VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Essentials of Data Analytics - (CSE3506)

Faculty – Lakshmi Pathi Jakkamputi **Sir**

Lab-2

Harshanth k Prakash

19BCE1293

L21-22 **Slot**

Tasks for Week-2: Time-series Forecasting

Understand time-series operations/functions and forecast the annual gdp growth rate of India based on given instructions.

AIM

To understands time-series operations/functions and forecast the annual GDP growth rate of India and also GOLD rate data based on given instructions.

Algorithm

1. Start
2. Read the dataset from host pc.
3. Import forecast and tseries libraries.
4. Set the starting, end and frequency with the data columns.
5. Perform acf,pacf and adf tests.
6. Forecast the model after arima.
7. Find the accuracy of the model.
8. Stop.

Result

Case 1: Gold data model.

```
> goldf=forecast(goldmodel,level=c(95),h=24)
> goldf
   Point Forecast    Lo 95    Hi 95
37       5081.371 4767.741 5395.001
38       5138.743 4695.203 5582.283
39       5196.114 4652.891 5739.338
40       5253.486 4626.226 5880.746
41       5310.857 4609.559 6012.155
42       5368.229 4599.995 6136.462
43       5425.600 4595.813 6255.387
44       5482.971 4595.892 6370.051
45       5540.343 4599.453 6481.233
46       5597.714 4605.929 6589.500
47       5655.086 4614.892 6695.279
48       5712.457 4626.011 6798.904
49       5769.829 4639.019 6900.638
50       5827.200 4653.704 7000.696
51       5884.571 4669.887 7099.255
52       5941.943 4687.423 7196.463
53       5999.314 4706.184 7292.444
54       6056.686 4726.066 7387.305
55       6114.057 4746.975 7481.139
56       6171.429 4768.832 7574.025
57       6228.800 4791.566 7666.034
58       6286.171 4815.116 7757.227
59       6343.543 4839.426 7847.660
60       6400.914 4864.447 7937.382
```

```
> accuracy(goldmodel)
                     ME     RMSE      MAE         MPE     MAPE      MASE
Training set 0.08218409 155.5098 116.6965 -0.1799051 2.960037 0.9286895
                    ACF1
Training set -0.07882193
```

```
> adf.test(goldts)

        Augmented Dickey-Fuller Test

data:  goldts
Dickey-Fuller = -2.3526, Lag order = 3, p-value = 0.4359
alternative hypothesis: stationary
```

```
> goldmodel = auto.arima(goldts,ic="aic",trace=TRUE)

 ARIMA(2,1,2) with drift         : Inf
 ARIMA(0,1,0) with drift         : 457.5809
 ARIMA(1,1,0) with drift         : 459.3633
 ARIMA(0,1,1) with drift         : 459.385
 ARIMA(0,1,0)                    : 459.9305
 ARIMA(1,1,1) with drift         : 461.3121

 Best model: ARIMA(0,1,0) with drift
```

Case 2: GDP rate model

```
> gdpf=forecast(gdpmodel,level=c(95),h=11)
> gdpf
     Point Forecast     Lo 95    Hi 95
2021       5.177274 -1.376684 11.73123
2022       5.177274 -1.401989 11.75654
2023       5.177274 -1.427197 11.78174
2024       5.177274 -1.452309 11.80686
2025       5.177274 -1.477327 11.83187
2026       5.177274 -1.502250 11.85680
2027       5.177274 -1.527082 11.88163
2028       5.177274 -1.551821 11.90637
2029       5.177274 -1.576470 11.93102
2030       5.177274 -1.601029 11.95558
2031       5.177274 -1.625500 11.98005
```

```
> accuracy(gdpf)
                     ME     RMSE      MAE      MPE     MAPE      MASE         ACF1
Training set 0.2704179 3.287709 2.345416 121.6616 161.0542 0.7720211 -0.02667223
>
```

```
        Augmented Dickey-Fuller Test

data:  gdps
Dickey-Fuller = -4.7448, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

```
> gdpmodel = auto.arima(gdps,ic="aic",trace=TRUE)

 ARIMA(2,1,2) with drift         : Inf
 ARIMA(0,1,0) with drift         : 341.4397
 ARIMA(1,1,0) with drift         : 332.4653
 ARIMA(0,1,1) with drift         : Inf
 ARIMA(0,1,0)                    : 339.554
 ARIMA(2,1,0) with drift         : 326.0715
 ARIMA(3,1,0) with drift         : 327.9755
 ARIMA(2,1,1) with drift         : Inf
 ARIMA(1,1,1) with drift         : Inf
 ARIMA(3,1,1) with drift         : Inf
 ARIMA(2,1,0)                    : 324.2097
 ARIMA(1,1,0)                    : 330.5929
 ARIMA(3,1,0)                    : 326.1139
 ARIMA(2,1,1)                    : 317.8228
 ARIMA(1,1,1)                    : 316.651
 ARIMA(0,1,1)                    : 314.6516
 ARIMA(0,1,2)                    : 316.6508
 ARIMA(1,1,2)                    : 316.6275

 Best model: ARIMA(0,1,1)
```

Inference

Case 1: Gold data model.

The expected values for the **coming 24 months** are printed using time-series operations/functions and forecast command. According to the dickey-fuller test, p value was found to be 0.4359. Since it is >0.05 conclude that the given values are not stationary. To convert non-stationary stationary to stationary we need to differentiate compares all the models and chooses the best one. Best model is found to be (0,1,0).

Case 2: GDP rate model

The expected values for the **coming 10 years** are printed using time-series operations/functions and forecast command. According to the dickey-fuller test the p value was found to be 0.01. Since it is <0.05 we can conclude that the given values are stationary. AIC compares all the models and choose the best Arima model. It found (0,1,1) is best one and the model is accepted.

Program

Case 1: Gold data model.

```
setwd('D:/6th Sem Works/A2- EDA/LAB/Lab2')
gold <- read.csv("gold.csv")
library(forecast)
library(tseries)
View(gold)
goldts <- ts(gold$Price,start = min(gold$Month) ,end = max(gold$Month), frequency = 1)
class(goldts)
plot(goldts)
acf(goldts)
pacf(goldts)
adf.test(goldts)
goldmodel = auto.arima(goldts,ic="aic",trace=TRUE)
goldf=forecast(goldmodel,level=c(95),h=24)
goldf
accuracy(goldmodel)
```

Case 2: GDP rate model

```
setwd('D:/6th Sem Works/A2- EDA/LAB/Lab2')
gdp <- read.csv("gdp.csv")
library(forecast)
library(tseries)
View(gdp)
gdps <- ts(gdp$GDP_gr,start = min(gdp$Year) ,end = max(gdp$Year), frequency = 1)
class(gdps)
plot(gdps)
acf(gdps)
pacf(gdps)
adf.test(gdps)
gdpmodel = auto.arima(gdps,ic="aic",trace=TRUE)
gdpf=forecast(gdpmodel,level=c(95),h=11)
gdpf
accuracy(gdpf)
```