

# **LAPORAN PERAMALAN (*FORECASTING*) DATA RUNTUN WAKTU PENJUALAN TOKO GLOBAL**



Dibuat Oleh :  
Kelompok 5

Jasmine Husna Sanditya	2006571034
Joan Bidadari Annandale	2006571085
Muhammad Jauhar Hakim	2006463982

Program Studi Statistika  
Departemen Matematika  
Fakultas Matematika dan Ilmu Pengetahuan Alam  
Semester Genap 2022  
Universitas Indonesia

## KATA PENGANTAR

Puji dan syukur kami panjatkan ke hadirat Tuhan YME, karena atas limpahan rahmat dan karunianya, kami berhasil menyelesaikan laporan praktikum PERAMALAN (*FORECASTING*) DATA RUNTUN WAKTU PENJUALAN TOKO GLOBAL ini dengan baik. Laporan ini kami susun untuk melengkapi tugas akhir Metode Peramalan Program Studi Statistika Fakultas Ilmu Pengetahuan dan Alam Universitas Indonesia.

Terlebih dahulu, kami mengucapkan terima kasih kepada Ibu Mila Novita, S.Si, M.Si, selaku Dosen Metode Peramalan yang telah memberikan tugas ini sehingga dapat menambah pengetahuan dan wawasan sesuai dengan bidang studi yang kami tekuni ini. Kami juga mengucapkan terima kasih kepada semua pihak yang tidak dapat kami sebutkan semua, terima kasih atas bantuannya sehingga kami dapat menyelesaikan tugas akhir ini.

Namun, dalam penyusunan laporan ini kami menyadari masih terdapat banyak kekurangan. Oleh karena itu, saran dan kritik yang membangun sangat kami harapkan. Akhir kata, semoga laporan ini bermanfaat bagi penyusun dan seluruh pihak pada umumnya.

Depok, 10 Juni 2022

Penyusun

# DAFTAR ISI

KATA PENGANTAR .....	I
DAFTAR ISI.....	II
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	2
1.3. Tujuan Penelitian.....	2
1.4. Pembatasan Masalah .....	2
BAB II TINJAUAN PUSTAKA .....	3
2.1. Analisis Runtun Waktu .....	3
2.2. Metode ARIMA Box-Jenkins .....	3
2.2.1. Proses White Noise.....	3
2.2.2. Model AR (Autoregressive) .....	4
2.2.3. Model MA (Moving Average) .....	4
2.2.4. Model ARMA (Autoregressive Moving Average).....	4
2.2.5. Model ARIMA (Autoregressive Integrated Moving Average) .....	5
2.3. Uji Stasioneritas .....	5
2.3.1. Korelogram.....	5
2.3.2. Uji Unit Akar.....	8
2.4. Spesifikasi Model ARIMA.....	9
2.5. Estimasi Parameter.....	12
2.5.1. Metode Momen.....	12
2.5.2. Metode Least-Square.....	12
2.5.3. Metode Maximum Likelihood.....	12
2.6. Diagnosis Model .....	13
2.6.1. Uji Independensi Ljung Box.....	13
2.6.2. Uji Normalitas Jarque-Bera.....	14
2.6.4. Overfitting Model.....	15
2.6.5. Cross Validation .....	15
2.7. Forecasting .....	16

BAB III METODOLOGI PENELITIAN .....	17
3.1. Jenis Peramalan .....	17
3.2. Sumber dan Penjelasan Data .....	17
3.3. Langkah Kerja .....	18
 BAB IV HASIL DAN PEMBAHASAN .....	19
4.1. Preprocessing Data .....	19
4.2. Identifikasi Stasioneritas Data .....	20
4.2.1. Uji Stasioner Data berdasarkan Plot .....	20
4.2.2. Uji Stasioner Data dengan ADF .....	21
4.2.3. Melakukan Differencing Data .....	21
4.2.4. Uji Stasioner Data Differencing berdasarkan Plot .....	22
4.2.5. Uji Stasioner Data Differencing dengan ADF .....	23
4.3. Spesifikasi Model .....	23
4.4. Diagnostik Model .....	26
4.4.1. Uji Independensi Ljung Box .....	26
4.4.2. Uji Normalitas Jarque-Bera .....	27
4.4.3. Overfitting Model .....	28
4.5. Analisis Model .....	29
4.6. <i>Forecasting</i> (Peramalan) .....	31
 BAB V KESIMPULAN .....	36
5.1. Kesimpulan .....	36
5.2. Saran .....	36
 DAFTAR PUSTAKA .....	37

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Suatu perusahaan yang bergerak dalam bidang jasa penjualan atau pendistribusian produk, selalu menginginkan keberhasilan dalam aktivitasnya dimasa yang akan datang. Ini menunjukkan bahwa setiap perusahaan selalu berusaha untuk tetap berkembang dalam bidang usahanya dimasa depan. Semakin ketatnya persaingan di antara perusahaan-perusahaan tersebut sebuah perusahaan harus memiliki strategi perusahaan yang tepat dan efektif dalam memenangkan persaingan yang terjadi.

Perusahaan yang memiliki kemampuan untuk dapat menetapkan keputusan yang tepat dalam menghadapi masa depan yang penuh ketidakpastian adalah perusahaan yang dapat meraih apa yang menjadi tujuannya. Salah satu hal yang paling penting untuk mewujudkan hal tersebut adalah memperkirakan atau meramal (*forecasting*) besarnya penjualan atau permintaan pelanggan akan barang yang dihasilkan.

Peramalan (*forecasting*) adalah kegiatan mengestimasi apa yang akan terjadi pada masa yang akan datang. Peramalan diperlukan karena terdapat perbedaan kesenjangan waktu (*Timelag*) antara kesadaran akan dibutuhkannya suatu kebijakan baru dengan waktu pelaksanaan kebijakan tersebut. Apabila perbedaan waktu tersebut panjang maka peran peramalan begitu penting dan sangat dibutuhkan, terutama dalam penentuan kapan terjadinya suatu sehingga dapat dipersiapkan tindakan yang perlu dilakukan. Heizer dan Render (2015:113) mendefinisikan peramalan (*forecasting*) adalah suatu seni dan ilmu pengetahuan dalam memprediksi peristiwa pada masa mendatang. Peramalan akan melibatkan pengambilan data historis (penjualan tahun lalu) dan memproyeksi mereka ke masa yang akan datang dengan model matematika.

## 1.2. Rumusan Masalah

Akan dilakukan proses peramalan (*forecasting*) untuk meramalkan permintaan konsumen sehingga dapat memperkirakan penjualan dimasa yang akan datang. Terdapat beberapa model yang dapat diajukan dalam meramalkan permintaan konsumen. Akan dipilih model terbaik yang diusahakan memiliki error terkecil dan dapat meramalkan masa depan dengan baik. Peramalan (*forecasting*) sangat penting dilakukan agar tidak terjadinya rugi akibat barang lama yang rusak ataupun basi sehingga diperlukan data mengenai seberapa banyak stok yang harus disiapkan.

## 1.3. Tujuan Penelitian

- a. Mampu mengetahui manfaat dari peramalan (*forecasting*)
- b. Mampu memahami metode dan teknik dalam peramalan (*forecasting*) beserta penerapannya
- c. Dapat digunakan sebagai pencegahan kerugian penjualan toko di masa depan

## 1.4. Pembatasan Masalah

Peramalan (*forecasting*) akan dilakukan dengan menggunakan data penjualan global toko super yang sudah berlangsung selama 4 tahun yaitu dari tahun 2015 hingga tahun 2018. Di dalam *dataset* penjualan global toko super terdapat waktu pemesanan, waktu pengantaran, nama pembeli, asal pembeli, kategori barang, dan harga barang yang dipesan.

Peramalan (*forecasting*) akan dilakukan dengan menjumlahkan penjualan per bulan terlebih dahulu sehingga akan didapatkan data runtun waktu dengan jeda tiap bulan. Lalu akan dilakukan peramalan dengan dasarnya adalah total penjualan yang dimulai dari bulan Januari 2015 hingga bulan Desember 2018.

Peramalan (*forecasting*) akan dilakukan dengan menggunakan metode atau model Box-Jenkins yang merupakan salah satu metode peramalan runtun waktu. Metode ini menggunakan nilai di masa lalu sebagai variabel dependen dan variabel independen diabaikan. Terdapat tiga langkah utama dalam proses pembuatan model ini yaitu *model spesification*, *model fitting*, dan *model diagnostics*.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Analisis Runtun Waktu

Analisis runtun waktu adalah sebuah analisis data yang dilakukan dengan menganalisis suatu periode waktu pada masa lalu untuk meramalkan masa depan. Dalam analisis runtun waktu akan dilakukan analisis hubungan antara variabel yang akan diprediksi terhadap variabel waktu. Waktu yang biasa digunakan bersifat diskrit. Beberapa contoh dari waktu diskrit adalah jam, hari, bulan, dan tahun. Metode menganalisis runtun waktu yang paling sering digunakan adalah metode ARIMA Box-Jenkins yang memanfaatkan data stasioner.

#### 2.2. Metode ARIMA Box-Jenkins

ARIMA (Autoregressive Integrated Moving Average) adalah suatu model yang digunakan untuk memprediksi data di masa depan berdasarkan data dari masa lalu. Model ARIMA biasanya ditulis sebagai ARIMA  $(p,d,q)$  dengan  $p$  adalah order untuk model AR  $(p)$ ,  $d$  merupakan banyaknya differencing yang dilakukan pada data, dan  $q$  merupakan order dari model MA  $(q)$ . Berikut penjelasan jenis-jenis model dari ARIMA Box-Jenkins;

##### 2.2.1. Proses White Noise

Proses white noise adalah barisan variabel acak  $\{e_t\}$  yang tidak berkorelasi. Proses white noise bersifat stasioner dan memiliki  $\mu = 0$  dan  $Var(e_t) = \sigma_e^2$  sehingga dapat dinotasikan dengan  $e_t \sim WN(0, \sigma_e^2)$ . Proses ini merupakan faktor pembangun dalam proses runtun waktu lainnya.

### 2.2.2. Model AR (Autoregressive)

Model autoregressive memiliki order sampai ke- $p$  yang dinyatakan sebagai AR ( $p$ ). Model AR ( $p$ ) dapat dinyatakan sebagai:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t$$

$Y_t$  = Nilai pada periode ke- $t$ ,

$\phi_p$  = Parameter autoregressive ke- $p$ ,

$Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$  = variabel bebas atau nilai masa lalu deret waktu yang berkaitan,  $e_t$  = White noise alias suku error ke- $t$  yang dihasilkan proses acak yang independen terhadap  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ .

### 2.2.3. Model MA (Moving Average)

Proses MA (Moving Average) merupakan fungsi linier dari error yang dihasilkan dari proses white noise peramalan sekarang dan periode-periode sebelumnya secara berurutan dengan suatu bobot  $\theta$  (parameter). Model Moving Average memiliki order sampai ke- $q$  yang dinyatakan sebagai MA ( $q$ ). Berikut merupakan model MA ( $q$ ).

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

$Y_t$  = nilai pada periode ke- $t$ ,

$\theta q$  = parameter Moving Average ke- $q$ ,

$e_{t-1}, e_{t-2}, \dots, e_{t-q}$  = variabel bebas atau nilai error masa lalu deret waktu yang berkaitan

$e_t$  = suku error pada saat  $t$  yang dihasilkan proses acak (white noise).

### 2.2.4. Model ARMA (Autoregressive Moving Average)

Gabungan antara model AR ( $p$ ) dan MA ( $q$ ), dinotasikan dengan ARMA ( $p, q$ ) disebut sebagai model ARMA. Model ARMA biasa dituliskan sebagai berikut

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$



### 2.2.5. Model ARIMA (Autoregressive Integrated Moving Average)

ARIMA merupakan model campuran dari proses AR ( $p$ ) dan MA ( $q$ ) yang telah dilakukan differencing  $d$  kali. Notasi dari model ini adalah ARIMA ( $p,d,q$ ). Proses runtun waktu  $Y_t$  yang dilakukan differencing  $d$  kali dinotasikan dengan  $W_t = \nabla^d Y_t$ . Model ARIMA ( $p,d,q$ ) dituliskan dalam bentuk seperti berikut.

$$W_t = \Delta^d Y_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

Jika  $p = 0$ , maka model ARIMA ( $0,d,q$ ) disebut sebagai model IMA (Integrated Moving Average) yang dinotasikan dengan IMA ( $d,q$ ). Jika  $q = 0$  maka model ARIMA ( $p,d,0$ ) disebut sebagai ARI (Autoregressive Integrated) yang dinotasikan dengan ARI ( $p,d$ ).

## 2.3. Uji Stasioneritas

### 2.3.1. Korelogram

Korelogram adalah teknik identifikasi kestasioneran data time series melalui fungsi autokorelasi (ACF) yang memberikan gambaran bagaimana korelasi antara data yang berdekatan. Fungsi autokorelasi dengan lag ke- $k$  adalah sebagai berikut.

$$\rho_k = \frac{Cov(Y_t, Y_{tk})}{\sigma_{Y_t} \sigma_{Y_{tk}}}; |\rho_k| \leq 1$$

Untuk proses yang stasioner, kita memiliki  $Var(Y_t) = Var(Y_{t+k})$  sehingga fungsi autokorelasi ditulis sebagai berikut:

$$\rho_k = \frac{Cov(Y_t, Y_{tk})}{\sigma_{Y_t}^2}; \gamma_k \gamma_0$$

Korelogram akan dihasilkan dengan membuat plot antara  $\rho_k$  dengan  $k$ . Tetapi, yang dapat dihitung hanyalah fungsi otokorelasi sampel (Sample Autocorrelation Function), yaitu

$$\hat{\rho}_k = r_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \text{ di mana}$$

$$\hat{\gamma}_k = \frac{\sum (Y_t - \bar{Y}_t)(Y_{t+k} - \bar{Y}_{t+k})}{n}, \hat{\gamma}_0 = \frac{\sum (Y_t - \bar{Y}_t)^2}{n}$$

Plot antara  $r_k$  dengan  $k$  disebut sampel korelogram. Untuk uji stasioner, plot ini bersifat subjektif.

Secara statistik, terdapat tiga pengujian stasioner, yaitu :

### 1. Uji Bartlett

Uji Bartlett ini digunakan untuk mengamati signifikansi  $r_k$  satu per satu.

#### a) Hipotesis

$$H_0: \rho_k = 0, k > 0$$

$$H_1: \rho_k \neq 0$$

#### b) Statistik uji:

Dengan  $(r_k \sim N(\rho_k, Var(r_k)))$ , maka

$$\frac{r_k - \rho_k}{\sqrt{Var(r_k)}} \sim N(0,1)$$

#### c) Aturan keputusan:

$H_0$  gagal diterima bila

$$\left| \frac{r_k - \rho_k}{\sqrt{Var(r_k)}} \right| > Z_{\frac{\alpha}{2}}$$

#### d) Kesimpulan:

Bila  $H_0$  ditolak maka dapat disimpulkan bahwa runtun waktu yang dimiliki tidak berasal dari proses *white noise*.

### 2. Uji Box-Pierce

Uji Box-Pierce digunakan untuk mengetahui apakah nilai dari  $\rho_k$  pada sekumpulan waktu berbeda dengan nol.

#### a) Hipotesis

$$H_0: \forall \rho_k = 0,$$

$$H_1: \exists \rho_k \neq 0,$$

b) Statistik uji:

$$Q = T \sum_{k=1}^m r_k^2$$

$T$  = Banyak pengamatan

$m$  = panjang lag  $\left(\frac{1}{3}T\right)$

c) Aturan keputusan:

$H_0$  gagal diterima bila  $Q > X_{m,a}^2$

d) Kesimpulan:

Bila  $H_0$  ditolak maka dapat disimpulkan bahwa runtun waktu yang dimiliki tidak berasal dari proses *white noise*.

### 3. Uji Ljung-Box

Uji ini mirip dengan uji Box-Pierce tetapi uji ini dilakukan untuk observasi  $n$  yang lebih besar. Selain itu, bentuk  $r_k^2$  diganti menjadi  $r_k^{*2} = \frac{n+2}{n-2} r_k^2$

a) Hipotesis

$$H_0: \forall \rho_k = 0$$

$$H_1: \exists \rho_k \neq 0$$

b) Statistik uji:

$$LB = (n)(n+2) \sum_{k=1}^m \left( \frac{r_k^{*2}}{n-k} \right)^2$$

c) Aturan keputusan:

$H_0$  ditolak bila  $LB > X_{m,a}^2$

d) Kesimpulan:

Bila  $H_0$  ditolak maka dapat disimpulkan bahwa runtun waktu yang dimiliki tidak berasal dari proses *white noise*.

### 2.3.2. Uji Unit Akar

#### 1. Uji Dickey Fuller (DF) pada model AR (1):

$$Y_t = \phi Y_{t-1} + e_t$$

Setelah kita mencari perubahan  $Y$ -nya, didapat:

$$Y_t - Y_{t-1} = (\phi - 1)Y_{t-1} + e_t$$

$$\Delta Y_t = \pi Y_{t-1} + e_t$$

Maka kita akan melakukan uji terhadap hipotesis berikut.

#### 1. Hipotesis

$H_0: \pi = 0$  (data tidak stasioner atau mempunyai unit root)

$H_1: \pi < 0$  (data stasioner)

#### 2. Statistik Uji

$$\tau = \frac{\hat{\theta} - 1}{se(\hat{\theta})} = \frac{\hat{\pi}}{se(\hat{\pi})}$$

Uji- $\tau$  terhadap hipotesis ini mengikuti distribusi  $t$ .

#### 3. Aturan Keputusan

$H_0$  ditolak bila  $|\tau| > |DF_{tabel}|$ .

#### 4. Kesimpulan

Jika  $H_0$  ditolak, maka dapat disimpulkan bahwa data stasioner atau tidak memiliki unit root. Dengan asumsi  $u_t$  tidak berkorelasi, uji DF dapat dilakukan dengan menggunakan model dengan intersep atau model dengan intersep dan memasukkan variabel bebas waktu ( $t$ ).

## 2. Uji Augmented Dickey-Fuller (ADF)

Uji ADF merupakan pengembangan dari uji DF dan digunakan untuk model  $AR(p)$  dengan  $p > 1$ .

### 1. Hipotesis

$H_0: \pi = 0$  (data tidak stasioner atau mempunyai unit root)

$H_1: \pi < 0$  (data stasioner)

### 2. Statistik Uji

$$\tau = \frac{\hat{\theta} - 1}{se(\hat{\theta})} = \frac{\hat{\pi}}{se(\hat{\pi})}$$

Uji- $\tau$  terhadap hipotesis ini mengikuti distribusi t.

### 3. Aturan Keputusan

$H_0$  ditolak jika  $|\tau| > |ADF_{tabel}|$ .

### 4. Kesimpulan

Jika  $H_0$  ditolak, maka dapat disimpulkan bahwa data stasioner atau tidak memiliki unit root. Dengan asumsi  $u_t$  tidak berkorelasi, uji ADF dapat dilakukan dengan menggunakan model dengan intersep dan tren, model dengan intersep saja, atau model tanpa intersep dan tren.

## 2.4. Spesifikasi Model ARIMA

Akan dilakukan identifikasi untuk nilai  $p$ ,  $d$ , dan  $q$  terhadap model runtun yang stasioner agar dapat ditentukan model yang sesuai untuk ARIMA. Metode-metode yang akan dilakukan adalah; Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), Extended Autocorrelation Function (EACF), Aike Information Criterion (AIC) dan Bayesian Information Criterion (BIC).

## 1. Autocorrelation Function (ACF)

Untuk runtun waktu teramati  $Y_1, Y_2, \dots, Y_n$  maka

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

Untuk  $k = 1, 2, \dots$ , maka  $r_k$  merupakan taksiran parameter untuk  $\rho_k$ . ACF bertujuan untuk mengidentifikasi model  $MA(q)$  tetapi memiliki kekurangan yaitu tidak dapat mengidentifikasikan model  $AR(p)$  secara akurat. Grafik ACF merupakan visualisasi dari uji hipotesis  $H_0: \rho_k = 0$ . Bila di grafik ACF terdapat tiang pancang yang melebihi garis putus-putus pada lag- $q$  maka dapat disimpulkan bahwa terdapat  $\rho_k$  yang signifikan sehingga model  $MA(q)$  dapat dipertimbangkan. Apabila terbentuk pola menyusut secara eksponensial atau adanya gelombang sinus yang kurang jelas, maka  $AR(p)$  bisa menjadi model yang bisa dipertimbangkan.

## 2. Partial Autocorrelation Function (PACF)

PACF adalah fungsi yang didefinisikan sebagai korelasi antara  $Y_t$  dan  $Y_{t-k}$  setelah menghilangkan efek dari variabel  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$ . PACF didefinisikan sebagai berikut

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \dots + \phi_{kk}\rho_{j-k}$$

PACF dapat mengidentifikasi model  $AR(p)$  dengan baik tetapi tidak dapat mengidentifikasikan model  $MA(q)$  secara akurat. Apabila terdapat suatu tiang pancang yang sangat jelas (terdapat tiang pancang yang melebihi garis putus-putus) pada lag- $p$ , maka model  $AR(p)$  dapat dipertimbangkan. Apabila terbentuk pola menyusut secara eksponensial atau adanya gelombang sinus yang kurang jelas, maka  $AR(p)$  bisa menjadi model yang bisa dipertimbangkan

### 3. Extended Autocorrelation Function (EACF)

EACF diajukan sebagai salah satu metode untuk menentukan orde dari model ARMA(p, q) dikarenakan untuk model ARMA (p, q) tidak dapat ditentukan ordenya dengan mudah hanya dengan menggunakan ACF dan PACF sederhana.. Metode EACF komponen AR akan mengabaikan dari data sehingga hanya tersisa komponen MA untuk diidentifikasi oleh ACF. EACF adalah autokorelasi sampel dari  $W_{t,k,j}$  yang definisinya dapat ditulis sebagai berikut:

$$W_{t,k,j} = Y_t - \widetilde{\phi}_1 Y_{t-1} - \cdots - \widetilde{\phi}_k Y_{t-k}$$

### 4. Memilih model terbaik

Setelah dilakukannya ACF, PACF, EACF untuk mencari model-model yang dapat digunakan, akan dicari model yang terbaik dengan memperhatikan kriteria-kriteria yang memaksimalkan nilai dari estimasi Likelihood melalui dua metode berikut.

#### i. Aike Information Criterion (AIC)

Model yang dipilih adalah yang meminimalkan

$$AIC = -2 \log (\text{maksimum likelihood}) + 2k$$

Dimana  $k = p + q + 1$  jika model mengandung suku konstanta dan  $k = p + q$  untuk yang lainnya.

#### ii. Bayesian Information Criterion (BIC)

Model yang dipilih adalah yang meminimalkan

$$BIC = -2 \log (\text{maksimum likelihood}) + k \log n$$

Dimana  $k = p + q + 1$  jika model mengandung suku konstanta dan  $k = p + q$  untuk yang lainnya.

## 2.5. Estimasi Parameter

Setelah melalui pemilihan model dengan memperhatikan nilai AIC dan BIC maka akan didapatkan suatu model  $ARIMA(p,d,q)$  yang ingin digunakan sebagai model. Nilai  $d$  sudah didapatkan dari banyaknya differencing yang dijalankan pada data. Lalu akan dikerjakan mengenai tiga metode untuk menaksirkan nilai  $ARMA(p,q)$ .

### 2.5.1. Metode Momen

Metode momen bekerja dengan menyelesaikan persamaan dari momen sampel dengan momen teoritis. Metode ini cukup bagus untuk model  $AR$ . Contoh penggunaannya adalah menaksir mean proses stasioner dengan mean sampel.

### 2.5.2. Metode Least-Square

Metode ini meminimumkan jumlah kuadrat dari fungsi dibawah ini.

$$SS = \sum_{t=2}^n (Y_t - \mu - \phi(Y_{t-1} - \mu))^2$$

dimana selisih kuadrat dari  $Y_t - \mu$  dengan  $\phi(Y_{t-1} - \mu)$  akan diminimumkan untuk mencari  $\hat{\phi}$ .

### 2.5.3. Metode Maximum Likelihood

Metode ini bekerja dengan mencari nilai maksimum dari fungsi Likelihood. Metode ini bekerja dengan baik untuk model yang memiliki runtun waktu yang pendek dan model yang bersifat musiman. Namun, dikarenakan metode ini memaksimumkan fungsi Likelihood maka perlu diketahui joint pdf dari fungsi Likelihood yang akan dimaksimumkan.



Metode estimasi parameter yang digunakan pada penelitian ini adalah Conditional Sum Square - Maximum Likelihood Estimation (MLE). Metode ini memiliki konvergensi yang lebih cepat dibandingkan dengan metode MLE dengan Unconditional Sum Square sehingga lebih efisien untuk dipilih. Sebagai contoh, untuk kasus  $AR(1)$ , CSSMLE akan memaksimumkan fungsi likelihood  $L(\phi, \mu, \sigma_e^2)$ , di mana

$$L(\phi, \mu, \sigma_e^2) = (2\pi\sigma_e^2)^{-\frac{n}{2}} (1 - \phi^2)^{-\frac{1}{2}} \exp \left[ -\frac{1}{2\sigma_e^2} S(\phi, \mu) \right]$$

Dimana

$$S(\phi, \mu) = \sum_{t=2}^n [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2$$

Sedangkan untuk Unconditional Sum Square – MLE,

$$S(\phi, \mu) = \sum_{t=2}^n [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2 + (1 - \phi^2)(Y_1 - \mu)^2$$

Perhatikan bahwa  $(1 - \phi^2)(Y_1 - \mu)^2$  mengakibatkan persamaan  $\frac{\partial S(\phi, \mu)}{\partial \theta} = 0$

dan  $\frac{\partial S(\phi, \mu)}{\partial \mu} = 0$  menjadi non-linier terhadap  $\phi$  dan  $\mu$  yang mempengaruhi konvergensi dan nilai parameter.

## 2.6. Diagnosis Model

Bagian ini bertujuan untuk mengecek apakah model sudah cukup baik. Apabila model dianggap belum cukup baik maka model akan dilakukan modifikasi.

### 2.6.1. Uji Independensi Ljung Box

Langkah-langkah pengujian:

1. Hipotesis:

$H_0$ : Tidak ada korelasi antar residual

$H_1$ : Terdapat korelasi antar residual

2. Statistik uji:

$$Q = n(n+2) \sum_{k=1}^m \frac{r_k^2}{n-k}$$

3. Aturan keputusan:

$H_0$  ditolak jika  $Q > X_{\alpha,m}^2$  atau  $p\text{-value} < \alpha$

### 2.6.2. Uji Normalitas Jarque-Bera

Langkah-langkah pengujian:

1. Hipotesis:

$H_0$ : Data berdistribusi Normal

$H_1$ : Data tidak berdistribusi Normal

2. Statistik uji:

$$JB = n \left[ \frac{S^2}{6} + \frac{(K-3)^2}{24} \right]$$

$K$  = Koefisien Kurtosis

$S$  = Koefisien Skewness

3. Aturan keputusan:

$H_0$  ditolak jika  $JB > X_{\alpha,2}^2$  atau  $p\text{-value} < \alpha$

### 2.6.3 Uji Kestasioneran Augmented Dickey-Fuller (ADF)

Uji ADF merupakan pengembangan dari uji DF dan digunakan untuk model  $AR(p)$  dengan  $p > 1$ .

#### 1. Hipotesis

$H_0 : \pi = 0$  (data tidak stasioner atau mempunyai unit root)

$H_1 : \pi < 0$  (data stasioner)

#### 2. Statistik Uji

$$\tau = \frac{\hat{\theta} - 1}{se(\hat{\theta})} = \frac{\hat{\pi}}{se(\hat{\pi})}$$

Uji- $\tau$  terhadap hipotesis ini mengikuti distribusi t.

#### 3. Aturan Keputusan

$H_0$  ditolak jika  $|\tau| > |ADF_{tabel}|$ .

#### 4. Kesimpulan

Jika  $H_0$  ditolak, maka dapat disimpulkan bahwa data stasioner.

### 2.6.4. Overfitting Model

Jika hasil dari dua pengujian sebelumnya memperlihatkan bahwa model cukup baik, maka akan dilakukan overfit dengan parameter model yang lebih besar. Dapat dianggap bahwa model yang digunakan adalah model terbaik apabila pada saat dilakukan overfit didapatkan signifikansi parameter yang rendah. Setelah itu akan digunakan prinsip parsimonin yaitu pemilihan parameter yang lebih sederhana. Saat melakukan overfit sebaiknya tidak meningkatkan order AR dan MA secara bersamaan untuk mendapatkan hasil yang paling tepat dan pengecekan yang paling efisien.

### 2.6.5. Cross Validation

Data aktual akan dipecahkan menjadi dua bagian yaitu data estimasi dan data prediksi. Bila nilai aktual masuk pada interval kepercayaan nilai prediksi saat dibandingkan maka dapat disimpulkan bahwa model sudah sesuai.

## 2.7. Forecasting

Data aktual akan dibentuk menjadi model untuk memprediksi data di masa depan menggunakan model ARIMA yang sudah ditentukan. Akan diharapkan hasil prediksi hampir sama dengan data aktual di masa depan atau dengan kata lain memiliki nilai residual yang rendah. Didefinisikan ekspektasi dari forecast sepanjang 1 waktu kedepan dengan metode minimum mean square error sebagai berikut:

$$\hat{Y}_t(l) = E(Y_{t+1}|(Y_1, Y_2, \dots, Y_t)|)$$

## BAB III

### METODOLOGI PENELITIAN

#### 3.1. Jenis Peramalan

Jenis peramalan yang kami gunakan adalah ex-post forecasting. Ex-post forecasting adalah jenis peramalan yang menggunakan informasi di luar waktu dimana ramalan tersebut dibuat. Analisis ex-post mempertimbangkan nilai awal dan akhir dan juga kemajuan dan penurunan dari sebuah nilai seiring berjalannya waktu. Hal itu digunakan untuk memprediksi nilai yang ada di masa depan. Selain itu metode yang kami gunakan adalah metode ARIMA Box-Jenkins. Metode ARIMA Box-Jenkins menggunakan pendekatan iteratif dalam mengidentifikasi suatu model sehingga merupakan model yang paling tepat.

#### 3.2. Sumber dan Penjelasan Data

Data yang kami gunakan adalah data *'Superstore Sales Dataset'* diambil dari situs kaggle yaitu <https://www.kaggle.com/datasets/rohitsahoo/sales-forecasting>. Data ini berisikan sales dari sebuah retail dalam kurun waktu empat tahun. Data terdiri dari nama customer, barang yang dibeli, tanggal pembelian, asal kota, sales, dan lain sebagainya. Berikut adalah cuplikan dari data yang kami gunakan:

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country
1	CA-2017-152156	08/11/2017	11/11/2017	Second Class	CG-12520	Claire Gite	Consumer	United States
2	CA-2017-152156	08/11/2017	11/11/2017	Second Class	CG-12520	Claire Gite	Consumer	United States
3	CA-2017-138688	12/06/2017	16/06/2017	Second Class	DV-13845	Darrin Van Huff	Corporate	United States
4	US-2016-108966	11/10/2016	18/10/2016	Standard Class	SO-28335	Sean O'Donnell	Consumer	United States
5	US-2016-108966	11/10/2016	18/10/2016	Standard Class	SO-28335	Sean O'Donnell	Consumer	United States

City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales
Henderson	Kentucky	42428	South	FUR-BD-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.96
Henderson	Kentucky	42428	South	FUR-OH-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	731.94
Los Angeles	California	90036	West	OFF-LA-10000240	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters by Universal	14.62
Fort Lauderdale	Florida	33311	South	FUR-TA-10000577	Furniture	Tables	Bretford CR4508 Series Slim Rectangular Table	957.5775
Fort Lauderdale	Florida	33311	South	OFF-ST-10000768	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	22.368

### 3.3. Langkah Kerja

Berikut adalah langkah kerja untuk melakukan forecasting data:

1. Identifikasi stasioneritas data, jika data tidak stasioner maka lakukan differencing hingga data stasioner.
2. Lakukan spesifikasi model ARIMA dari data.
3. Estimasi parameter dari model yang dimiliki data.
4. Lakukan diagnostik model, jika model tidak sesuai maka ulang lagi dari tahap pertama hingga model sesuai.
5. Lakukan forecasting data untuk mendapatkan prediksi data.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1. Preprocessing Data

Akan dilakukan penjumlahan penjualan berdasarkan bulan yaitu dari bulan Januari 2015 hingga bulan Desember 2018. Dengan menggunakan program R didapatkan :

```
#Mengubah kolom Order.Date menjadi tipe datetime
data[['Order.Date']] <- as.POSIXct(data[['Order.Date']], format =
"%d/%m/%Y")

#Mengubah Format tanggal nya menjadi tahun-bulan-hari
#Drop semua kolom kecuali Order.Date dan Sales
databaru <- subset(data, select = c("Order.Date", "Sales"))
head(databaru)
      Order.Date  Sales
1  2017-11-08  261.9600
2  2017-11-08  731.9400
3  2017-06-12   14.6200
4  2016-10-11  957.5775
5  2016-10-11   22.3680
6  2015-06-09   48.8600

#memeriksa jumlah missing value
sum(is.na(databaru))
0

#Groupby bulan dan tahun berdasarkan jumlah sales
data2 <- databaru %>% mutate(month = format(Order.Date, "%m"), year =
format(Order.Date, "%Y")) %>% group_by(month, year) %>%
summarise(total = sum(Sales))

data2$Date <- as.yearmon(paste(data2$year, data2$month), "%Y %m")
```

```
#Mengurutkan dan Mengganti Nama Kolom

data3 <- subset(data2, select = c("Date", "total"))

data3 <- rename(data3, Total_Sales=total)

data4 <- data3[order(data3$Date),]

#Mengubah data menjadi data runtun waktu

datats <- ts(data4$Total_Sales, start=2015, frequency=12)

datats
```

	A Time Series: 4 x 12											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2015	14205.707	4519.892	55205.797	27906.855	23644.303	34322.936	33781.543	27117.536	81623.527	31453.393	77907.661	68167.058
2016	18066.958	11951.411	32339.318	34154.469	29959.531	23599.374	28608.259	36818.342	63133.606	31011.737	75249.399	74543.601
2017	18542.491	22978.815	51165.059	38679.767	56656.908	39724.486	38320.783	30542.200	69193.391	59583.033	79066.496	95739.121
2018	43476.474	19920.997	58863.413	35541.910	43825.982	48190.728	44825.104	62837.848	86152.888	77448.131	117938.155	83030.389

## 4.2. Identifikasi Stasioneritas Data

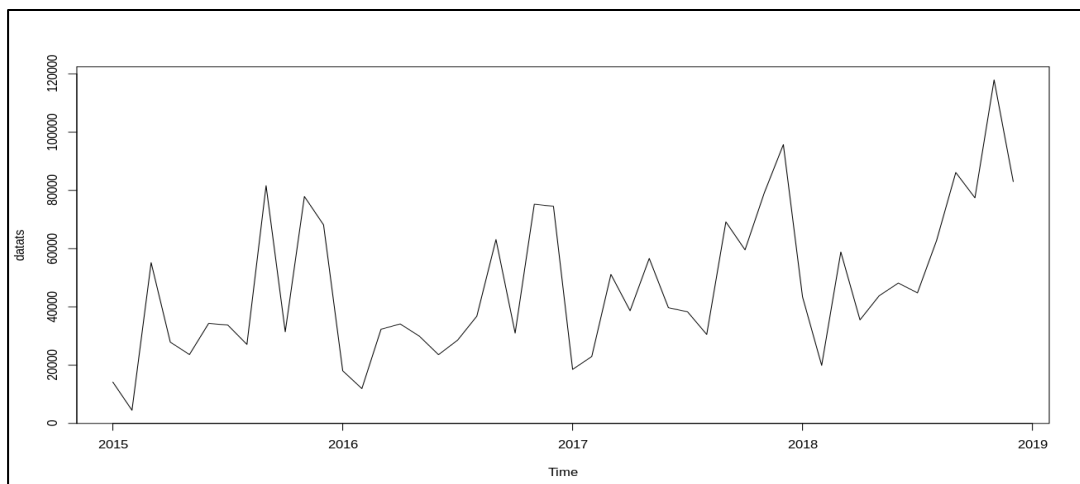
### 4.2.1. Uji Stasioner Data berdasarkan Plot

Akan dilakukan visualisasi dari data runtun waktu dengan plot garis dengan sumbu x sebagai indeks waktunya dan sumbu y sebagai nilai dari data dan pada penelitian kali ini sumbu y nya adalah total penjualan per bulan.

```
options(repr.plot.width = 13, repr.plot.height = 7, repr.plot.res = 100)

plot(datats)
```

Sehingga akan dihasilkan visualisasi plot sebagai berikut :





Dapat dilihat dari gambar sebelumnya bahwa plot tidak terlihat stasioner karena terdapat pola yang naik dan sedikit ada pola berulang. Melihat kestasioneran data berdasarkan plot masih sangat subjektif sehingga akan dilanjutkan dengan melakukan uji stasioner data secara statistik.

#### 4.2.2. Uji Stasioner Data dengan ADF

Akan dilakukan pengujian dengan uji ADF atau uji Augmented Dickey-Fuller untuk menguji kestasioneran data secara statistik. Pada pengujian dengan ADF digunakan  $H_0$  : Data tidak stasioner dan  $H_1$ : Data stasioner.

```
#Pengujian ADF
adf.test(datats)

Augmented Dickey-Fuller Test
data: datats
Dickey-Fuller = -3.3858, Lag order = 3, p-value = 0.06943
alternative hypothesis: stationary
```

Didapatkan statistik uji  $\tau = -3.3858$  dengan  $p - value = 0.06943$ . Dengan menggunakan tingkat signifikansi  $\alpha = 5\% = 0.05$ , maka akan didapatkan  $p - value = 0.06943 > 0.05 = \alpha$ , sehingga  $H_0$  tidak ditolak, yang artinya belum ada bukti untuk mengatakan bahwa data runtun waktunya merupakan data yang stasioner.

#### 4.2.3. Melakukan Differencing Data

Karena data runtun waktu sebelumnya belum stasioner, akan dilakukan differencing satu kali dan diharapkan data yang sudah differencing akan stasioner.

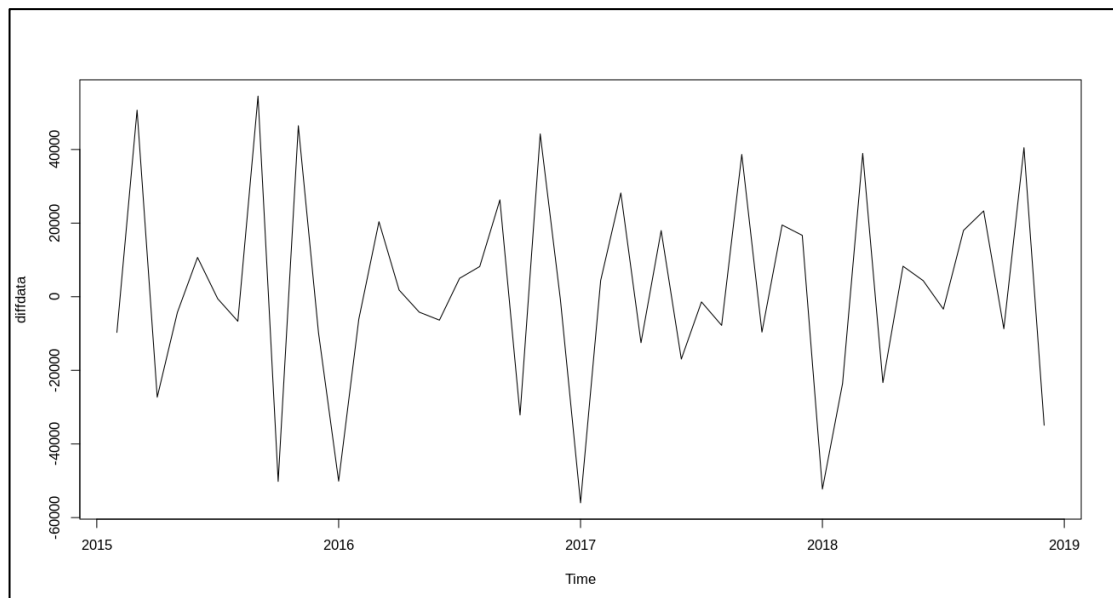
```
#Differencing satu kali
diffdata <- diff(datats, differences=1)
diffdata
```

A Time Series: 4 × 12												
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2015		-9685.8150	50685.9050	-27298.9420	-4262.5520	10678.6326	-541.3926	-6664.0065	54505.9903	-50170.1338	46454.2677	-9740.6022
2016	-50100.1009	-6115.5466	20387.9074	1815.1501	-4194.9380	-6360.1565	5008.8850	8210.0832	26315.2638	-32121.8685	44237.6620	-705.7983
2017	-56001.1102	4436.3240	28186.2440	-12485.2920	17977.1410	-16932.4220	-1403.7030	-7778.5827	38651.1906	-9610.3579	19483.4628	16672.6252
2018	-52262.6470	-23555.4766	38942.4154	-23321.5027	8284.0721	4364.7455	-3365.6237	18012.7440	23315.0400	-8704.7568	40490.0238	-34907.7662

#### 4.2.4. Uji Stasioner Data Differencing berdasarkan Plot

Akan dibuat plot runtun waktu kembali tetapi untuk data yang telah didifferencing.

```
options(repr.plot.width = 13, repr.plot.height = 7, repr.plot.res = 100)
plot(diffdata)
```



#### 4.2.5. Uji Stasioner Data Differencing dengan ADF

Pada pengujian dengan ADF digunakan  $H_0$  : Data tidak stasioner dan  $H_1$ : Data stasioner.

```
#Pengujian ADF Data Runtun Waktu Differencing
adf.test(diffdata)

Augmented Dickey-Fuller Test

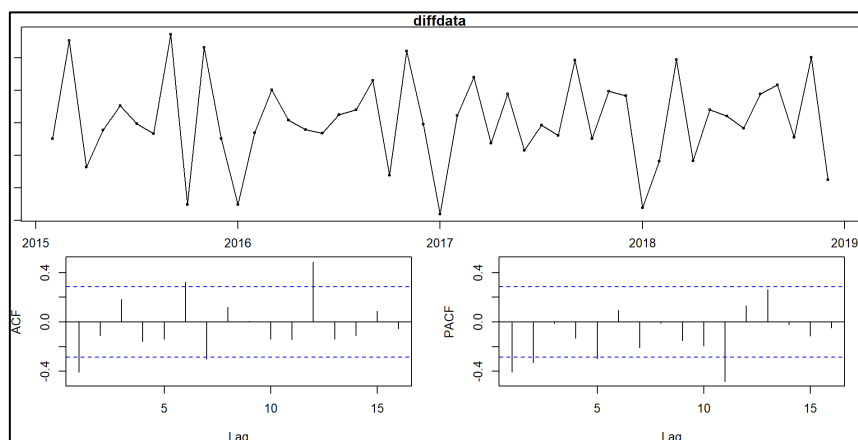
data:  diffdata
Dickey-Fuller = -4.2286, Lag order = 3, p-value = 0.01
alternative hypothesis: stationary
```

Didapatkan statistik uji  $\tau = -4.2286$  dengan  $p - value = 0.01$  . Dengan menggunakan tingkat signifikansi  $\alpha = 5\% = 0.05$  , maka akan didapatkan  $p - value = 0.01 < 0.05 = \alpha$ , sehingga  $H_0$  ditolak, yang artinya dapat dikatakan bahwa data runtun waktunya merupakan data yang stasioner.

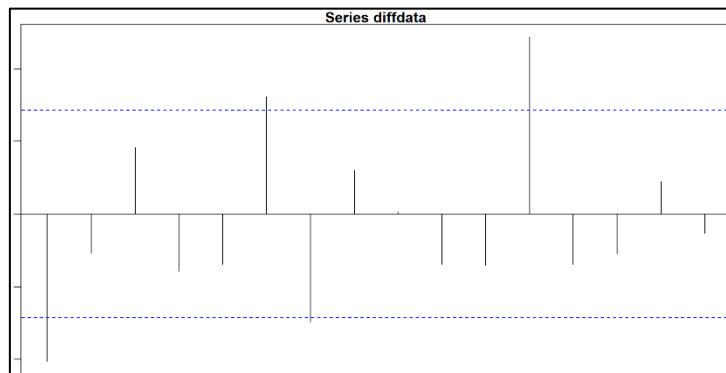
#### 4.3. Spesifikasi Model

Untuk melakukan spesifikasi model dengan menggunakan software R, akan digunakan fungsi `tsdisplay()` atau `acf()` dan `pacf()` untuk data yang sudah di differencing sebagai berikut:

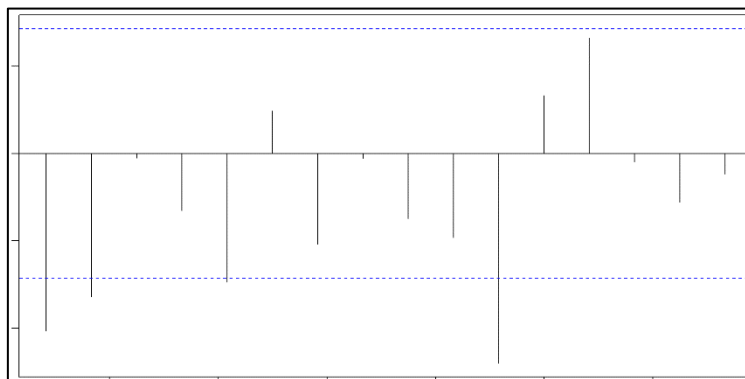
```
tfdisplay(diffdata)
```



```
acf(diffdata)
```



```
pacf(diffdata)
```



Lalu run fungsi `eacf()`:

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

Dari hasil `eacf(diffdata)` akan diambil tiga kandidat model yaitu ARI(2,1), IMA(1,1), dan ARIMA(1,1,1) :

```
#fit model

modell1 <- Arima(datats,order=c(0,1,1),include.constant = TRUE)
modell2 <- Arima(datats,order=c(1,1,1),include.constant = TRUE)
modell3 <- Arima(datats,order=c(2,1,0),include.constant = TRUE)
cbind(modell1,modell2,modell3)
```

modell1	modell2	modell3	
coef	numeric,2	numeric,3	numeric,3
sigma2	490128253	478138948	574104749
var.coef	numeric,4	numeric,9	numeric,9
mask	logical,2	logical,3	logical,3
loglik	-537.843	-536.4951	-539.3085
aic	1081.686	1080.99	1086.617
arma	integer,7	integer,7	integer,7
residuals	ts,48	ts,48	ts,48
call	expression	expression	expression
series	"datats"	"datats"	"datats"
code	0	0	0
n.cond	0	0	0
nobs	47	47	47
model	list,10	list,10	list,10
aicc	1082.244	1081.943	1087.569
bic	1087.236	1088.391	1094.018
xreg	integer,48	integer,48	integer,48
x	ts,48	ts,48	ts,48
fitted	ts,48	ts,48	ts,48

Dari hasil di atas, model ARIMA(1,1,1) memiliki nilai AIC terkecil dengan nilai 1080.99 sementara model IMA(1,1) memiliki nilai BIC terkecil dengan nilai 1087.236. Sehingga akan diambil model terbaik yaitu ARIMA(1,1,1) :

```
fit1 <- Arima(datats,order=c(1,1,1),include.constant = TRUE)
fit1
>
```

```
Series: datats
ARIMA(1,1,1) with drift

Coefficients:
          ar1          ma1          drift
      0.2391    -1.000    910.3310
s.e.  0.1445     0.063    284.6572

sigma^2 = 478138948:  log likelihood = -536.5
AIC=1080.99  AICc=1081.94  BIC=1088.39
```

#### 4.4. Diagnostik Model

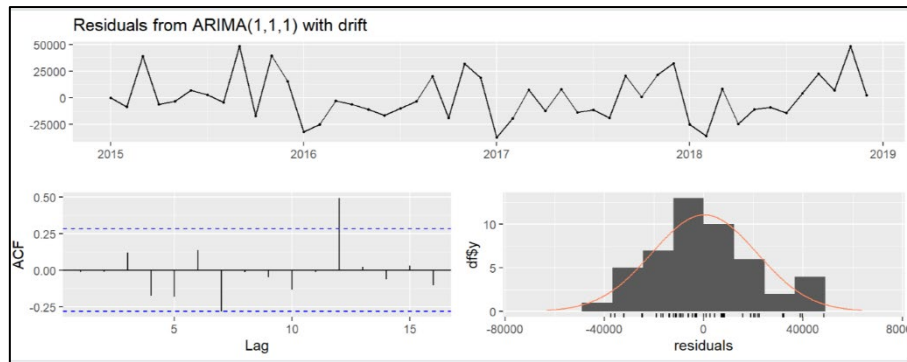
##### 4.4.1. Uji Independensi Ljung Box

Akan dilakukan uji independensi menggunakan Ljung box dengan hipotesis,

$H_0$ : Residual independen

$H_1$ : Residual tidak independen

```
checkresiduals(fit1)
> Ljung-Box test
data:  Residuals from ARIMA(1,1,1) with drift
Q* = 11.332, df = 7, p-value = 0.1248
Model df: 3.    Total lags used: 10
```



Dengan hasil di atas dapat disimpulkan bahwa residual independen.

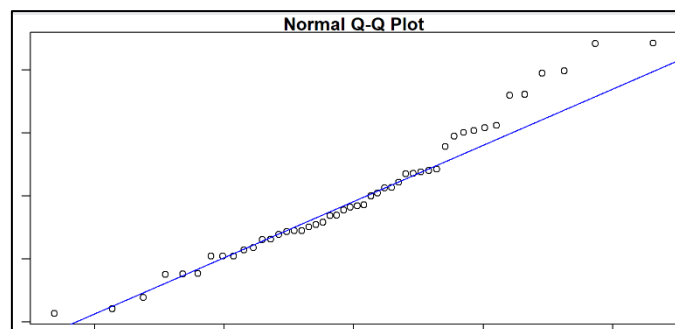
#### 4.4.2. Uji Normalitas Jarque-Bera

Akan dilakukan uji normalitas menggunakan Jarque-Bera dengan hipotesis:

$H_0$ : Data berdistribusi normal

$H_1$ : Tidak demikian

```
qqnorm(fit1$residuals)
qqline(fit1$residuals, col = 'blue')
```



```
jb.norm.test(fit1$residuals)
> Jarque-Bera test for normality
data: fit1$residuals
JB = 2.3387, p-value = 0.164
```

Dari hasil yang didapat dapat disimpulkan bahwa data berdistribusi normal.

#### 4.4.3. Overfitting Model

Untuk model ARIMA(1,1,1) akan dipilih dua overfit model yaitu,

```
Overfit1 <- Arima(datats,order=c(2,1,1),include.constant = TRUE)
Overfit2<- Arima(datats,order=c(1,1,2),include.constant = TRUE)
Cbind(fit1,Overfit1,Overfit2)
```

```
>
      fit1      Overfit1  Overfit2
coef    numeric,3  numeric,4  numeric,4
sigma2  478138948  587308374  487705269
var.coef numeric,9  numeric,16 numeric,16
mask     logical,3  logical,4  logical,4
loglik   -536.4951  -539.3039  -536.5075
aic      1080.99    1088.608    1083.015
arma     integer,7  integer,7  integer,7
residuals ts,48     ts,48     ts,48
call     expression expression expression
series   "datats"   "datats"  "datats"
code     0          0          0
n.cond   0          0          0
nobs     47         47         47
model    list,10    list,10    list,10
aicc     1081.943    1090.071    1084.478
bic      1088.391    1097.859    1092.266
xreg     integer,48 integer,48 integer,48
x        ts,48      ts,48      ts,48
fitted   ts,48      ts,48      ts,48
```

Dari hasil AIC dan BIC yang didapat, model ARIMA(1,1,1) masih yang terbaik.



Maka selanjutnya akan dilakukan uji t dengan hipotesis:

$H_0$ : Parameter = 0

$H_1$ : Tidak demikian

```
uji <- Overfit1$coef[['ar2']]/0.2521
daerah_kritis <- qt(0.025, (length(datats)-1))
daerah_kritis; uji
>
[1] -2.011741
[1] -1.446028
```

```
#Overfit2
uji2 <- Overfit2$coef[['ma2']]/0.6465
daerah_kritis2 <- qt(0.025, (length(datats)-1))
daerah_kritis2; uji2
>
[1] -2.011741
[1] -0.6068817
```

Maka dari hasil yang didapat, diketahui bahwa parameter AR2 dan MA2 adalah 0.

#### 4.5. Analisis Model

Dari perhitungan yang sudah dilakukan, model ARIMA(1,1,1) adalah model yang terbaik untuk data yang digunakan. Model ARIMA(1,1,1) pada data memiliki nilai AIC dan BIC yang lebih rendah dibandingkan model yang lain. Selain itu model ARIMA(1,1,1) memiliki residual yang independen juga data yang berdistribusi normal. Pada overfitting model juga, ARIMA(1,1,1) memiliki nilai AIC dan BIC yang lebih baik dan setelah melakukan uji t parameter dari MA2 dan AR2 tidak signifikan dalam model (parameter AR2 dan MA2 = 0).

Diketahui bahwa proses umum ARIMA (1,1,1)

$$\Delta Y_t = \mu + \phi \eta_t$$

Dengan

$$\eta_t = \phi_1 \eta_{t-1} + e_t + \phi_1 e_{t-1}$$

Dengan menyubstitusi  $Y_t - Y_{t-1} - \mu = \eta_t$  didapatkan

$$\begin{aligned} Y_t - Y_{t-1} - \mu &= \phi_1(Y_{t-1} - Y_{t-2} - \mu) + e_t + \phi_1 e_{t-1} \\ Y_t &= (1 - \phi_1)\mu + (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + e_t + \phi_1 e_{t-1} \end{aligned}$$

Sehingga didapat

$$\begin{aligned} Y_t &= (1 - \widehat{\phi}_1)\hat{\mu} + (1 + \widehat{\phi}_1)Y_{t-1} - \widehat{\phi}_1 Y_{t-2} + e_t + \widehat{\theta}_1 e_{t-1} \\ Y_t &= (1 - 0.2391)(910.331) + (1 + 0.2391)Y_{t-1} - (0.2391)Y_{t-2} + e_t + (-1)e_{t-1} \\ Y_t &= 692.67 + (1.2391)Y_{t-1} - (0.2391)Y_{t-2} + e_t - e_{t-1} \end{aligned}$$

Sehingga didapatkan persamaan peramalan untuk 1 bulan ke depan yaitu

$$\widehat{Y}_t(1) = E[Y_{t+1}|\Omega_t] = 692.67 + (1.2391)Y_t - (0.2391)Y_{t-1} - e_t$$

Persamaan peramalan untuk 2 bulan ke depan yaitu

$$\widehat{Y}_t(2) = E[Y_{t+2}|\Omega_t] = 692.67 + (1.2391)E[Y_{t+1}|\Omega_t] - (0.2391)Y_t$$

Dan persamaan peramalan untuk  $l > 2$  bulan ke depan yaitu

$$\widehat{Y}_t(l) = E[Y_{t+l}|\Omega_t] = 692.67 + (1.2391)E[Y_{t+l-1}|\Omega_t] - (0.2391)E[Y_{t+l-2}|\Omega_t]$$

Dengan  $\Omega_t$  adalah  $Y_1, Y_2, \dots, Y_t$ .

#### 4.6. Forecasting (Peramalan)

Akan dilakukan *Cross-Validation* dengan membagi data menjadi 2 yaitu data *train* dan data *test*.

##### 1. Membagi data train dari Januari 2015 hingga Desember 2017

```
#Membagi data train dari Januari 2015 hingga Desember 2017  
train <- window(datats, end=c(2018,0))  
Train
```

A Time Series: 3 x 12												
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2015	14205.707	4519.892	55205.797	27906.855	23644.303	34322.936	33781.543	27117.536	81623.527	31453.393	77907.661	68167.058
2016	18066.958	11951.411	32339.318	34154.469	29959.531	23599.374	28608.259	36818.342	63133.606	31011.737	75249.399	74543.601
2017	18542.491	22978.815	51165.059	38679.767	56656.908	39724.486	38320.783	30542.200	69193.391	59583.033	79066.496	95739.121

##### 2. Membagi data test dari Januari 2018 hingga Desember 2018

```
test <- window(datats, start=c(2018,1))  
Test
```

A Time Series: 1 x 12												
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2018	43476.47	19921.00	58863.41	35541.91	43825.98	48190.73	44825.10	62837.85	86152.89	77448.13	117938.15	83030.39

##### 3. Melakukan forecasting untuk data train

```
train_fit <- Arima(train,order=c(1,1,1),include.constant = TRUE)  
forecast_train <- forecast(train_fit,12)  
train  
forecast train
```

A Time Series: 3 x 12												
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2015	14205.707	4519.892	55205.797	27906.855	23644.303	34322.936	33781.543	27117.536	81623.527	31453.393	77907.661	68167.058
2016	18066.958	11951.411	32339.318	34154.469	29959.531	23599.374	28608.259	36818.342	63133.606	31011.737	75249.399	74543.601
2017	18542.491	22978.815	51165.059	38679.767	56656.908	39724.486	38320.783	30542.200	69193.391	59583.033	79066.496	95739.121
Point Forecast												
Jan 2018	67412.40	39231.87	95592.92	24314.00	110510.8							
Feb 2018	62959.27	34164.20	91754.34	18921.02	106997.5							
Mar 2018	62921.96	34080.94	91762.98	18813.42	107030.5							
Apr 2018	63701.43	34854.18	92548.68	19583.37	107819.5							
May 2018	64631.98	35783.66	93480.30	20512.28	108751.7							
Jun 2018	65590.47	36741.95	94438.98	21470.47	109710.5							
Jul 2018	66554.12	37705.57	95402.68	22434.07	110674.2							
Aug 2018	67518.74	38670.18	96367.30	23398.67	111638.8							
Sep 2018	68483.53	39634.97	97332.09	24363.46	112603.6							
Oct 2018	69448.36	40599.79	98296.92	25328.28	113568.4							
Nov 2018	70413.19	41564.62	99261.75	26293.11	114533.3							
Dec 2018	71378.02	42529.45	100226.58	27257.94	115498.1							

#### 4. Menggabungkan data forecast train dengan data test

```
cbind(test, forecast_train)
```

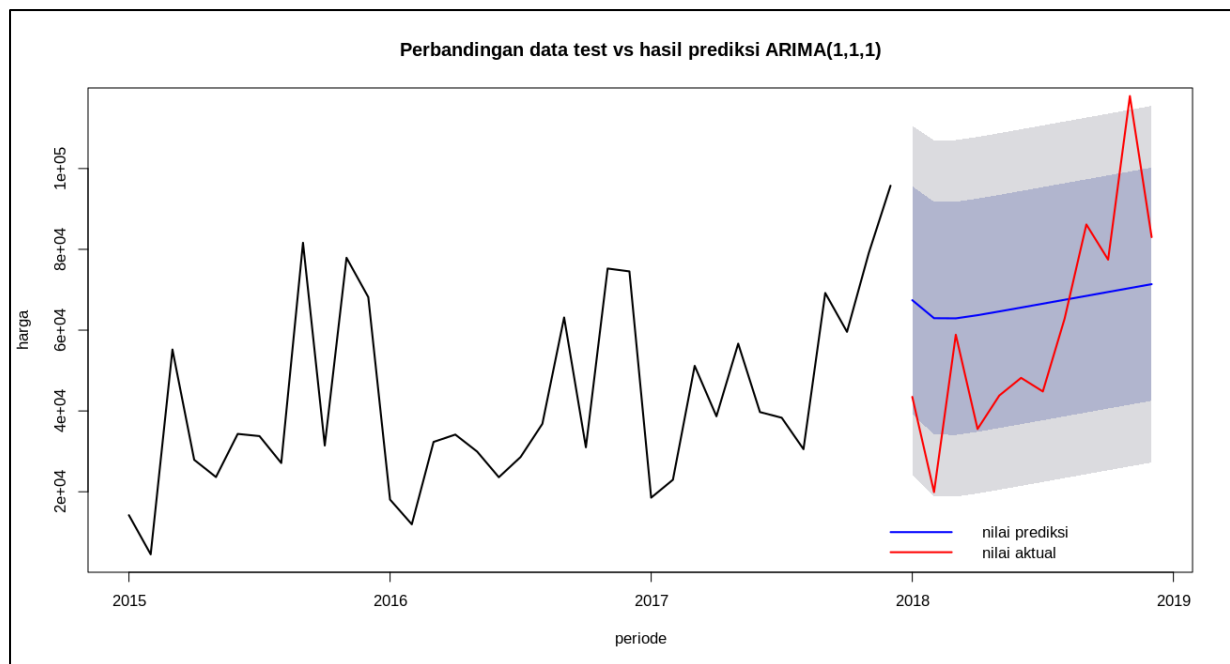
✕

A Time Series: 12 × 6

	test	forecast_train.Point Forecast	forecast_train.Lo 80	forecast_train.Hi 80	forecast_train.Lo 95	forecast_train.Hi 95
Jan 2018	43476.47	67412.40	39231.87	95592.92	24314.00	110510.8
Feb 2018	19921.00	62959.27	34164.20	91754.34	18921.02	106997.5
Mar 2018	58863.41	62921.96	34080.94	91762.98	18813.42	107030.5
Apr 2018	35541.91	63701.43	34854.18	92548.68	19583.37	107819.5
May 2018	43825.98	64631.98	35783.66	93480.30	20512.28	108751.7
Jun 2018	48190.73	65590.47	36741.95	94438.98	21470.47	109710.5
Jul 2018	44825.10	66554.12	37705.57	95402.68	22434.07	110674.2
Aug 2018	62837.85	67518.74	38670.18	96367.30	23398.67	111638.8
Sep 2018	86152.89	68483.53	39634.97	97332.09	24363.46	112603.6
Oct 2018	77448.13	69448.36	40599.79	98296.92	25328.28	113568.4
Nov 2018	117938.15	70413.19	41564.62	99261.75	26293.11	114533.3
Dec 2018	83030.39	71378.02	42529.45	100226.58	27257.94	115498.1

#### 5. Membuat visualisasi forecast bersama data actualnya

```
plot(forecast_train,
     fcol="blue",lwd=2,
     main = 'Perbandingan data test vs hasil prediksi ARIMA(1,1,1)',
     xlab = 'periode',
     ylab = 'harga')
lines(test,
     col = 'red',
     lwd = 2)
legend('bottomright',
     col=c('blue','red'),
     legend = c('nilai prediksi','nilai aktual'),
     lwd=2,
     bty='n')
```



Hasil forecast tentunya tidak sempurna tetapi memiliki hasil yang cukup untuk dapat digunakan untuk forecasting selanjutnya tentu dengan syarat untuk beberapa periode awal.

## 6. Mencari rata-rata

```
#Mencari rata-rata
mean(abs(test-forecast_train$mean))
>
20721.198584364
```

Didapatkan mean senilai 20721.198584364

## 7. Lalu akan dilakukan forecasting untuk memprediksi 6 periode selanjutnya.

```
#Forecast final
fit #pake semua data
forecast_final <- forecast(fit,h=6)
forecast_final
```

```

>
#Forecast final
Series: datats
ARIMA(1,1,1) with drift

Coefficients:
          ar1          ma1          drift
      0.2391    -1.000    910.3310
s.e.    0.1445     0.063    284.6572

sigma^2 = 478138948:  log likelihood = -536.5
AIC=1080.99  AICc=1081.94  BIC=1088.39

```

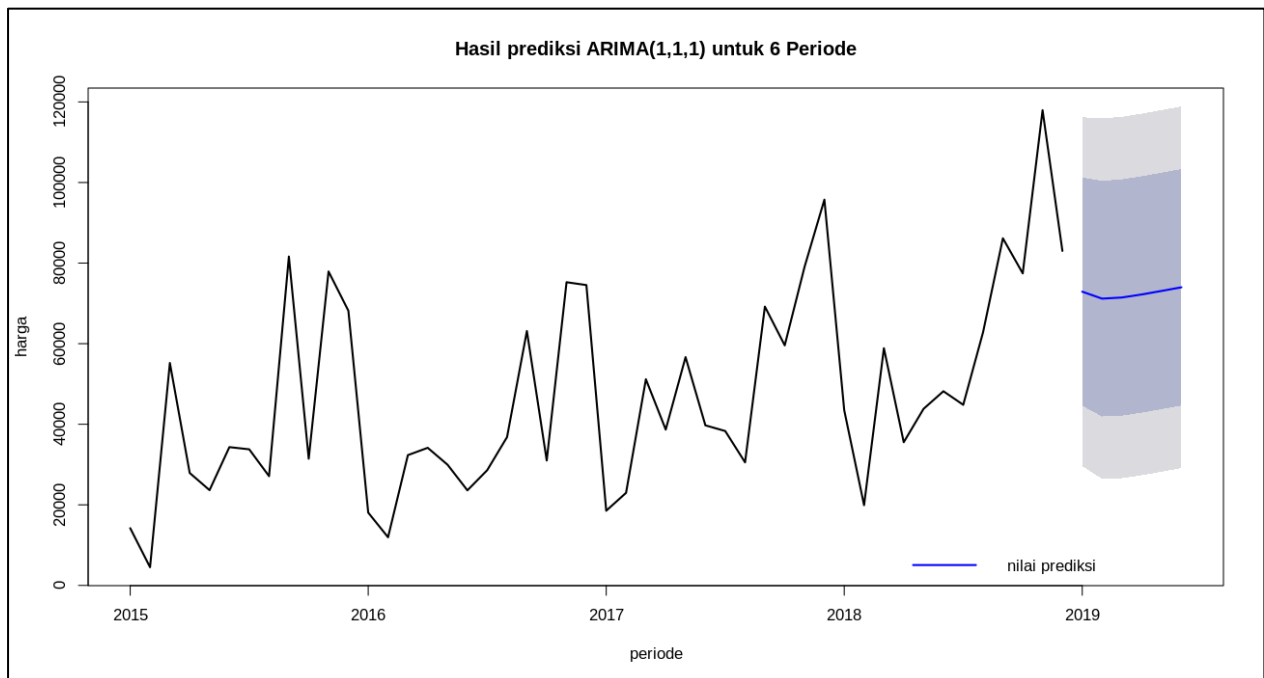
	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2019	72904.86	44595.58	101214.1	29609.55	116200.2
Feb 2019	71177.01	41937.74	100416.3	26459.40	115894.6
Mar 2019	71456.67	42133.65	100779.7	26610.99	116302.3
Apr 2019	72216.23	42880.92	101551.5	27351.74	117080.7
May 2019	73090.52	43752.71	102428.3	28222.21	117958.8
Jun 2019	73992.24	44653.85	103330.6	29123.05	118861.4

Di atas didapatkan hasil forecasting untuk 6 periode berikutnya.

## 8. Membuat visualisasi forecasting untuk data final.

```
#Visualisasi forecast data final

plot(forecast_final,
     fcol="blue",lwd=2,
     main = 'Hasil prediksi ARIMA(1,1,1) untuk 6 Periode',
     xlab = 'periode',
     ylab = 'harga')
legend('bottomright',
     col=c('blue'),
     legend = c('nilai prediksi'),
     lwd=2,
     bty='n')
```



## BAB V

### KESIMPULAN

#### 5.1. Kesimpulan

Berdasarkan pembahasan di atas, didapatkan kesimpulan sebagai berikut :

1. Berdasarkan analisis data menggunakan metode ARIMA Box-Jenkins, model terbaik untuk meramalkan jumlah penjualan adalah model ARIMA(1,1,1)
2. Model runtun waktu yang terbaik untuk data runtun waktu penjualan toko global adalah

$$Y_t = 692.67 + (1.2391)Y_{t-1} - (0.2391)Y_{t-2} + e_t - e_{t-1}$$

3. Berdasarkan hasil forecasting, penjualan toko global akan meningkat ke depannya.

#### 5.2. Saran

Berdasarkan kesimpulan di atas, penulis dapat memberikan saran:

1. Peneliti selanjutnya yang ingin membuat peramalan untuk data ini untuk sebaiknya menggunakan metode peramalan lainnya selain ARIMA yang memungkinkan memberikan hasil yang lebih baik.
2. Meskipun hasil dari model yang dipilih terasa sudah cukup memuaskan, ini tidak mengurangi kemungkinan bahwa penulis dapat saja memilih model berdasarkan kesubjektifan sehingga untuk peneliti selanjutnya dianjurkan untuk menggunakan model lain melalui metode lain.



## DAFTAR PUSTAKA

Cryer, Jonathan D. dan Kung-Sik Chan. (2002). *Time Series Analysis: With Applications in R* Second Edition. New York: Springer Science Business Media

Jay Heizer & Barry Render (2015). “Manajemen Operasi Edisi Ke-11. Jakarta: Salemba Empat”

Munaroh, Siti (2010). “Analisis Model ARIMA Box-Jenkins Pada Data Fluktuasi Harga Emas”. Malang: UIN Malang.

Rohit Sahoo (2020). “Time Series Forecasting Superstore Sales Dataset”  
<https://www.kaggle.com/datasets/rohitsahoo/sales-forecasting>

Utama, Made Suyana & Wirawan, I Gusti Putu Nata (2014). “Model *Box-Jenkins* Dalam Rangka Peramalan Produk Domestik Regional Bruto Provinsi Bali”. *Jurnal Buletin Studi Ekonomi*, Vol. 19, No