.

predict() – Used to predict the model based on the results of the various fitting model used.

Implementation of ARIMA model in R

In R programming, arima() function is used to perform this technique. ARIMA model is used to fit a univariate data. auto.arima() function returns the best ARIMA model by searching over many models.

Syntax:
auto.arima(x)

Parameters:
x: represents univariate time series object

To know about more optional parameters, use below command m the console: help("auto.arima'J

Example 1:
In this example, let's predict the next 10 sale values by using BJsales dataset present in R packages. This dataset is already a time series object, so there is no need to apply ts() function.

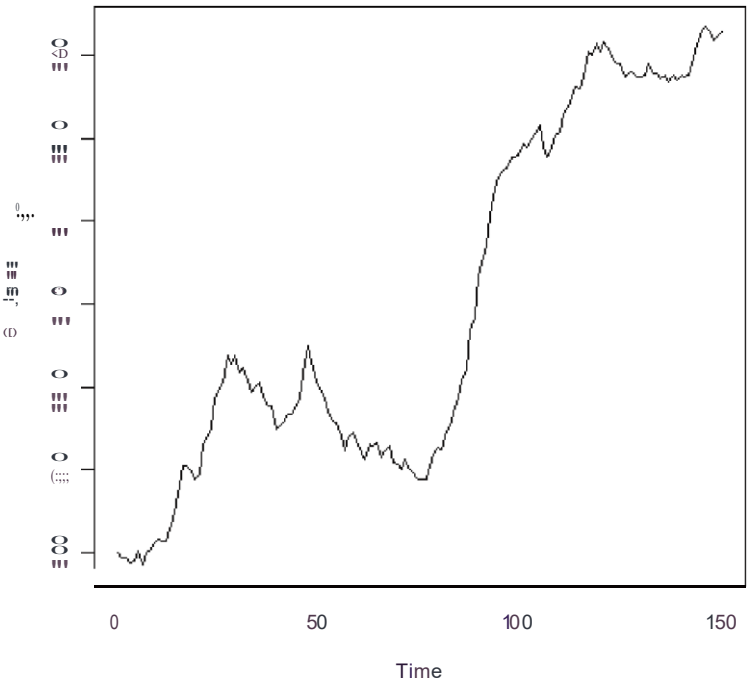
```
# R program to illustrate Time Series Analysis Using ARIMA model in R

# Install the library for forecast()
install.packages("forecast")


# library required for forecasting
library(forecast)

# Output to be created as png file
png(file = "TimeSeriesGFG.png")

# Plotting graph without forecasting
plot(BJsales, main = "Graph without forecasting", col.main =
    "darkgreen")

# Saving the file dev.off()

# Output to be created as png file
png(file = "TimeSeriesARIMAGFG.png")

# Fitting model using arima model fit
<- auto.arima(BJsales)
```

```
# Next 10 forecasted values
forecastedValues <- forecast(fit, 10)

# Print forecasted values
print(forecastedValues)

plot(forecastedValues, main = "Graph with forecasting",
    col.main = "darkgreen")

# saving the file
dev.off()

pre<-predict(fit)
print(pre)
```

Output:

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
151	262.8620	261.1427	264.5814	260.2325	265.4915
152	263.0046	260.2677	265.7415	258.8189	267.1903
153	263.1301	259.4297	266.8304	257.4709	268.7893
154	263.2405	258.5953	267.8857	256.1363	270.3447
155	263.3377	257.7600	268.9153	254.8074	271.8680
156	263.4232	256.9253	269.9211	253.4855	273.3608
157	263.4984	256.0941	270.9028	252.1744	274.8224
158	263.5647	255.2691	271.8602	250.8778	276.2516
159	263.6229	254.4529	272.7930	249.5986	277.6473
160	263.6742	253.6474	273.7011	248.3395	279.0089

Graph without forecasting



Explanation:

output is produced by executing the above code. 10 next values are predicted by using **forecast()** function based on ARIMA model of BJsales dataset. First graph shows the visuals of **BJ**sales without forecasting and the second graph shows the visuals of **BJ**sales with forecasted values.

Conclusion:-

1. The difference between the actual value of the time series and the forecasted value called-

   Forecast Error

2. Use of ARIMA ()-

   This function in R is used for fitting Autoregressive Integrated Moving Average (ARIMA) models, a popular statistical method for forecasting time series data. ARIMA models capture trends, seasonality, and random variations within the data.

3. Use of forecast()-

   The forecast() function is used to generate forecasts for future time periods.