

pemodelan

Kamilah Nurul Azizah

2023-10-05

Library

```
library(ggplot2)
library(tsibble)
library(tseries)
library(MASS)
library(dplyr)
library(TTR)
library(forecast)
library(lmtest)
library(orcutt)
library(HoRM)
library(dLagM)
library(dynlm)
library(MLmetrics)
library(car)
library(TSA)
library(aTSA)
library(graphics)
```

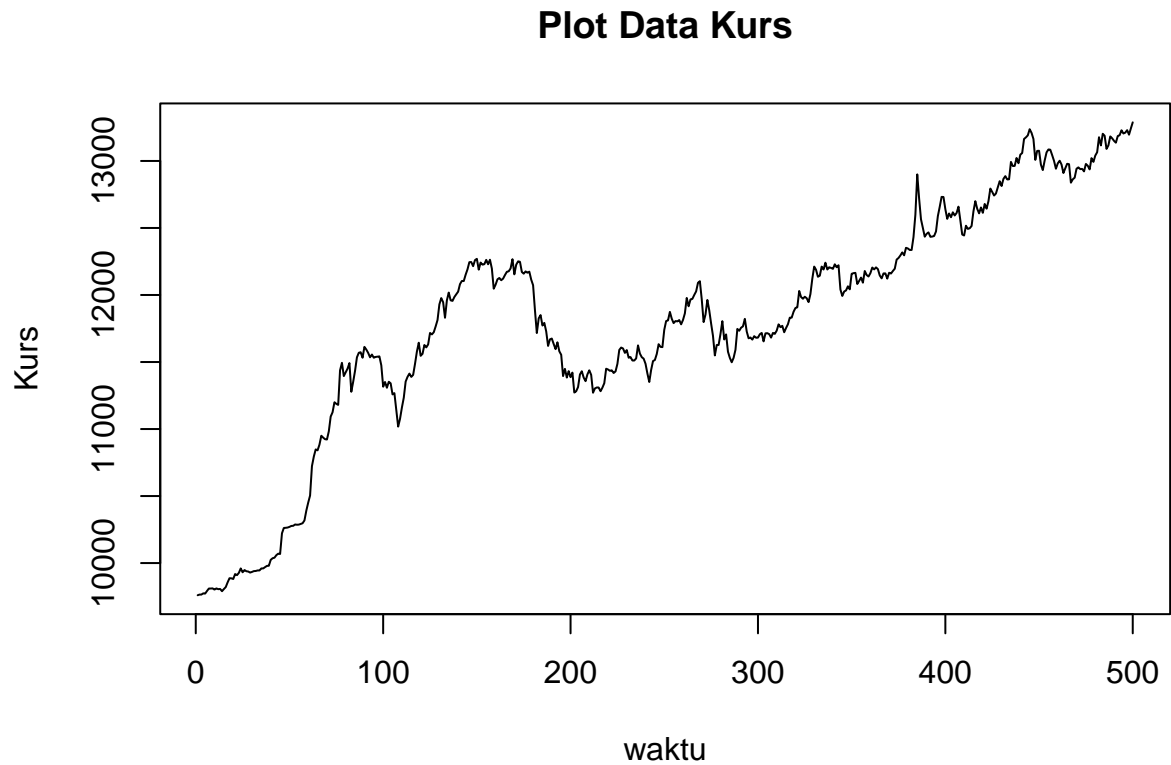
Data Asli

Digunakan data kurs yang dalam hal ini hanya digunakan data 500 periode awal

```
library(rio)
datakurs<-import("https://raw.githubusercontent.com/rizkynurhambali/Praktikum-MPDW-2324/main/Pertemuan%201-5/datakurs.csv")
datakurs <- datakurs[1:500,]
datakurs.ts<-ts(datakurs)
```

Eksplorasi Data

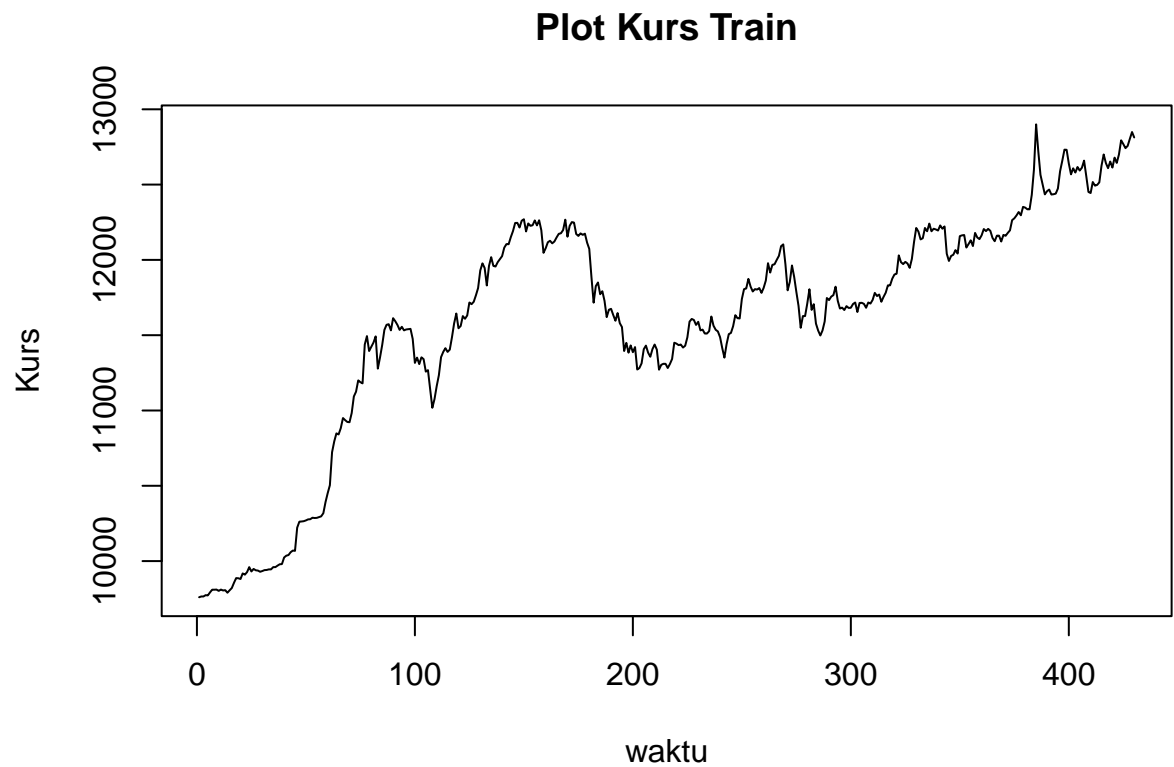
```
plot.ts(datakurs.ts, lty=1, xlab="waktu", ylab="Kurs", main="Plot Data Kurs")
```



Plot Data Penuh

Berdasarkan plot data deret waktu, terlihat bahwa data cenderung memiliki trend yang naik. Berdasarkan pola data, pembagian data latih dan data uji ditetapkan dengan proporsi 86%:14%.

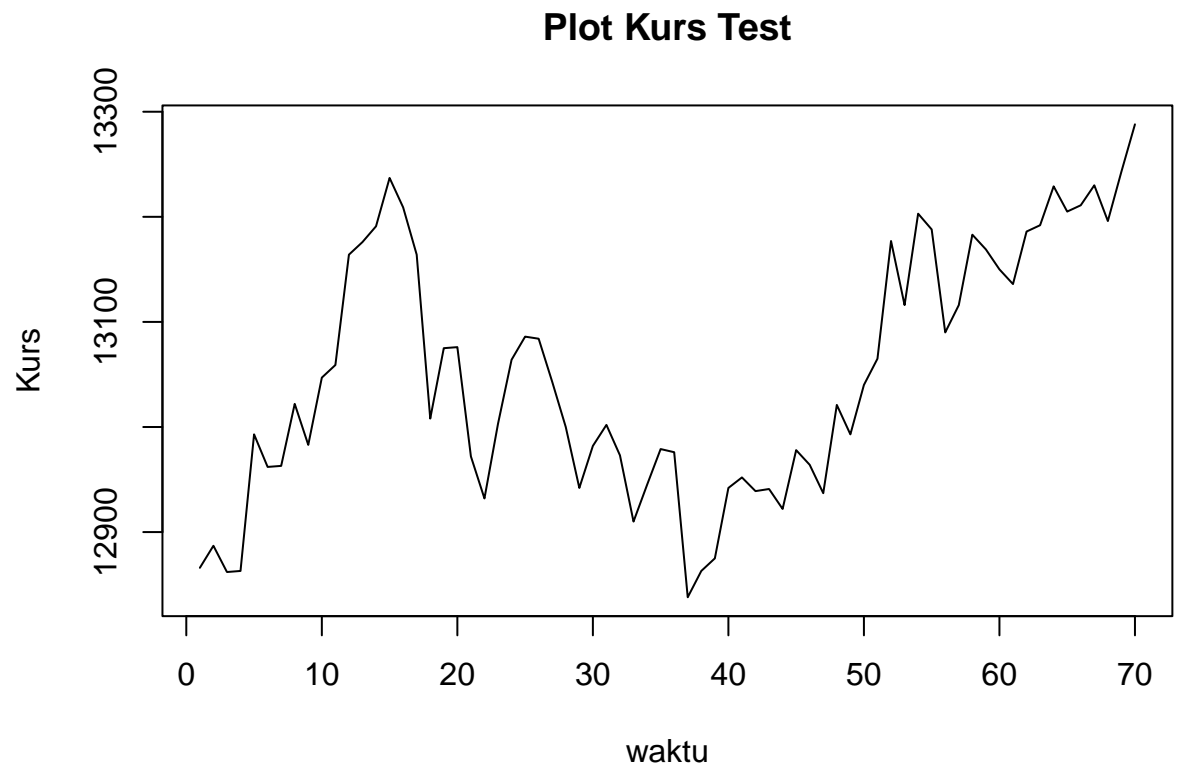
```
kurstrain<-datakurs[1:430]  
train.ts<-ts(kurstrain)  
plot.ts(train.ts, lty=1, xlab="waktu", ylab="Kurs", main="Plot Kurs Train")
```



Plot Data Latih

Berdasarkan plot data deret waktu pada data latih, terlihat bahwa data cenderung memiliki trend yang naik dan cenderung tidak bergerak pada nilai tengah tertentu. Hal ini mengindikasikan bahwa data tidak stasioner dalam rata-rata.

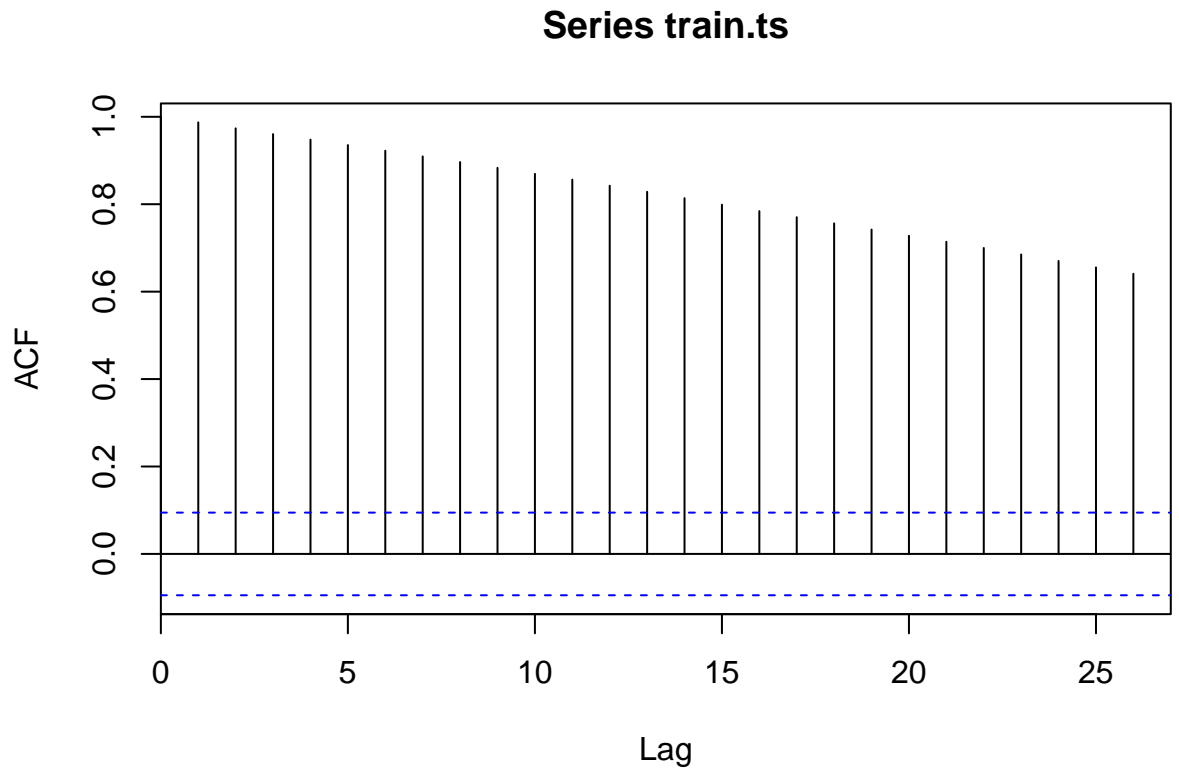
```
kurstest<-datakurs[431:500]  
test.ts<-ts(kurstest)  
plot.ts(test.ts, lty=1, xlab="waktu", ylab="Kurs", main="Plot Kurs Test")
```



Plot Data Uji

Uji Stasioneritas Data

```
acf(train.ts)
```



Plot ACF

Berdasarkan plot ACF, terlihat bahwa plot ACF data menurun secara perlahan (*tails of slowly*). Hal ini juga menjadi indikasi bahwa data tidak stasioner dalam rata-rata.

```
tseries::adf.test(train.ts)
```

Uji ADF

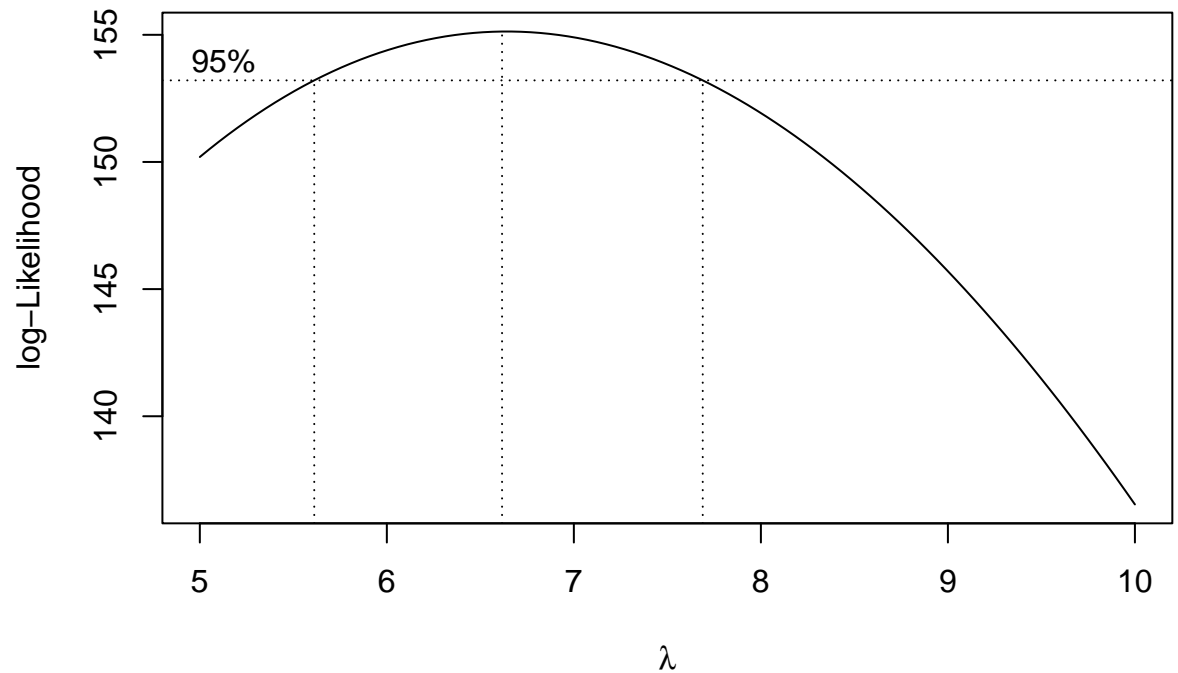
```
##
## Augmented Dickey-Fuller Test
##
## data: train.ts
## Dickey-Fuller = -2.0526, Lag order = 7, p-value = 0.5553
## alternative hypothesis: stationary
```

H_0 : Data tidak stasioner dalam rata-rata

H_1 : Data stasioner dalam rata-rata

Berdasarkan uji ADF tersebut, didapat *p-value* sebesar 0.5553 yang lebih besar dari taraf nyata 5% sehingga tak tolak H_0 dan menandakan bahwa data tidak stasioner dalam rata-rata. Hal ini sesuai dengan hasil eksplorasi menggunakan plot time series dan plot ACF, sehingga ketidakstasioneran model kedepannya harus ditangani.

```
index <- seq(1:430)
bc = boxcox(train.ts~index, lambda = seq(5,10,by=1))
```



Plot Box-Cox

```
#Nilai Rounded Lambda
lambda <- bc$x[which.max(bc$y)]
lambda
```

```
## [1] 6.616162
```

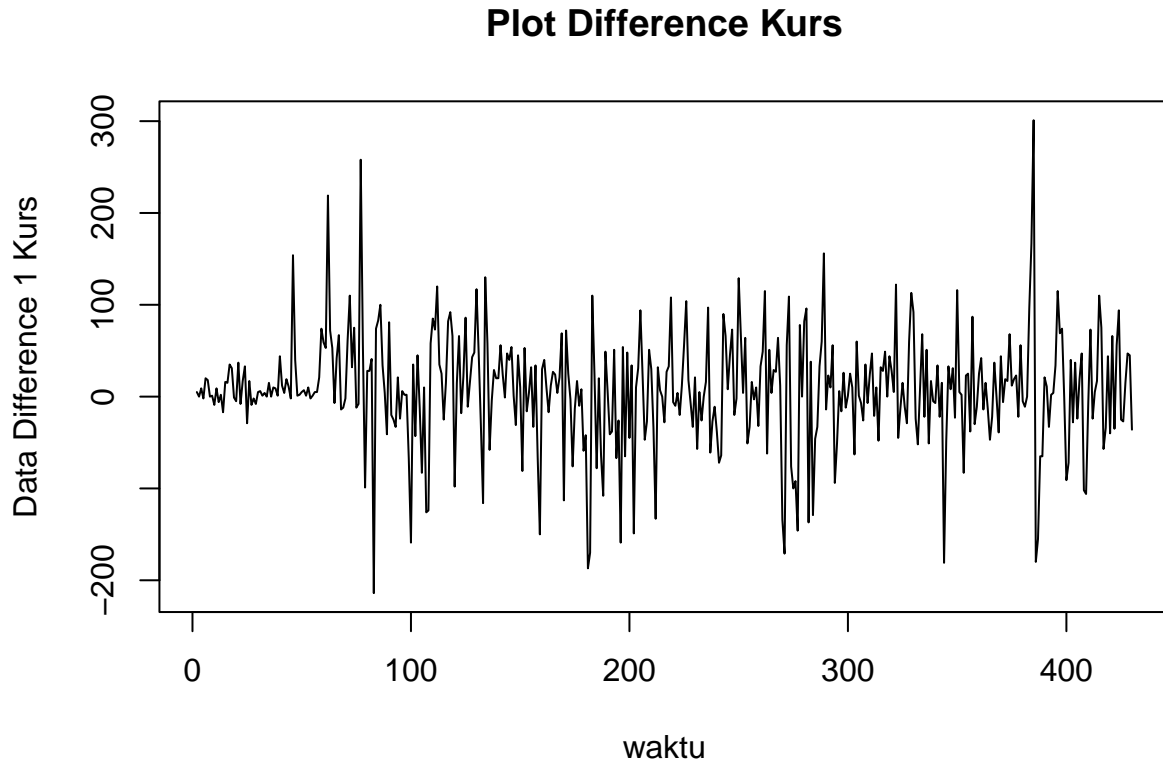
```
#SK
bc$x[bc$y > max(bc$y) - 1/2 * qchisq(.95,1)]
```

```
## [1] 5.656566 5.707071 5.757576 5.808081 5.858586 5.909091 5.959596 6.010101
## [9] 6.060606 6.111111 6.161616 6.212121 6.262626 6.313131 6.363636 6.414141
## [17] 6.464646 6.515152 6.565657 6.616162 6.666667 6.717172 6.767677 6.818182
## [25] 6.868687 6.919192 6.969697 7.020202 7.070707 7.121212 7.171717 7.222222
## [33] 7.272727 7.323232 7.373737 7.424242 7.474747 7.525253 7.575758 7.626263
## [41] 7.676768
```

Plot Boxcox menunjukkan nilai *rounded value* (λ) optimum sebesar **6,64** dan pada selang kepercayaan 95% nilai memiliki batas bawah **0,48** dan batas atas **5,27**. Selang tersebut memuat nilai satu sehingga dapat dikatakan bahwa data bangkitan stasioner dalam ragam.

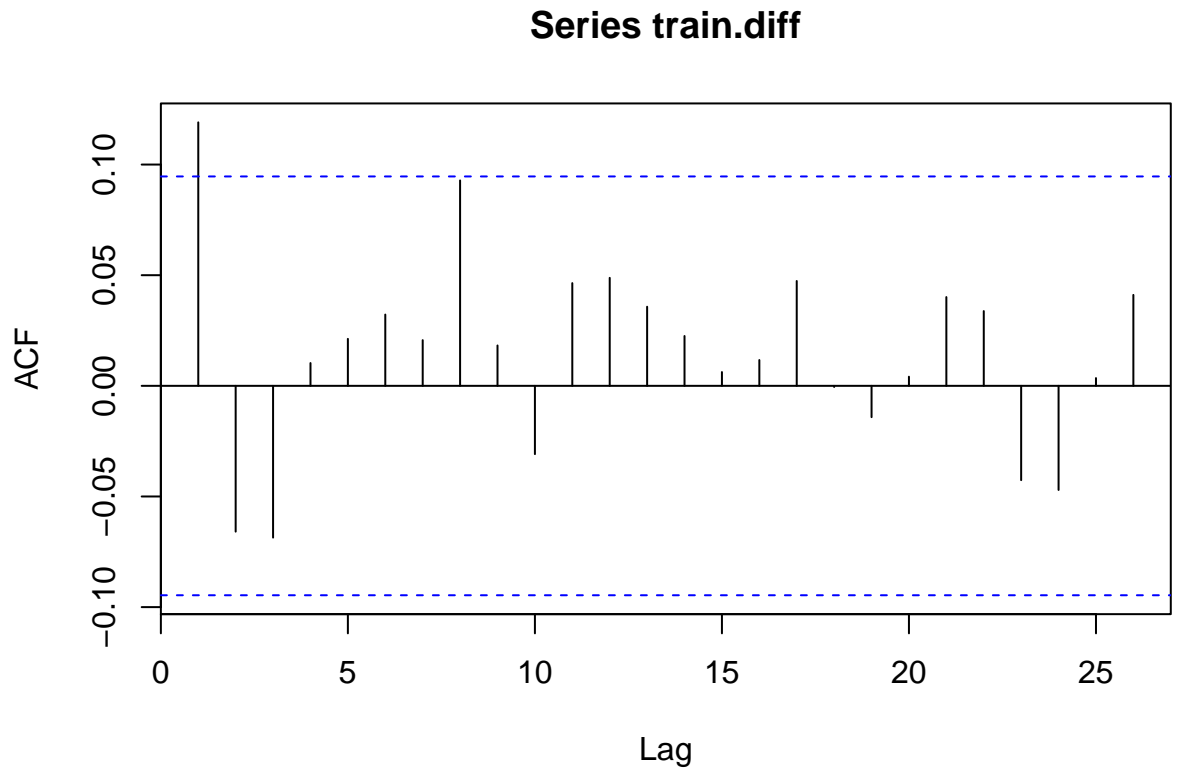
Penanganan Ketidakstasioneran Data

```
train.diff<-diff(train.ts,differences = 1)
plot.ts(train.diff, lty=1, xlab="waktu", ylab="Data Difference 1 Kurs", main="Plot Difference Kurs")
```



Berdasarkan plot data deret waktu, terlihat bahwa data sudah stasioner dalam rata-rata ditandai dengan data bergerak pada nilai tengah tertentu (tidak terdapat trend ataupun musiman pada data)

```
acf(train.diff)
```



Plot ACF

Berdasarkan plot tersebut, terlihat bahwa plot ACF cuts off pada lag ke 1. Hal ini menandakan data sudah stasioner dalam rata-rata dan ketidakstasioneran data telah berhasil tertangani.

```
tseries::adf.test(train.diff)
```

Uji ADF

```
## Warning in tseries::adf.test(train.diff): p-value smaller than printed p-value
```

```
##
## Augmented Dickey-Fuller Test
##
## data: train.diff
## Dickey-Fuller = -6.3673, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

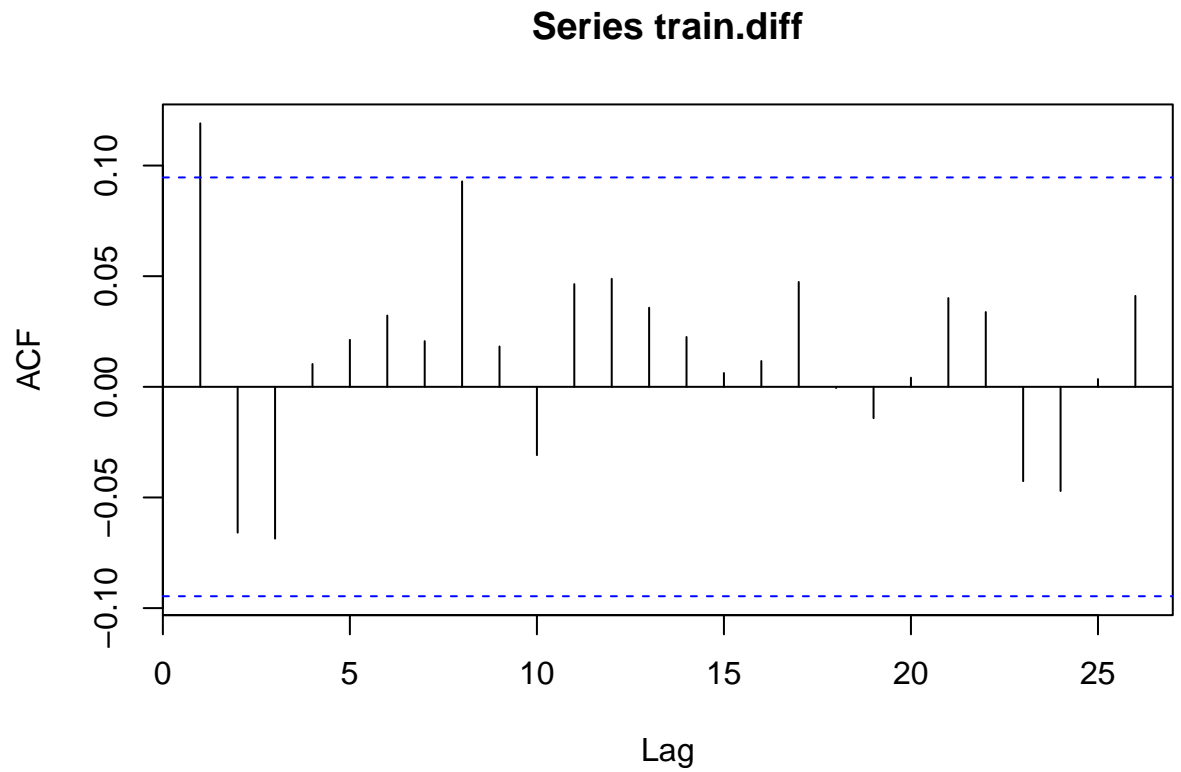
H_0 : Data tidak stasioner dalam rata-rata

H_1 : Data stasioner dalam rata-rata

Berdasarkan uji ADF tersebut, didapat *p-value* sebesar 0.01 yang lebih kecil dari taraf nyata 5% sehingga tolak H_0 atau data stasioner dalam rata-rata. Hal ini sesuai dengan hasil eksplorasi menggunakan plot time series dan plot ACF, sehingga dalam hal ini ketidakstasioneran data sudah berhasil ditangani dan dapat dilanjutkan ke pemodelan.

Identifikasi Model

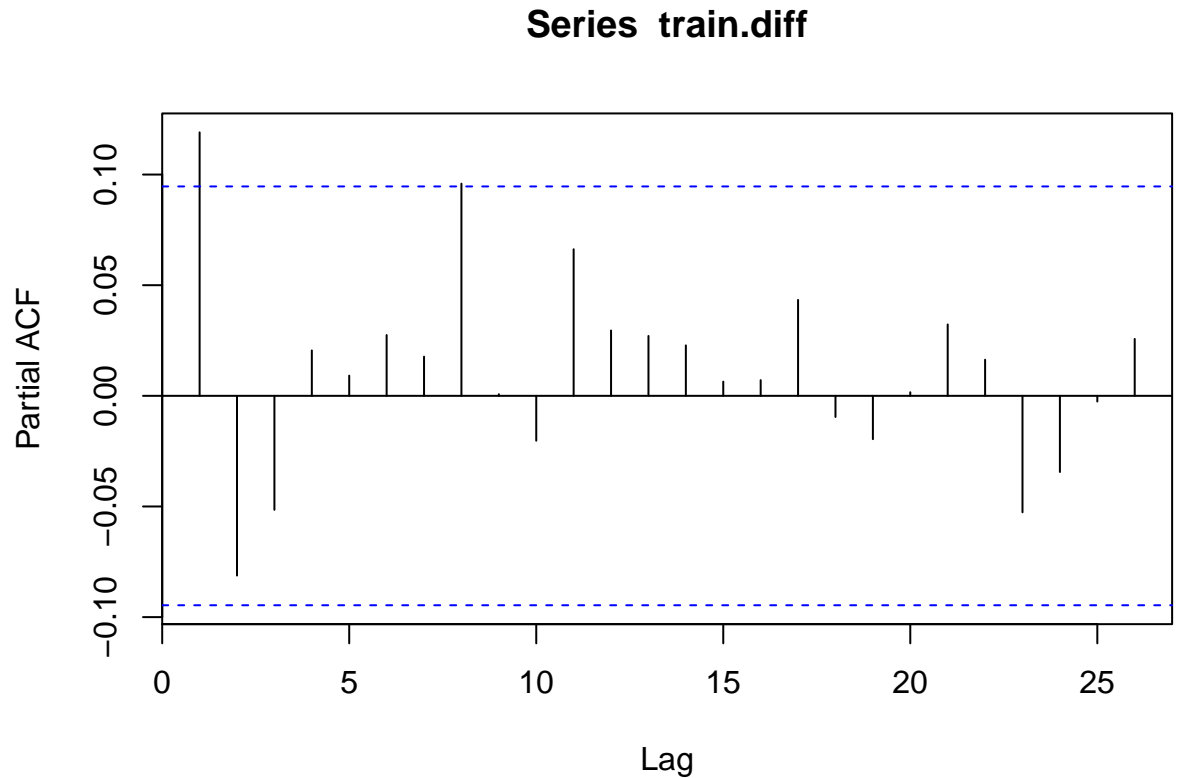
```
acf(train.diff)
```



Plot ACF

Berdasarkan plot tersebut, terlihat bahwa plot ACF cenderung *cuts off* pada lag ke 1, sehingga jika plot PACF dianggap *tails of*, maka model tentatifnya adalah ARIMA(0,1,1).

```
pacf(train.diff)
```



Plot PACF

Berdasarkan plot tersebut, terlihat bahwa plot PACF cenderung *cuts off* pada lag ke 1, sehingga jika plot ACF dianggap *tails of*, maka model tentatifnya adalah ARIMA(1,1,0).

Jika baik plot ACF maupun plot PACF keduanya dianggap tails of, maka model yang terbentuk adalah ARIMA(1,1,1)

```
eacf(train.diff)
```

Plot EACF

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o o o o o o o o o o o o o
## 1 x x o o o o o o o o o o o
## 2 x x o o o o o o o o o o o
## 3 x x o o o o o o o o o o o
## 4 x x x o o o o o o o o o o
## 5 x x o o o o o o o o o o o
## 6 x x o o x x o o o o o o o
## 7 x x o x x x x o o o o o o
```

Identifikasi model menggunakan plot EACF dilakukan dengan melihat ujung segitiga pada pola segitiga nol. Dalam hal ini model tentatif yang terbentuk adalah ARIMA(0,1,2), ARIMA(1,1,2), ARIMA(2,1,2), dan ARIMA(3,1,2).

Pendugaan Parameter Model Tentatif

```
model1.da=Arima(train.diff, order=c(0,1,1),method="ML")
summary(model1.da) #AIC=4753.18
```

ARIMA(0,1,1)

```
## Series: train.diff
## ARIMA(0,1,1)
##
## Coefficients:
##          ma1
##        -0.9879
## s.e.    0.0148
##
## sigma^2 = 3835: log likelihood = -2374.59
## AIC=4753.18  AICc=4753.21  BIC=4761.3
##
## Training set error measures:
##              ME      RMSE      MAE MPE MAPE      MASE      ACF1
## Training set -0.6762224 61.78387 43.22489 NaN  Inf 0.7363284 0.112352
```

```
lmtest::coeftest(model1.da) #seluruh parameter signifikan
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ma1 -0.987941    0.014806 -66.727 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model2.da=Arima(train.diff, order=c(1,1,0),method="ML")
summary(model2.da) #AIC=4917.41
```

ARIMA(1,1,0)

```
## Series: train.diff
## ARIMA(1,1,0)
##
## Coefficients:
##          ar1
##        -0.3940
## s.e.    0.0444
##
## sigma^2 = 5676: log likelihood = -2456.7
## AIC=4917.41  AICc=4917.43  BIC=4925.52
```

```
##
## Training set error measures:
##           ME      RMSE      MAE MPE MAPE      MASE      ACF1
## Training set -0.05787648 75.16292 54.08917 NaN  Inf 0.9213996 -0.121856
```

```
lmtest::coeftest(model2.da) #seluruh parameter signifikan
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1 -0.394037    0.044414 -8.8719 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model3.da=Arima(train.diff, order=c(1,1,1),method="ML")
summary(model3.da) #AIC=4761.39
```

ARIMA(1,1,1)

```
## Series: train.diff
## ARIMA(1,1,1)
##
## Coefficients:
##          ar1      ma1
##      0.1207 -0.9966
## s.e. 0.0491 0.0210
##
## sigma^2 = 3782: log likelihood = -2371.6
## AIC=4749.21 AICc=4749.27 BIC=4761.39
##
## Training set error measures:
##           ME      RMSE      MAE MPE MAPE      MASE      ACF1
## Training set -2.082868 61.28528 43.2876 -Inf  Inf 0.7373967 0.006544589
```

```
lmtest::coeftest(model3.da) #seluruh parameter signifikan
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1 0.120655    0.049064  2.4591 0.01393 *
## ma1 -0.996606    0.020970 -47.5253 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model4.da=Arima(train.diff, order=c(0,1,2),method="ML")
summary(model4.da) #AIC=4748.3
```

ARIMA(0,1,2)

```
## Series: train.diff
## ARIMA(0,1,2)
##
## Coefficients:
##          ma1      ma2
##      -0.8600 -0.1359
## s.e.   0.0512  0.0507
##
## sigma^2 = 3775: log likelihood = -2371.15
## AIC=4748.3  AICc=4748.35  BIC=4760.47
##
## Training set error measures:
##              ME      RMSE      MAE  MPE MAPE      MASE      ACF1
## Training set -2.059629 61.22216 43.24997 -Inf  Inf  0.7367557 -0.009277841
```

```
lmtest::coeftest(model4.da) #seluruh parameter signifikan
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ma1 -0.860021  0.051204 -16.796 < 2.2e-16 ***
## ma2 -0.135945  0.050725  -2.680  0.007361 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model5.da=Arima(train.diff, order=c(1,1,2),method="ML")
summary(model5.da) #AIC=4749.85
```

ARIMA(1,1,2)

```
## Series: train.diff
## ARIMA(1,1,2)
##
## Coefficients:
##          ar1      ma1      ma2
##      -0.1723 -0.6912 -0.3021
## s.e.   0.2443  0.2348  0.2305
##
## sigma^2 = 3782: log likelihood = -2370.92
## AIC=4749.85  AICc=4749.94  BIC=4766.08
##
```

```
## Training set error measures:
##           ME      RMSE      MAE  MPE MAPE      MASE      ACF1
## Training set -1.798497 61.21036 43.15664 -Inf  Inf 0.7351659 -0.006549048
```

```
lmtest::coeftest(model5.da) #terdapat parameter tidak signifikan
```

```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1 -0.17233    0.24425 -0.7056 0.480467
## ma1 -0.69120    0.23481 -2.9436 0.003244 **
## ma2 -0.30214    0.23054 -1.3106 0.189995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model6.da=Arima(train.diff, order=c(2,1,2),method="ML")
summary(model6.da) #AIC=4749.52
```

ARIMA(2,1,2)

```
## Series: train.diff
## ARIMA(2,1,2)
##
## Coefficients:
##          ar1      ar2      ma1      ma2
##          0.4635 -0.1308 -1.3295 0.3386
## s.e. 0.3091 0.0565 0.3089 0.3073
##
## sigma^2 = 3777: log likelihood = -2369.76
## AIC=4749.52 AICc=4749.66 BIC=4769.81
##
## Training set error measures:
##           ME      RMSE      MAE  MPE MAPE      MASE      ACF1
## Training set -0.4957896 61.09641 43.1911 NaN  Inf 0.7357528 -0.004047784
```

```
lmtest::coeftest(model6.da) #terdapat parameter tidak signifikan
```

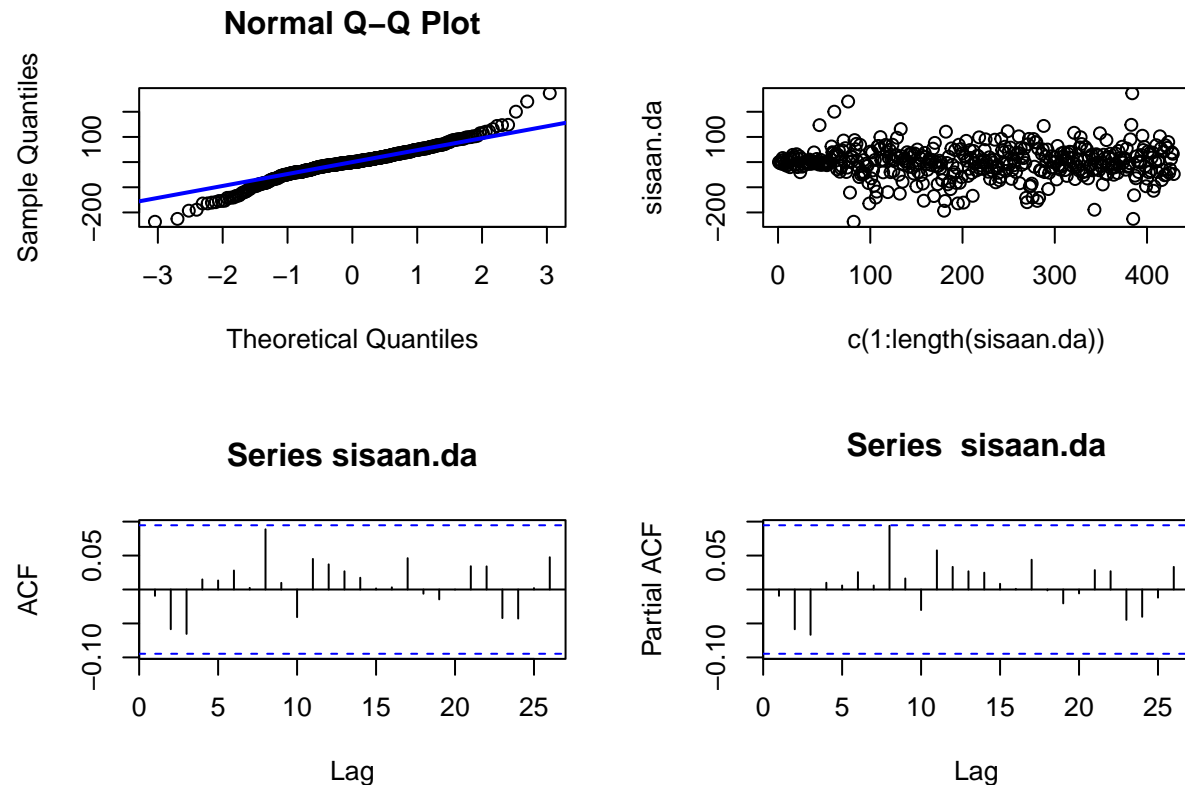
```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1 0.463451    0.309061 1.4995 0.13373
## ar2 -0.130837    0.056523 -2.3148 0.02063 *
## ma1 -1.329455    0.308856 -4.3045 1.674e-05 ***
## ma2 0.338625    0.307281 1.1020 0.27046
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Berdasarkan pendugaan parameter di atas, nilai AIC terkecil dimiliki oleh model ARIMA(0,1,2) dan parameter model ARIMA(0,1,2) juga seluruhnya signifikan sehingga model yang dipilih adalah model ARIMA(0,1,2).

Analisis Sisaan

Model terbaik hasil identifikasi kemudian dicek asumsi sisaannya. Sisaan model ARIMA harus memenuhi asumsi normalitas, kebebasan sisaan, dan kehomogenan ragam. Diagnostik model dilakukan secara eksplorasi dan uji formal.

```
#Eksplorasi
sisaan.da <- model4.da$residuals
par(mfrow=c(2,2))
qqnorm(sisaan.da)
qqline(sisaan.da, col = "blue", lwd = 2)
plot(c(1:length(sisaan.da)),sisaan.da)
acf(sisaan.da)
pacf(sisaan.da)
```



Eksplorasi Sisaan

```
par(mfrow = c(1,1))
```

Berdasarkan plot kuantil-kuantil normal, secara eksplorasi ditunjukkan sisaan tidak menyebar normal ditandai dengan titik titik yang cenderung tidak mengikuti garis 45°. Kemudian dapat dilihat juga lebar pita sisaan yang cenderung tidak sama menandakan bahwa sisaan memiliki ragam yang heterogen. Plot ACF dan PACF sisaan ARIMA(0,0,2) juga tidak signifikan pada 20 lag awal yang menandakan saling bebas. Kondisi ini akan diuji lebih lanjut dengan uji formal.

```
#1) Sisaan Menyebar Normal
ks.test(sisaan.da,"pnorm") #tak tolak H0 > sisaan menyebar normal
```

Uji Formal

```
##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  sisaan.da
## D = 0.46816, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

Selain dengan eksplorasi, asumsi tersebut dapat diuji menggunakan uji formal. Pada tahapan ini uji formal yang digunakan untuk normalitas adalah uji Kolmogorov-Smirnov (KS). Hipotesis pada uji KS adalah sebagai berikut.

H_0 : Sisaan menyebar normal

H_1 : Sisaan tidak menyebar normal

Berdasarkan uji KS tersebut, didapat *p-value* sebesar 0.00 yang kurang dari taraf nyata 5% sehingga tolak H_0 dan menandakan bahwa sisaan tidak menyebar normal. Hal ini sesuai dengan hasil eksplorasi menggunakan plot kuantil-kuantil normal.

```
#2) Sisaan saling bebas/tidak ada autokorelasi
Box.test(sisaan.da, type = "Ljung") #tak tolak H0 > sisaan saling bebas
```

```
##
## Box-Ljung test
##
## data:  sisaan.da
## X-squared = 0.037186, df = 1, p-value = 0.8471
```

Selanjutnya akan dilakukan uji formal untuk kebebasan sisaan menggunakan uji Ljung-Box. Hipotesis yang digunakan adalah sebagai berikut.

H_0 : Sisaan saling bebas

H_1 : Sisaan tidak saling bebas

Berdasarkan uji Ljung-Box tersebut, didapat *p-value* sebesar 0.8471 yang lebih besar dari taraf nyata 5% sehingga tak tolak H_0 dan menandakan bahwa sisaan saling bebas. Hal ini berbeda dengan eksplorasi.

```
#3) Sisaan homogen
Box.test((sisaan.da)^2, type = "Ljung") #tak tolak H0 > sisaan homogen
```

```
##
## Box-Ljung test
##
## data:  (sisaan.da)^2
## X-squared = 29.906, df = 1, p-value = 4.536e-08
```


Hipotesis yang digunakan untuk uji kehomogenan ragam adalah sebagai berikut.

H_0 : Ragam sisaan homogen

H_1 : Ragam sisaan tidak homogen

Berdasarkan uji Ljung-Box terhadap sisaan kuadrat tersebut, didapat *p-value* sebesar 0.000 yang kurang dari taraf nyata 5% sehingga tak tolak H_0 dan menandakan bahwa ragam sisaan tidak homogen.

```
#4) Nilai tengah sisaan sama dengan nol
t.test(sisaan.da, mu = 0, conf.level = 0.95) #tak tolak h0 > nilai tengah sisaan sama dengan 0

##
## One Sample t-test
##
## data:  sisaan.da
## t = -0.69638, df = 428, p-value = 0.4866
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -7.872876  3.753619
## sample estimates:
## mean of x
## -2.059629
```

Terakhir, dengan uji-t, akan dicek apakah nilai tengah sisaan sama dengan nol. Hipotesis yang diujikan sebagai berikut.

H_0 : nilai tengah sisaan sama dengan 0

H_1 : nilai tengah sisaan tidak sama dengan 0

Berdasarkan uji-t tersebut, didapat *p-value* sebesar 0.4866 yang lebih besar dari taraf nyata 5% sehingga tak tolak H_0 dan menandakan bahwa nilai tengah sisaan sama dengan nol. Hal ini berbeda dengan eksplorasi.

Peramalan

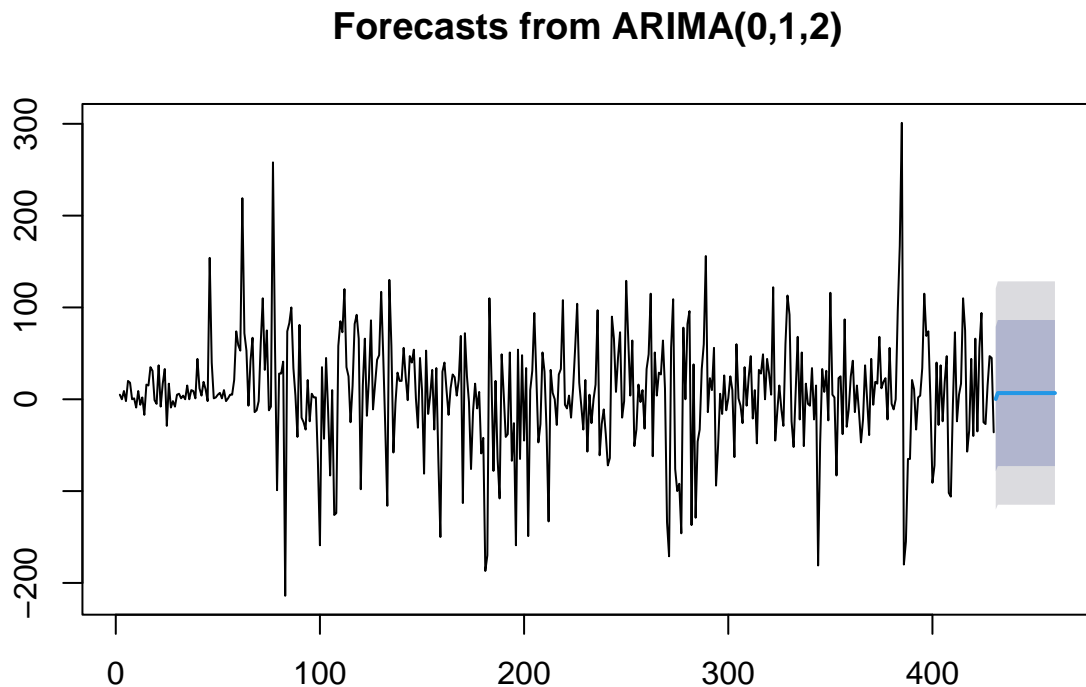
Peramalan dilakukan menggunakan fungsi `forecast()` . Contoh peramalan berikut ini dilakukan untuk 30 hari ke depan.

```
#---FORECAST---#
ramalan.da <- forecast::forecast(model4.da, h = 30)
ramalan.da

##      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
## 431      0.3239709 -78.42499  79.07294 -120.1122  120.7601
## 432      6.7730086 -72.74745  86.29346 -114.8430  128.3891
## 433      6.7730086 -72.74808  86.29410 -114.8440  128.3900
## 434      6.7730086 -72.74872  86.29473 -114.8450  128.3910
## 435      6.7730086 -72.74935  86.29537 -114.8459  128.3920
## 436      6.7730086 -72.74998  86.29600 -114.8469  128.3929
## 437      6.7730086 -72.75062  86.29664 -114.8479  128.3939
## 438      6.7730086 -72.75125  86.29727 -114.8489  128.3949
## 439      6.7730086 -72.75189  86.29790 -114.8498  128.3958
## 440      6.7730086 -72.75252  86.29854 -114.8508  128.3968
## 441      6.7730086 -72.75316  86.29917 -114.8518  128.3978
```

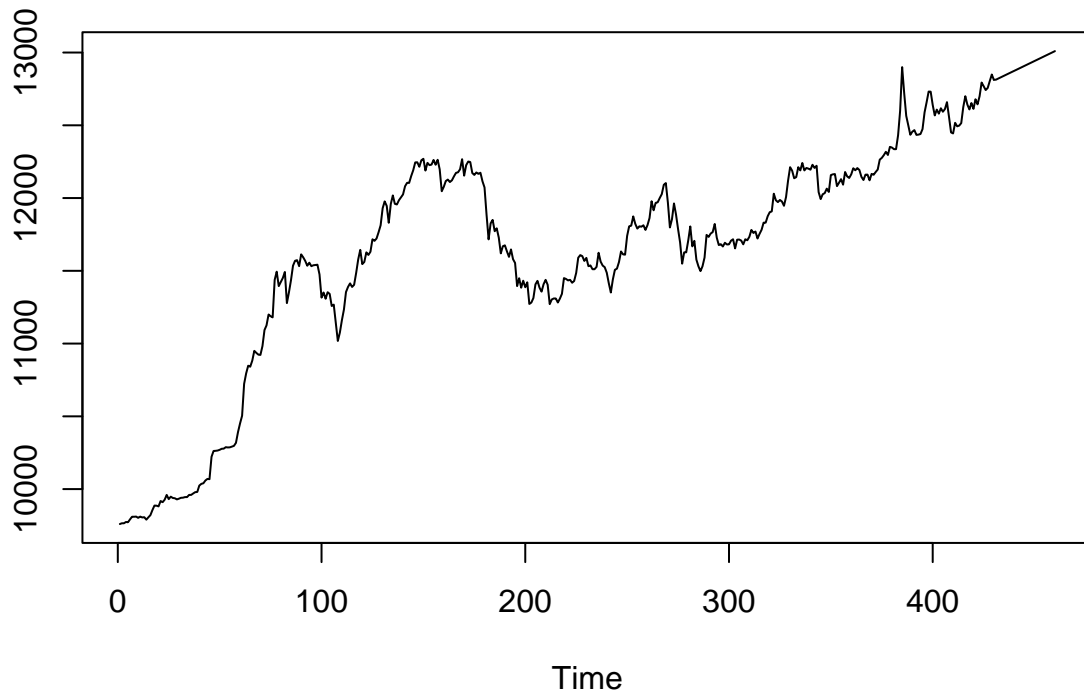
```
## 442      6.7730086 -72.75379 86.29981 -114.8527 128.3988
## 443      6.7730086 -72.75442 86.30044 -114.8537 128.3997
## 444      6.7730086 -72.75506 86.30108 -114.8547 128.4007
## 445      6.7730086 -72.75569 86.30171 -114.8556 128.4017
## 446      6.7730086 -72.75633 86.30234 -114.8566 128.4026
## 447      6.7730086 -72.75696 86.30298 -114.8576 128.4036
## 448      6.7730086 -72.75760 86.30361 -114.8586 128.4046
## 449      6.7730086 -72.75823 86.30425 -114.8595 128.4055
## 450      6.7730086 -72.75886 86.30488 -114.8605 128.4065
## 451      6.7730086 -72.75950 86.30552 -114.8615 128.4075
## 452      6.7730086 -72.76013 86.30615 -114.8624 128.4085
## 453      6.7730086 -72.76077 86.30678 -114.8634 128.4094
## 454      6.7730086 -72.76140 86.30742 -114.8644 128.4104
## 455      6.7730086 -72.76204 86.30805 -114.8653 128.4114
## 456      6.7730086 -72.76267 86.30869 -114.8663 128.4123
## 457      6.7730086 -72.76330 86.30932 -114.8673 128.4133
## 458      6.7730086 -72.76394 86.30996 -114.8683 128.4143
## 459      6.7730086 -72.76457 86.31059 -114.8692 128.4152
## 460      6.7730086 -72.76521 86.31122 -114.8702 128.4162
```

```
data.ramalan.da <- ramalan.da$mean
plot(ramalan.da)
```



Berdasarkan hasil plot ramalan di atas, dapat dilihat bahwa ramalan ARIMA(0,0,2) cenderung stabil hingga akhir periode. Selanjutnya, dapat dicari nilai akurasi antara hasil ramalan dengan data uji sebagai berikut.

```
pt_1 <- train.ts[430] #nilai akhir data latih
hasil.forc.Diff <- data.ramalan.da
hasil <- diffinv(hasil.forc.Diff, differences = 1) + pt_1
#has.1 sama hasilnta dengan: cumsum(c(pt_1,hasil.forc.Diff))
ts.plot(train.ts,hasil)
```



```
perbandingan.da<-matrix(data=c(head(test.ts, n=30), hasil[-1]),
                          nrow = 30, ncol = 2)
colnames(perbandingan.da)<-c("Aktual","Hasil Forecast")
perbandingan.da
```

```
##      Aktual Hasil Forecast
## [1,] 12866      12813.32
## [2,] 12887      12820.10
## [3,] 12862      12826.87
## [4,] 12863      12833.64
## [5,] 12993      12840.42
## [6,] 12962      12847.19
## [7,] 12963      12853.96
## [8,] 13022      12860.74
## [9,] 12983      12867.51
## [10,] 13047      12874.28
## [11,] 13059      12881.05
## [12,] 13164      12887.83
## [13,] 13176      12894.60
```

```
## [14,] 13191      12901.37
## [15,] 13237      12908.15
## [16,] 13209      12914.92
## [17,] 13164      12921.69
## [18,] 13008      12928.47
## [19,] 13075      12935.24
## [20,] 13076      12942.01
## [21,] 12972      12948.78
## [22,] 12932      12955.56
## [23,] 13003      12962.33
## [24,] 13064      12969.10
## [25,] 13086      12975.88
## [26,] 13084      12982.65
## [27,] 13043      12989.42
## [28,] 13000      12996.20
## [29,] 12942      13002.97
## [30,] 12982      13009.74
```

```
accuracy(ts(hasil[-1]), head(test.ts, n=30))
```

```
##           ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
## Test set 118.9674 156.6908 126.4518 0.9079581 0.9657541 0.7929994 2.646555
```