

ANALISIS DATA
CURAH HUJAN KOTA BANDUNG TAHUN 2017-2022



Disusun Oleh:

Luis Ricardo Pandiangan	(2108541050)
Ragil Mona Bianca	(2108541052)
Misel Eriyani Rohi	(2108541068)
Sinsin Triana Q	(2108541077)
Ni Putu Dian Astutik	(2108541083)

PROGRAM STUDI MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS UDAYANA
2023

Tugas ini adalah hasil pekerjaan kelompok kami sendiri dan belum pernah dikerjakan atau digunakan untuk tugas mata kuliah lain atau keperluan lain. Jika terdapat analisis yang sama persis titik-koma (plagiat) maka kami akan mendapat sanksi diberi nilai nol (0).

Nama Anggota I : Luis Ricardo Pandiangan

NIM Anggota I : 2108541050

Tanda tangan Anggota I :

Nama Anggota II : Ragil Mona Bianca

NIM Anggota II : 2108541052

Tanda tangan Anggota II :

Nama Anggota III : Misel Eriyani Rohi

NIM Anggota III : 2108541068

Tanda tangan Anggota III :

Nama Anggota IV : Sinsin Triana Kian Q

NIM Anggota IV : 2108541077

Tanda tangan Anggota IV :

Nama Anggota V : Ni Putu Dian Astutik

NIM Anggota V : 2108541083

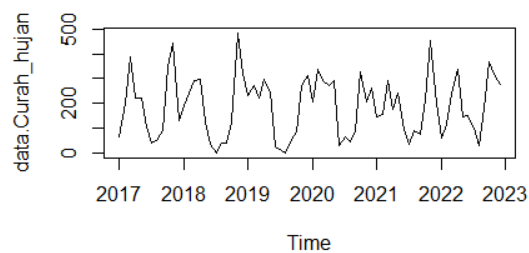
Tanda tangan Anggota V :

ANALISIS DATA CURAH HUJAN KOTA BANDUNG TAHUN 2017-2022

Data yang digunakan adalah data jumlah curah hujan di Kota Bandung, Jawa Barat yang diperoleh dari situs resmi BPS Kota Bandung, Jawa Barat tahun 2017 hingga tahun 2022.

1.1 Plot Data

Langkah pertama adalah memplot data untuk mengetahui ke-stasioneran data dengan visualisasi. Pada output R-Studio diperoleh plot deret waktu dari data curah hujan di Kota Bandung seperti berikut:

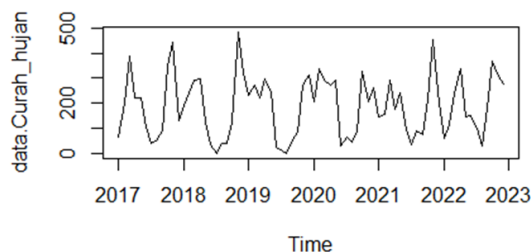


Dapat terlihat dari plot data tersebut tidak adanya tren namun terbentuk pola musiman yang berulang dalam selang waktu 12 periode.

1.2 Spesifikasi Model

a. Uji kestasioneran

- Plot



Dapat terlihat dari plot data tersebut tidak adanya tren namun terbentuk pola musiman yang berulang dalam selang waktu 12 periode. Diperoleh bahwa data stasioner namun masih terdapat komponen musiman.

- Augmented Dickey Fuller Test

Kemudian akan diuji kestasioneran menggunakan Uji Dickey-Fuller (Augmented Dickey-Fuller atau ADF), dengan hipotesis sebagai berikut:

- Hipotesis nol (H_0) : Deret waktu tidak stasioner
- Hipotesis alternatif H_1) : Deret waktu stasioner

Statistik uji dari uji ADF dibandingkan dengan nilai kritis, dengan program R beserta output R-Studio sebagai berikut:

```
Augmented Dickey-Fuller Test
data: data.Curah_hujan
Dickey-Fuller = -4.0338, Lag order = 4,
p-value = 0.01315
alternative hypothesis: stationary
```

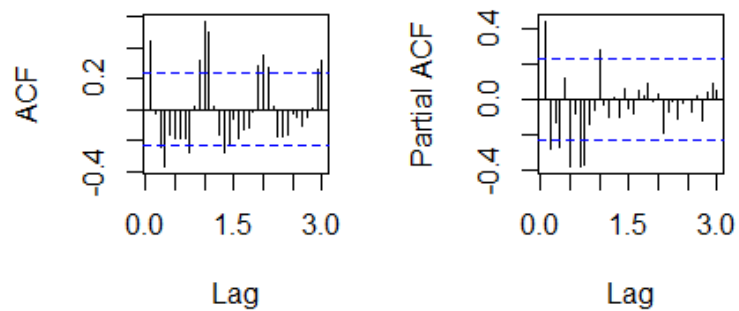
dari hasil uji ADF diperoleh p-value sebesar 0,01315 menggunakan tingkat signifikansi 5% ($\alpha = 0,05$) diperoleh $0,01315 < 0,05$ sehingga hasil keputusan yang diperoleh adalah menolak H_0 artinya data curah hujan Kota Bandung tahun 2017 hingga 2022 merupakan data stasioner.

b. Menentukan orde p, d, q, P, D, dan Q yang sesuai.

1) Model Non musiman

Untuk menentukan orde nonmusiman digunakan data asli karena pada uji formal diperoleh hasil bahwa data stasioner. Jika dilakukan dengan menggunakan ACF dan PACF data curah hujan dengan program R beserta outputnya sebagai berikut:

Series data.Curah_huji; Series data.Curah_huj



Dengan plot ACF dan PACF membentuk grafik eksponensial sehingga mengindikasikan grafik ARMA, sehingga diperlukan metode lain untuk menentukan orde p dan q. Menggunakan metode EACF untuk identifikasi p dan q yang sesuai, diperoleh output R sebagai berikut:

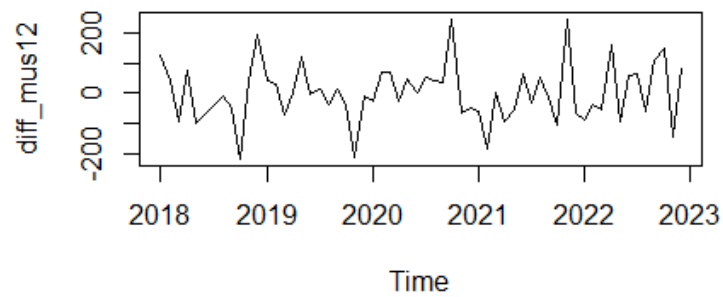
```
eacf(data.Curah_hujan)

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

Sehingga pada komponen nonmusiman menghasilkan kandidat-kandidat model ARMA(0,1), ARMA(1,1) dan ARMA(2,1).

2) Komponen musiman

Selanjutnya pembentukan komponen musiman dari data asli. Diperoleh plot pada pembentukan musiman dari data asli sebagai berikut:

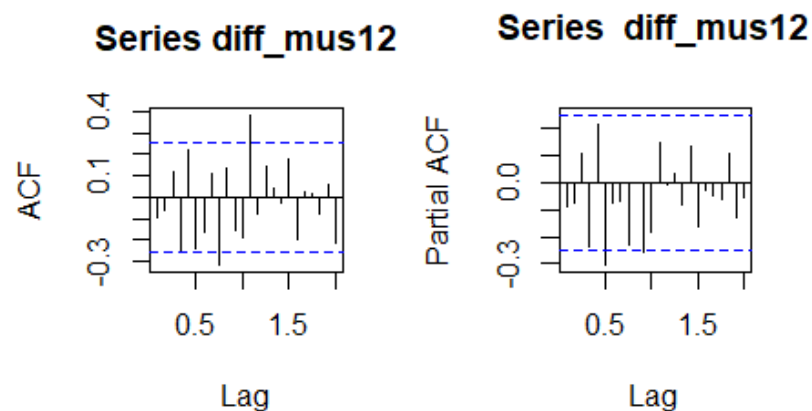


Kemudian diuji formal kestasioneran dengan uji ADF diperoleh bahwa:

Augmented Dickey-Fuller Test

```
data: diff_mus12
Dickey-Fuller = -4.6218, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

Nilai p-value 0,01 yang lebih kecil daripada tingkat signifikan 0.05 sehingga diperoleh bahwa data pembentukan musiman dari data asli merupakan model stasioner. Diperoleh plot ACF dan PACF pada pembentukan musiman dari data asli sebagai berikut:



Dengan plot ACF dengan garis signifikan pada lag 12 dan plot PACF yang meluruh secara perlahan mengindikasikan bahwa model musimanya merupakan model ARMA. Sehingga menggunakan metode EACF untuk identifikasi p dan q yang sesuai, diperoleh output R sebagai berikut;

```
eacf(diff_mus12)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o o o o o o o o x o o o x o
1 x o o o o o o o o o o o x o
2 x o o o o x o o o o o o x o
3 x o x o o o o o o o o x o
4 x o o o o o o o o o o o x o
5 x o o o o o o o o o o o x o
6 o o x o o o o o o o o o o
7 x o x o o o o o o o o o o
```

Sehingga pada komponen musiman menghasilkan kandidat-kandidat model ARIMA(0,0,1)₁₂, ARIMA(1,0,1)₁₂ atau ARIMA(2,0,1)₁₂.

3) Berdasarkan metode Grid Search

Grid Search mencoba berbagai kombinasi nilai p,d,q, P, D dan Q untuk menemukan kombinasi yang memberikan kinerja terbaik berdasarkan suatu kriteria evaluasi, seperti AIC (Akaike Information Criterion) atau BIC (Bayesian Information Criterion).

Dengan rentang nilai parameter sebagai berikut:

```
p_values <- c(0, 1, 2, 3, 4, 5)
d_values <- c(0, 1)
q_values <- c(0, 1, 2, 3, 4, 5,)
P_values <- c(0, 1, 2, 3, 4, 5)
D_values <- c(0, 1)
Q_values <- c(0, 1, 2, 3, 4, 5)
```

Diperoleh hasil menggunakan metode Grid Search sebagai berikut:

```
print(best_model)
Series: data.Curah_hujan
ARIMA(4,1,4) (1,0,0) [12]

Coefficients:
          ar1          ar2          ar3          ar4          ma1          ma2
      -0.4476   0.4000   0.4848   -0.3515   -0.3284   -1.3732
s.e      0.1540   0.1177   0.1506    0.1796    0.1567    0.1750
          ma3          ma4          sar1
      -0.2574   0.9644   0.6571
s.e.      0.1191   0.1373   0.1276

sigma^2 = 5382:  log likelihood = -408.4
AIC=836.8    AICc=840.46    BIC=859.42
```

Sehingga kandidat model bertambah ARIMA(4,1,4)(1,0,0)₁₂

Berikut kandidat-kandidat model yang menurut Anda cocok untuk data.

Model ke-	Nonmusiman	Musiman	Kandidat
1	(0,0,1)	(0,0,1)	(0,0,1)(0,0,1) ₁₂
2	(0,0,1)	(1,0,1)	(0,0,1)(1,0,1) ₁₂
3	(0,0,1)	(2,0,1)	(0,0,1)(2,0,1) ₁₂
4	(1,0,1)	(0,0,1)	(1,0,1)(0,0,1) ₁₂
5	(1,0,1)	(1,0,1)	(1,0,1)(1,0,1) ₁₂
6	(1,0,1)	(2,0,1)	(1,0,1)(2,0,1) ₁₂
7	(2,0,1)	(0,0,1)	(2,0,1)(0,0,1) ₁₂
8	(2,0,1)	(1,0,1)	(2,0,1)(1,0,1) ₁₂
9	(2,0,1)	(2,0,1)	(2,0,1)(2,0,1) ₁₂
10	(4,1,4)	(1,0,0)	(4,1,4)(1,0,0) ₁₂

1.3 Estimasi Model

Berdasarkan spesifikasi model diperoleh tabel nilai AIC dari kandidat model sebagai berikut.

Model ke-	Nonmusiman	Musiman	Kandidat	AIC
1	(0,0,1)	(0,0,1)	(0,0,1)(0,0,1)12	866.14
2	(0,0,1)	(1,0,1)	(0,0,1)(1,0,1)12	857
3	(0,0,1)	(2,0,1)	(0,0,1)(2,0,1)12	858
4	(1,0,1)	(0,0,1)	(1,0,1)(0,0,1)12	864.27
5	(1,0,1)	(1,0,1)	(1,0,1)(1,0,1)12	853
6	(1,0,1)	(2,0,1)	(1,0,1)(2,0,1)12	853
7	(2,0,1)	(0,0,1)	(2,0,1)(0,0,1)12	854.42
8	(2,0,1)	(1,0,1)	(2,0,1)(1,0,1)12	854
9	(2,0,1)	(2,0,1)	(2,0,1)(2,0,1)12	855
10	(4,1,4)	(1,0,0)	(4,1,4)(1,0,0)12	834.8

Sehingga dipilih model ke-5 atau ARIMA(1,0,1)(1,0,1)₁₂ karena memiliki nilai AIC terkecil kedua dengan model yang lebih sederhana, dengan model matematis sebagai berikut:

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})y_t = (1 + \theta_1 B)(1 + \Theta_1 B^{12})e_t$$

$$(1 - \phi_1 B - \Phi_1 B^{12} + \phi_1 \Phi_1 B^{13})y_t = (1 + \theta_1 B + \Theta_1 B^{12} + \theta_1 \Theta_1 B^{13})e_t$$

$$y_t - \phi_1 y_{t-3} - \Phi_1 y_{t-12} + \phi_1 \Phi_1 y_{t-13} = e_t + \theta_1 e_{t-1} + \Theta_1 e_{t-12} + \theta_1 \Theta_1 e_{t-13}$$

$$y_t = e_t + \theta_1 e_{t-1} + \Theta_1 e_{t-12} + \theta_1 \Theta_1 e_{t-13} + \phi_1 y_{t-3} + \Phi_1 y_{t-12} - \phi_1 \Phi_1 y_{t-13}$$

dengan program dan output R sebagai berikut:

```
Call:
arima(x = data.Curah_hujan, order = c(1, 0, 1),
seasonal = list(order = c(1, 0, 1), period = 12))
Coefficients:
          ar1          ma1          sar1          sma1      intercept
        -0.637      0.913      0.999     -0.952             187
s.e.       0.145      0.077      0.006      0.184             30

sigma^2 estimated as 5029:  log likelihood = -421,
aic = 853
```

1.4 Diagnostik Model

Diagnostik model terhadap kandidat model yang terpilih dengan melakukan:

(a) Uji kenormalan sisaan,

Untuk memastikan bahwa data berdistribusi normal dilakukan uji Shapiro-Wilk.

Hipotesis nol (H_0) : data berdistribusi normal.

Hipotesis alternatif (H_1) : data tidak berdistribusi normal.

Dengan program dan output sebagai berikut:

```
Shapiro-Wilk normality test
data:  data.Curah_hujan5$residuals
W = 1, p-value = 0.1
```

Diperoleh nilai p-value = 0.1 yang nilainya lebih besar daripada tingkat signifikansi ($\alpha=0.05$) sehingga diperoleh keputusan bahwa H_0 gagal ditolak atau diperoleh kesimpulan data berdistribusi normal. Sehingga memenuhi asumsi kenormalan residual.

(b) Uji autokorelasi

Uji ini berguna untuk memeriksa apakah terdapat korelasi residual yang signifikan pada berbagai lag dalam deret waktu residual

Hipotesis yang diuji dalam Uji Ljung-Box adalah sebagai berikut:

H_0 : Tidak ada autokorelasi dalam residual model pada lag-lag tertentu.

H_1 : Terdapat autokorelasi dalam residual model pada lag-lag tertentu.

Diperoleh hasil output R sebagai berikut:

```
Box-Ljung test

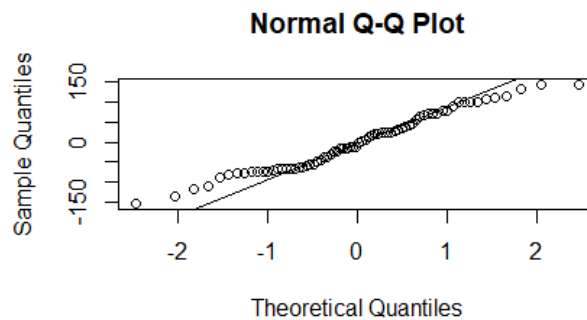
data:  residuals
X-squared = 0.03, df = 1, p-value = 0.9
```

Dengan nilai p-value = 0.9 yang lebih besar daripada nilai signifikan 0.05 maka dapat disimpulkan gagal menolak H_0 dan menyimpulkan bahwa tidak adanya autokorelasi dalam deret waktu pada lag-lag tertentu.

(c) Plot kuantil-kuantil (QQ plot), dan

Diperoleh plot residual untuk model yang terpilih sebagai berikut:

```
#KENORMALAN RESIDUAL  
qqnorm(residuals(data.Curah_hujan5))  
qqline(residuals(data.Curah_hujan5))
```



Dari plot terlihat bahwa titik-titik tersebut sebagian besar terletak di sepanjang garis diagonal lurus dengan sedikit penyimpangan di sepanjang masing-masing ekornya. Berdasarkan plot ini, kita dapat berasumsi bahwa kumpulan residual ini terdistribusi normal.

(d) Uji signifikansi koefisien atau Uji t.

Uji signifikansi koefisien, atau yang sering disebut sebagai Uji-t, digunakan untuk menentukan apakah koefisien regresi suatu variabel dalam model regresi linear signifikan secara statistik. Uji-t dilakukan dengan membandingkan nilai estimasi koefisien dengan standard error dari estimasi tersebut.

Hasil uji t yang diperoleh dengan software R sebagai berikut:

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
ar1	-0.63727	0.14455	-4.41	1.0e-05	***
ma1	0.91305	0.07728	11.81	< 2e-16	***
sar1	0.99919	0.00635	157.47	< 2e-16	***
smal	-0.95220	0.18356	-5.19	2.1e-07	***
intercept	186.85518	29.96008	6.24	4.5e-10	***

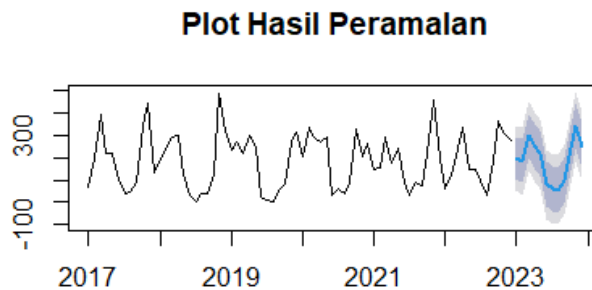
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Diperoleh bahwa nilai p kurang dari tingkat signifikansi 0.05 atau disimpulkan menolak hipotesis nol dan menyimpulkan bahwa koefisien tersebut signifikan secara statistik.

1.5 Peramalan

Plot peramalan terhadap untuk enam dua belas (12) bulan ke depan diperoleh dengan program R dan output sebagai berikut:

```
#Forecasting untuk 12 bulan kedepan
forecasting <- forecast(data.Curah_hujan, model=data.Curah_hujan5
, h=12)
plot(forecasting, main="Plot Hasil Peramalan")
```



Keterangan:

- Data Observasi (Garis Hitam): Garis biru menunjukkan data observasi curah hujan yang kita miliki.
- Peramalan (Garis Biru): Garis hitam mewakili hasil peramalan untuk 12 bulan ke depan.
- Interval Kepercayaan (Area Berwarna Abu-abu): Area abu-abu di sekitar garis hitam menunjukkan interval kepercayaan untuk peramalan. Ini memberikan perkiraan kisaran di mana kita dapat mengharapkan nilai sebenarnya berada dengan tingkat kepercayaan tertentu.

Adapun peramalan dengan menggunakan fungsi predict pada R. Untuk meramalkan 12 bulan kedepan di program R dan outputnya sebagai berikut:

```
od <- options(digits = 5)

fit <- arima(data.Curah_hujan, order=c(1,0,1), seasonal =
list(order = c(1,0,1)))

predict(fit, n.ahead = 12)

options(od)
```

dengan output:

\$pred

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
2023	198.7	186.8	295.3	247.2	216.4	80.6	59.4	55.2
	Sep	Oct	Nov	Dec				
	99.5	239.2	341.7	246.2				

\$se

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
2023	74.6	77.3	78.3	78.7	78.9	79.0	79.0	79.0	79.0
	Oct	Nov	Dec						
	79.0	79.0	79.0						

Sehingga dapat dibuat tabel hasil peramalanya sebagai berikut:

Bulan	Prediksi	Standart error
Januari	198.7	74.6
Febuari	186.8	77.3
Maret	295.3	78.3
April	247.2	78.7
Mei	216.4	78.9
Juni	80.6	79
Juli	59.4	79
Agustus	55.2	79
September	99.5	79
November	239.2	79
Desember	246.2	79

Data tersebut menyatakan hasil ramalan curah hujan Kota Bandung untuk 12 bulan kedepan atau sepanjang tahun 2023 berdasarkan data curah hujan di Kota Bandung pada tahun 2017-2022.

1.6 Daftar Pustaka.

Badan Pusat Statistik (BPS). 2022. Curah Hujan (mm) per Bulan di Kota Bandung (mm), 2017-2022. BPS Kota Bandung

Stephan Kolassa, callmeanythingyouwant. (2021, Juli 22). Selecting the best model parameters from grid search SARIMA. StackExchange. <https://stats.stackexchange.com/questions/535507/selecting-the-best-model-parameters-from-grid-search-sarima>

Tsay, R. and Tiao, G. (1984). "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models." Journal of the American Statistical Association, 79 (385), pp. 84-96

Lampiran syntax R-studio

```
library(tseries)
library(forecast)
library(lmtest)
library(lubridate)
library(TSA)

#####MEMANGGIL DATA CSV#####

setwd("D:/KULIAH RAGIL/SEMESTER 5/Analisis Deret Waktu/final
project")
data<-read.csv("curah_hujan_bandung.csv",
  sep=',',header=TRUE)
data.ts <- ts(data, start = decimal_date(ymd("2017-01-01")),
  frequency = 12)

#####DATA YANG DIGUNAKAN UNTUK DI ANALISIS#####

data.Curah_hujan=ts(data$Curah_hujan,          start          =
  decimal_date(ymd("2017-01-01")), frequency = 12)

#####PLOTING DATA ASLI#####

#memploting data curah hujan
plot.ts(data.Curah_hujan)

#####UJI STASIONER DATA ASLI#####

#uji ADF untuk menguji kestasioneran
adf.test(data.Curah_hujan) #stasioner
acf0 =acf(data.Curah_hujan, lag.max=15)
pacf0= pacf(data.Curah_hujan, lag.max=15)

#menentukan orde p,0,q
eacf(data.Curah_hujan)
#diperoleh model non musiman ARIMA(0,0,1)

##### PEMBENTUKAN MUSIMAN DARI DATA ASLI #####
diff_mus12 <- diff(data.Curah_hujan,lag=12)
plot.ts(diff_mus12)
par(mfrow=c(1,2))
acf(diff_mus12, lag.max=24)
pacf(diff_mus12, lag.max=24)
```

```

#menentukan orde P,0,Q
eacf(diff_mus12)

adf.test(diff_mus12)#stasioner
#jadi tidak perlu di differencing
##jadi kesimpulan musiman adalah modelnya SARIMA(0,0,1)12
##kesimpulan keseluruhan adalah ARIMA(0,0,1)x(0,0,1)12

##### METODE GRID SEARCH #####
# Tentukan rentang nilai parameter
p_values <- c(0, 1, 2,3,4,5)
d_values <- c(0, 1)
q_values <- c(0, 1, 2,3,4,5)
P_values <- c(0, 1, 2,3,4,5)
D_values <- c(0, 1)
Q_values <- c(0, 1, 2,3,4,5)

# Buat grid kombinasi parameter
parameter_grid <- expand.grid(p = p_values, d = d_values, q
                             = q_values,
                             P = P_values, D = D_values, Q
                             = Q_values)

# Inisialisasi variabel untuk menyimpan hasil evaluasi
best_aic <- Inf
best_model <- NULL

# Iterasi melalui kombinasi parameter
for (i in 1:nrow(parameter_grid)) {
  current_param <- parameter_grid[i, ]

  # Membangun model SARIMA
  current_model <- Arima(data.Curah_hujan, order =
    c(current_param$p, current_param$d, current_param$q),
    seasonal = list(order =
    c(current_param$P, current_param$D, current_param$Q)))

  # Evaluasi model menggunakan AIC
  current_aic <- AIC(current_model)

  # Membandingkan dengan model terbaik sejauh ini
  if (current_aic < best_aic) {
    best_aic <- current_aic
    best_model <- current_model
  }
}

```



```

# Menampilkan model terbaik
print(best_model)

###ESTIMASI PARAMETER DAN PEMILIHAN MODEL TERBAIK#####

data.Curah_hujan1 <- arima(data.Curah_hujan, order=c(0,0,1),
  seasonal = list(order = c(0,0,1), period = 12))
data.Curah_hujan1

data.Curah_hujan2 <- arima(data.Curah_hujan, order=c(0,0,1),
  seasonal = list(order = c(1,0,1), period = 12))
data.Curah_hujan2

data.Curah_hujan3 <- arima(data.Curah_hujan, order=c(0,0,1),
  seasonal = list(order = c(2,0,1), period = 12))
data.Curah_hujan3

data.Curah_hujan4 <- arima(data.Curah_hujan, order=c(1,0,1),
  seasonal = list(order = c(0,0,1), period = 12))
data.Curah_hujan4

data.Curah_hujan5 <- arima(data.Curah_hujan, order=c(1,0,1),
  seasonal = list(order = c(1,0,1), period = 12))
data.Curah_hujan5

data.Curah_hujan6 <- arima(data.Curah_hujan, order=c(1,0,1),
  seasonal = list(order = c(2,0,1), period = 12))
data.Curah_hujan6

data.Curah_hujan7<- arima(data.Curah_hujan, order=c(2,0,1),
  seasonal = list(order = c(0,0,1), period = 12))
data.Curah_hujan7

data.Curah_hujan8 <- arima(data.Curah_hujan, order=c(2,0,1),
  seasonal = list(order = c(1,0,1), period = 12))
data.Curah_hujan8

data.Curah_hujan9<- arima(data.Curah_hujan, order=c(2,0,1),
  seasonal = list(order = c(2,0,1), period = 12))
data.Curah_hujan9

data.Curah_hujan10<- arima(data.Curah_hujan, order=c(4,1,4),
  seasonal = list(order = c(1,0,0),period = 12))
data.Curah_hujan10

```

```
#####DIAGNOSTIK RESIDUAL#####
plot(rstandard(data.Curah_hujan5))

#KENORMALAN RESIDUAL
qqnorm(residuals(data.Curah_hujan5))
qqline(residuals(data.Curah_hujan5))

shapiro.test(data.Curah_hujan5$residuals)

#Uji Kesignifikanan parameter: Uji t
coeftest(data.Curah_hujan5)

#AUTOKORELASI RESIDUAL
residuals <- residuals(data.Curah_hujan5)
ljung_box_test <- Box.test(residuals, type = "Ljung-Box")
print(ljung_box_test)

#####Forecasting untuk 12 bulan kedepan#####
forecasting <- forecast(data.Curah_hujan,
  model=data.Curah_hujan5, h=12)
plot(forecasting, main="Plot Hasil Peramalan")

od <- options(digits = 3)
fit <- arima(data.Curah_hujan, order=c(1,0,1), seasonal =
  list(order = c(1,0,1), period=12))
predict(fit, n.ahead = 12)
options(od)
```