

**ANALISIS RUNTUN WAKTU MODEL ARIMA PADA DATA TEMPERATUR  
HARIAN KOTA WARSAW PADA PERIODE 1 JUNI 2022 - 31 DESEMBER  
2022**



Disusun Oleh:

Kelompok H

Rahmi Radhia Khalqi (1906375796)

Audrey Clarissa (2106724864)

Favian Sulthan Wafi (2106706205)

Fernaldy (2106706464)

**PROGRAM STUDI STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS INDONESIA  
DEPOK  
2023**

## KATA PENGANTAR

Puji dan syukur penulis ucapkan kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan laporan akhir yang berjudul “Analisis Runtun Waktu Model ARIMA pada Data Temperatur Harian Kota Warsaw Periode 1 Juni 2022 - 31 Desember 2022”. Penyusunan laporan akhir ini dilakukan dalam rangka memenuhi salah satu penilaian pada mata kuliah Metode Peramalan, program studi S1 Statistika TA 2022/2023.

Laporan akhir ini telah disusun dengan dedikasi penuh dan mendapatkan dukungan dari berbagai pihak yang berperan penting dalam kelancaran pembuatannya. Oleh karena itu, penulis ingin mengucapkan terima kasih banyak kepada semua pihak yang telah berkontribusi dalam penulisan laporan akhir ini, khususnya kepada Ibu Yekti Widyaningsih, selaku dosen pengampu mata kuliah Metode Peramalan.

Penulis dengan sepenuhnya menyadari bahwa laporan akhir ini masih memiliki kekurangan dan jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan masukan, kritik, dan saran dari para pembaca guna meningkatkan kualitas laporan akhir ini. Jika terdapat banyak kesalahan dalam laporan akhir ini, penulis dengan tulus meminta maaf yang sebesar-besarnya. Meskipun demikian, penulis berharap laporan akhir ini membawa manfaat baik bagi penulis maupun pembaca.

Depok, 17 Juni 2023

Tim Penulis

## DAFTAR ISI

KATA PENGANTAR.....	ii
DAFTAR ISI.....	iii
TABEL KONTRIBUSI.....	v
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah .....	2
1.3. Tujuan Penelitian .....	2
1.4. Manfaat Penelitian.....	2
BAB II LANDASAN TEORI .....	3
2.1. Peramalan.....	3
2.2. Konsep Dasar <i>Time Series</i> .....	3
2.3. Model <i>Autoregressive Integrated Moving Average</i> (ARIMA).....	5
2.3.1. Model <i>Autoregressive</i> (AR).....	5
2.3.2. Model <i>Moving Average</i> (MA).....	5
2.3.3. Model <i>Autoregressive Moving Average</i> (ARMA).....	6
2.3.4. Model <i>Autoregressive Integrated Moving Average</i> (ARIMA) .....	6
2.4. Stasioneritas dan Nonstasioneritas.....	7
2.4.1. Korelogram.....	8
2.4.2. Uji Unit Akar ( <i>Unit Root</i> ) .....	10
2.5. Spesifikasi Model .....	11
2.5.1. <i>Autocorrelation Function</i> (ACF) .....	11
2.5.2. <i>Partial Autocorrelation Function</i> (PACF) .....	11
2.5.3. <i>Extended Autocorrelation Function</i> (EACF) .....	12
2.5.4. <i>Akaike Information Criterion</i> (AIC) .....	13
2.5.5. <i>Bayesian Information Criterion</i> (BIC).....	13
2.6. Estimasi Parameter.....	13
2.7. <i>Diagnostic Checking</i> .....	14
2.7.1. Uji Independensi Ljung-Box .....	14
2.7.2. Uji Normalitas Jarque-Bera.....	15
2.7.3. <i>Overfitting</i> Model.....	16
2.7.4. <i>Cross Validation</i> .....	16
2.8. <i>Forecasting</i> .....	17
BAB III METODE PENELITIAN .....	18

3.1.	<b>Variabel Penelitian</b> .....	18
3.2.	<b>Data</b> .....	18
3.3.	<b>Metode Pengumpulan Data</b> .....	18
3.4.	<b>Metode Analisis Data</b> .....	19
<b>BAB IV ANALISIS DATA</b> .....		21
4.1.	<i>Preprocessing Data</i> .....	21
4.2.	<b>Membentuk Data <i>Time-Series</i></b> .....	23
4.3.	<b>Uji Stasioneritas Data</b> .....	23
4.4.	<b>Spesifikasi Model</b> .....	25
4.5.	<b>Estimasi Parameter</b> .....	27
4.6.	<b>Diagnosis Model</b> .....	28
4.6.1.	<b>Analisis Residual</b> .....	28
4.6.2.	<i>Overfitting Model</i> .....	32
4.7.	<i>Forecasting</i> .....	35
<b>BAB V PENUTUP</b> .....		39
5.1.	<b>Kesimpulan</b> .....	39

**TABEL KONTRIBUSI**

Anggota	Jenis Kontribusi	Persentase Kontribusi
Rahmi Radhia Khalqi (1906375796)	Aktif terlibat dalam diskusi. Menulis pendahuluan, sebagian landasan teori, menulis metode penelitian, merapihkan penulisan.	100%
Audrey Clarissa (2106724864)	Aktif terlibat dalam diskusi dan membuat landasan teori.	100%
Favian Sulthan Wafi (2106706205)	Aktif terlibat dalam diskusi. Mencari data, melakukan analisis data dan pembahasan dari <i>preprocessing</i> data sampai spesifikasi model, dan membuat kesimpulan.	100%
Fernaldy (2106706464)	Aktif terlibat dalam diskusi, melakukan analisis data, menambahkan bagian landasan teori, serta membuat kesimpulan.	100%

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Warsaw merupakan ibu kota dan kota terbesar di Polandia. Terletak di tepi sungai Wisła, Warsaw merupakan pusat politik, ekonomi, dan budaya negara tersebut. Kota ini memiliki sejarah yang kaya dengan berbagai pengaruh budaya, arsitektur, dan warisan sejarah yang signifikan. Warsaw merupakan pusat pemerintahan, dengan markas besar institusi nasional dan kantor-kantor pemerintah yang penting. Selain itu, kota ini juga memiliki sektor ekonomi yang maju, terutama dalam bidang keuangan, teknologi informasi, periklanan, perdagangan, dan pariwisata. Warsaw menjadi tujuan utama bagi wisatawan yang tertarik dengan sejarah, seni, dan kehidupan malam yang beragam.

Temperatur udara merupakan suatu data deret waktu (*time series*) karena didapat dari pengamatan dari waktu ke waktu. Temperatur udara harian merupakan faktor penting yang mempengaruhi kehidupan sehari-hari masyarakat, aktivitas pertanian, kesehatan, dan pengambilan keputusan terkait energi, transportasi, serta pengelolaan lingkungan. Kota Warsaw memiliki lingkungan yang menarik karena memiliki variasi temperatur udara yang signifikan sepanjang tahun dan pengaruh iklim yang bervariasi. Hal ini membuatnya menjadi lokasi yang menarik untuk melakukan analisis dan peramalan temperatur udara harian guna memahami pola cuaca dan perubahan iklim di kawasan tersebut.

Dengan memiliki pemahaman yang lebih mendalam tentang pola dan tren temperatur udara harian di Kota Warsaw, dapat membantu dalam perencanaan dan pengambilan keputusan yang lebih efektif. Informasi ramalan temperatur yang akurat juga dapat membantu dalam pengelolaan energi, sistem pendingin dan pemanasan, penyesuaian kegiatan *outdoor*, dan perencanaan kegiatan berbasis musim. Penelitian ini bertujuan untuk mengembangkan model peramalan (*forecasting*) yang dapat mengestimasi temperatur udara harian di Kota Warsaw. Dengan memperoleh prediksi temperatur udara harian yang tepat, pihak berkepentingan seperti meteorolog, penyedia layanan cuaca, peneliti lingkungan, serta sektor-sektor terkait lainnya dapat memperoleh informasi yang berharga untuk perencanaan, pengambilan keputusan, dan mitigasi terhadap perubahan iklim.

## **1.2. Rumusan Masalah**

Rumusan masalah pada penelitian ini adalah sebagai berikut.

1. Bagaimana proses analisis data temperatur harian Kota Warsaw menggunakan model runtun waktu?
2. Model runtun waktu apakah yang tepat digunakan untuk melakukan *forecasting* data temperatur harian Kota Warsaw?
3. Bagaimana penerapan metode peramalan ARIMA untuk memprediksi temperatur harian Kota Warsaw untuk periode 1 Januari 2023 – 3 Januari 2023

## **1.3. Tujuan Penelitian**

Tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut.

1. Mengetahui proses analisis data temperatur harian Kota Warsaw menggunakan model runtun waktu.
2. Mengetahui model runtun waktu apakah yang tepat digunakan untuk melakukan *forecasting* data temperatur harian Kota Warsaw.
3. Mengetahui prediksi temperatur harian Kota Warsaw untuk periode 1 Januari 2023 – 3 Januari 2023 melalui penerapan metode peramalan ARIMA.

## **1.4. Manfaat Penelitian**

Manfaat dari penelitian ini adalah sebagai berikut.

1. Memberikan informasi mengenai proses pengolahan data temperatur harian Kota Warsaw hingga memperoleh model runtun waktu yang sesuai untuk melakukan *forecasting* masa mendatang.
2. Memberikan informasi mengenai model runtun waktu terbaik untuk data temperatur harian Kota Warsaw.
3. Memberikan informasi mengenai prediksi temperatur harian Kota Warsaw pada periode 1 Januari 2023 – 3 Januari 2023.

## **BAB II**

### **LANDASAN TEORI**

Pada Bab II, akan dijelaskan pengertian-pengertian dasar yang digunakan sebagai landasan pembahasan pada bab selanjutnya, yaitu peramalan data runtun waktu (*time series*), konsep dasar *time series*, stasioneritas, dan nonstasioneritas. *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF), *white noise*, dan model-model ARIMA.

#### **2.1. Peramalan**

Peramalan pada dasarnya merupakan proses menyusun informasi tentang kejadian masa lampau yang berurutan untuk menduga kejadian di masa depan. Peramalan bertujuan mendapatkan ramalan yang dapat meminimumkan kesalahan meramal yang dapat diukur dengan *Mean Absolute Percent Error* (MAPE). Peramalan pada umumnya digunakan untuk memprediksi sesuatu yang kemungkinan besar akan terjadi misalnya kondisi permintaan, banyaknya curah hujan, kondisi ekonomi, dan lain-lain.

Atas dasar logika, langkah dalam metode peramalan secara umum adalah mengumpulkan data, menyeleksi dan memilih data, memilih model peramalan, menggunakan model terpilih untuk melakukan peramalan, evaluasi hasil akhir. Berdasarkan sifatnya, peramalan dibedakan menjadi:

##### **1. Peramalan Kualitatif**

Peramalan yang didasarkan atas data kualitatif pada masa lalu. Hasil peramalan kualitatif didasarkan pada pengamatan kejadian-kejadian di masa sebelumnya digabung dengan pemikiran dari penyusunnya.

##### **2. Peramalan Kuantitatif**

Peramalan yang didasarkan atas data kuantitatif masa lalu yang diperoleh dari pengamatan nilai-nilai sebelumnya. Hasil peramalan yang dibuat tergantung pada metode yang digunakan, menggunakan metode yang berbeda akan diperoleh hasil peramalan yang berbeda.

#### **2.2. Konsep Dasar *Time Series***

*Time series* adalah suatu rangkaian atau seri dari nilai-nilai suatu variabel atau hasil observasi, dalam hal ini adalah temperatur udara harian yang dicatat dalam jangka waktu yang berurutan. Metode *time series* adalah metode peramalan dengan menggunakan



analisis pola hubungan antara variabel yang akan diperkirakan dengan variabel waktu atau analisis *time series*, antara lain:

1. Metode *Smoothing*
2. Metode Box–Jenkins (ARIMA)
3. Metode Proyeksi *Trend* dengan Regresi

Hal yang perlu diperhatikan dalam melakukan peramalan adalah pada galat (*error*), yang tidak dapat dipisahkan dalam metode peramalan. Untuk mendapatkan hasil yang mendekati data asli, maka seorang peramal berusaha membuat *error*-nya sekecil mungkin.

Dengan adanya data *time series*, maka pola gerakan data dapat diketahui. Dengan demikian, data *time series* dapat dijadikan sebagai dasar untuk:

- a. Pembuatan keputusan pada saat ini
- b. Peramalan keadaan perdagangan dan ekonomi pada masa yang akan datang
- c. Perencanaan kegiatan untuk masa depan

Analisis data *time series* adalah analisis yang menerangkan dan mengukur berbagai perubahan atau perkembangan data selama satu periode. Analisis *time series* dilakukan untuk memperoleh pola data *time series* dengan menggunakan data masa lalu yang akan digunakan untuk meramalkan suatu nilai pada masa yang akan datang. Dalam *time series* terdapat 4 macam tipe pola data, yaitu sebagai berikut.

1) Horizontal

Tipe data horizontal ialah data yang memiliki observasi berubah-ubah di sekitar tingkatan atau rata-rata yang konstan. Sebagai contoh penjualan tiap bulan suatu produk tidak meningkat atau menurun secara konsisten pada suatu waktu.

2) Musiman (*Seasonal*)

Tipe data *seasonal* ialah data yang memiliki observasi dipengaruhi oleh musiman, yang ditandai dengan adanya pola perubahan yang berulang secara otomatis dari tahun ke tahun. Salah satu contohnya adalah pola data pembelian buku baru pada tahun ajaran baru.

3) *Trend*

Tipe data *trend* ialah data yang memiliki observasi naik atau menurun pada perluasan periode suatu waktu. Contoh data tipe *trend* adalah data populasi.

4) *Cyclical*

Tipe data *cyclical* ditandai dengan adanya fluktuasi bergelombang data yang terjadi di sekitar garis *trend*. Contoh data tipe *cyclical* adalah data-data pada kegiatan ekonomi dan bisnis.

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

Beberapa model ARIMA yang dapat digunakan pada data *time series* yaitu sebagai berikut.

### 2.3.1. Model Autoregressive (AR)

*Autoregressive* adalah suatu bentuk regresi tetapi bukan yang menghubungkan variabel tak bebas, melainkan menghubungkan nilai-nilai sebelumnya pada *time lag* (selang waktu) yang bermacam-macam. Jadi suatu model *autoregressive* akan menyatakan suatu ramalan sebagai fungsi nilai-nilai sebelumnya dari *time series* tertentu.

Model *Autoregressive* (AR) dengan order  $p$  dinotasikan dengan AR(p). Bentuk umum model AR(p) adalah:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

dengan,

$X_t$	: nilai variabel pada waktu ke- $t$
$X_{t-1}, X_{t-2}, \dots, X_{t-p}$	: nilai masa lalu dari <i>time series</i> yang bersangkutan pada waktu $t-1, t-2, \dots, t-p$
$\phi_i$	: koefisien regresi, $i: 1, 2, 3, \dots, p$
$\varepsilon_t$	: nilai <i>error</i> pada waktu ke- $t$
$p$	: order <i>AR</i>

Pada umumnya, order AR yang sering digunakan dalam analisis *time series* adalah  $p = 1$  atau  $p = 2$ , yaitu model *AR* (1) dan *AR* (2).

### 2.3.2. Model Moving Average (MA)

Model *Moving Average* dengan order  $q$ , dinotasikan MA( $q$ ), didefinisikan sebagai berikut.

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} ; \varepsilon_t \sim N(0, \sigma_t^2)$$

dengan,

$X_t$	: nilai variabel pada waktu ke- $t$
$\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$	: nilai-nilai dari <i>error</i> pada waktu $t, t-1, t-2, \dots, t-q$ dan $\varepsilon_t$ diasumsikan <i>White Noise</i> dan normal.
$\theta_i$	: koefisien regresi, $i: 1, 2, 3, \dots, q$
$\varepsilon_t$	: nilai <i>error</i> pada waktu ke- $t$
$q$	: order <i>MA</i>

Secara umum, order MA, yang sering digunakan dalam analisis *time series* adalah  $q = 1$  atau  $q = 2$ , yaitu MA(1) dan MA(2).

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

Model *Autoregressive Moving Average* (ARMA) merupakan suatu kombinasi dari model AR dan MA. Bentuk umum model ARMA sebagai berikut.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

dengan,

$X_t$	: nilai variabel pada waktu ke- $t$
$\phi_i$	: koefisien regresi ke- $i$ , $i = 1, 2, 3, \dots, p$
$p$	: order <i>AR</i>
$\theta_i$	: parameter model <i>MA</i> ke- $i$ , $i = 1, 2, 3, \dots, q$
$\varepsilon_t$	: nilai <i>error</i> pada waktu ke- $t$
$\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$	: <i>error</i> pada saat $t, t-1, t-2, \dots, t-q$ dan $\varepsilon_t$ diasumsikan <i>White Noise</i> dan normal.

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

ARIMA merupakan model ARMA yang telah melalui *differencing* sebanyak  $d$  kali yang dinyatakan dengan ARIMA(p,d,q). Secara umum, model ARIMA(p,d,q) untuk suatu data *time series* adalah sebagai berikut.

$$\begin{aligned} (1 - B)^d X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} \\ = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \end{aligned}$$

dengan,

$X_t$	: data observasi ke- $t$
$B$	: operator <i>back shift</i>
$(1 - B)^d X_t$	: <i>time series</i> yang stasioner pada pembedaan ke- $d$
$\varepsilon_t$	: nilai <i>error</i> pada waktu ke- $t$
$p$	: order <i>AR</i>
$d$	: order pembedaan
$q$	: order <i>MA</i>

Untuk  $p = 0$ , model ARIMA(0,d,q) disebut juga model *Integrated Moving Average* yang dinotasikan dengan IMA(d,q). Sementara itu, untuk  $q = 0$ , model ARIMA(p,d,0) disebut juga model *Autoregressive Integrated* yang dinotasikan dengan ARI(p,q). Apabila *differencing* pertama dilakukan terhadap model agar menjadi stasioner, maka model menjadi ARIMA (1,1,1) didefinisikan sebagai berikut.

$$(1 + \theta_1 B) \varepsilon_t$$

## 2.4. Stasioneritas dan Nonstasioneritas

Stasioneritas berarti bahwa tidak terdapat perubahan yang drastis pada data. Fluktuasi data berada disekitar suatu nilai rata-rata yang konstan, tidak tergantung pada waktu dan variansi dari fluktuasi tersebut. Data *time series* dikatakan stasioner jika rata-rata dan variansinya konstan, tidak ada unsur trend dalam data, dan tidak ada unsur musiman.

Apabila data tidak stasioner, maka perlu dilakukan modifikasi untuk menghasilkan data yang stasioner. Salah satu cara yang umum dipakai adalah metode pembedaan (*differencing*). Untuk menentukan apakah series stasioner, nonstasioner dapat dibantu dengan melihat plot dari *series* atau bentuk *difference*-nya. Proses *differencing* dapat dilakukan untuk beberapa periode sampai data stasioner, yaitu dengan cara mengurangi suatu data dengan data sebelumnya.

Apabila suatu *time series* tidak stasioner, maka data tersebut dapat dibuat lebih mendekati stasioner dengan melakukan pembedaan pertama.

$$\begin{aligned}X'_t &= X_t - X_{t-1} \\X'_t &= X_t - BX_t = (1 - B)X_t\end{aligned}$$

Sama halnya pada *differencing* orde kedua, yaitu *differencing* pertama dari *differencing* pertama sebelumnya.

$$\begin{aligned}X''_t &= X'_t - X'_{t-1} \\&= X_t - X_{t-1} - (X_{t-1} - X_{t-2}) \\&= X_t - 2X_{t-1} + X_{t-2} \\&= (1 - 2B + B^2)X_t \\&= (1 - B)^2X_t\end{aligned}$$

Tujuan dari menghitung *differencing* adalah untuk mencapai stasioneritas. Selanjutnya stasioneritas dibagi menjadi 2, yaitu:

### 1) Stasioner dalam *mean* (rata-rata)

Stasioner dalam *mean* adalah fluktuasi data berada di sekitar suatu nilai rata-rata yang konstan, tidak tergantung pada waktu dan variansi dari fluktuasi tersebut. Dari bentuk *plot* data seringkali dapat diketahui bahwa data tersebut stasioner atau tidak stasioner. Apabila dilihat dari *plot* ACF, maka nilai-nilai autokorelasi dari data stasioner akan turun menuju nol sesudah *time lag* (selisih waktu) kedua atau ketiga.

## 2) Stasioneritas dalam Variansi

Suatu data *time series* dikatakan stasioner dalam variansi apabila struktur data dari waktu ke waktu mempunyai fluktuasi data yang tetap atau konstan dan tidak berubah-ubah. Secara visual untuk melihat hal tersebut dapat dibantu dengan menggunakan plot *time series*, yaitu dengan melihat fluktuasi data dari waktu ke waktu.

Untuk melihat kestasioneran data dapat dilakukan melalui:

### 2.4.1. Korelogram

Korelogram merupakan teknik untuk mengidentifikasi kestasioneran data *time series* dengan melihat *Auto Correlation Function* (ACF) yang memberikan informasi korelasi antara  $X_t$  yang berdekatan. Formula ACF diberikan sebagai berikut.

$$\rho_k = \frac{Cov(X_t, X_{t+k})}{\sigma_{X_t} \sigma_{X_{t+k}}}; |\rho_k| \leq 1$$

Untuk proses yang stasioner,  $\sigma_{X_t} = \sigma_{X_{t+k}}$  sehingga formula di atas dapat dituliskan menjadi

$$\rho_k = \frac{Cov(X_t, X_{t+k})}{\sigma_X^2}; \frac{\gamma_k}{\gamma_0}.$$

Sementara itu, untuk menghitung autokorelasi sampel, digunakan formula berikut.

$$\hat{\rho}_k = r_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum (X_t - \bar{X}_t)(X_{t+k} - \bar{X}_t)}{\sum (X_t - \bar{X}_t)^2}$$

Plot antara  $\rho_k$  dan  $k$  disebut dengan korelogram.

Secara formal, kestasioneran data dapat diuji dengan tiga uji berikut.

#### 2.4.1.1. Uji Box Pierce

- Hipotesis

$$H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0$$

$$H_i: \text{Terdapat setidaknya satu } \rho_k \neq 0$$

- Taraf signifikansi

$$\alpha = 0,05$$

- Statistik Uji

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

di mana  $T$  = banyaknya pengamatan dan  $m = \text{panjang lag } \frac{1}{3}T$

- Aturan keputusan

$H_0$  ditolak apabila  $Q > \chi_{\alpha, m}^2$

- Kesimpulan

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

#### 2.4.1.2. Uji Bartlett

- Hipotesis

$H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0$

$H_i$ : Terdapat setidaknya satu  $\rho_k \neq 0$

- Taraf signifikansi

$\alpha = 0,05$

- Statistik Uji

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

- Aturan keputusan

$H_0$  ditolak apabila  $\frac{r_k - \rho_k}{\sqrt{\text{Var}(r_k)}} > Z_{\frac{\alpha}{2}}$

- Kesimpulan

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

#### 2.4.1.3. Uji Ljung-Box

- Hipotesis

$H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0$

$H_i$ : Terdapat setidaknya satu  $\rho_k \neq 0$

- Taraf signifikansi

$\alpha = 0,05$

- Statistik Uji

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

- Aturan keputusan

$H_0$  ditolak apabila  $LB > \chi_{\alpha, m}^2$

- Kesimpulan

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

## 2.4.2. Uji Unit Akar (*Unit Root*)

### 2.4.2.1. Uji Dickey Fuller (DF), model AR(1)

Model AR(1)  $X_t = \phi X_{t-1} + \varepsilon_t$  dimanipulasi menjadi

$$X_t - X_{t-1} = (\phi - 1)X_{t-1} + \varepsilon_t$$

$$\Delta X_t = \pi X_{t-1} + \varepsilon_t$$

sehingga berdasarkan model di atas dapat dibentuk uji hipotesis sebagai berikut.

- Hipotesis

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

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

- Taraf signifikansi

$$\alpha = 0,05$$

- Statistik Uji

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

- Aturan keputusan

$H_0$  ditolak apabila  $|\tau| > |DF_{tabel}|$  (tabel uji-t) atau  $p - value < \alpha$

- 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 dan memasukkan variable bebas waktu ( $t$ ).

### 2.4.2.2. Uji Augmented Dickey Fuller (ADF)

Uji ini merupakan pengembangan dari uji DF untuk model AR(P) dengan  $P > 1$ .

- Hipotesis

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

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

- Taraf signifikansi

$$\alpha = 0,05$$

- Statistik Uji

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

- Aturan keputusan

$H_0$  ditolak apabila  $|\tau| > |DF_{tabel}|$  (tabel uji-t) atau  $p - value < \alpha$

- 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 saja, model dengan intersep dan tren, maupun model tanpa intersep dan tren.

## 2.5. Spesifikasi Model

Identifikasi nilai  $p$ ,  $d$ , dan  $q$  yang sesuai dalam penentuan model terbaik dilakukan menggunakan metode sebagai berikut.

### 2.5.1. Autocorrelation Function (ACF)

Untuk runtun teramati  $X_1, X_2, \dots, X_n$ , maka taksiran parameter  $\rho_k$  adalah sebagai berikut.

$$r_k = \frac{\sum_{t=k+1}^n (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=k+1}^n (X_t - \bar{X})^2}; k = 1, 2, \dots$$

Plot ACF dibentuk dengan tujuan untuk mengamati pola dari  $r_k$  yang menyerupai karakteristik  $\rho_k$  pada model ARMA. Plot ACF yang menyusut secara eksponensial menandakan model AR. ACF merupakan indikator yang baik untuk menentukan order  $q$  dalam proses MA( $q$ ). Hal ini disebabkan oleh garis putus-putus pada grafik ACF yang menampilkan nilai kritis untuk menguji apakah autokorelasi sampel signifikan berbeda dengan nol atau tidak. Jika  $\rho_k = 0$  untuk  $k > q$ , maka model yang cocok adalah model MA( $q$ ).

### 2.5.2. Partial Autocorrelation Function (PACF)

Koefisien autokorelasi parsial pada lag ke- $k$  dengan notasi  $\phi_{kk}$  merupakan fungsi baru yang dibentuk dengan cara mendefinisikan korelasi antara  $Y_t$  dan  $Y_{t-k}$  setelah menghapuskan efek dari variabel-variabel di antara keduanya karena ACF tidak dapat menentukan order  $p$  dari AR( $p$ ). Plot PACF yang menurun secara eksponensial menandakan model MA. Sebaliknya dengan ACF, jika



$\phi_{kk} = 0$  untuk  $k > q$ , maka model yang cocok adalah model  $AR(p)$  sehingga PACF merupakan indikator yang baik untuk menentukan order dari proses  $AR(p)$ .

Fungsi PACF dapat ditulis sebagai berikut:

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \varphi_3 \rho_{k-3} + \cdots + \varphi_p \rho_{k-p}$$

Keterangan:

$k = \text{lag waktu}$  ;  $k = 1, 2, \dots, p$

$\rho = \text{Nilai dari ACF}$

$\varphi = \text{Nilai dari PACF}$

Rumus PACF dapat ditulis sebagai berikut.

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_{k-j}} ; k = 3, 4, \dots$$

Keterangan:

$$\phi_{kj} = \phi_{k-1,j} - \phi_{kk} \phi_{k-1,k-j} ; k = 2 \text{ dan } j = 1, 2, \dots, k-1$$

Pola ACF dan PACF model dapat diringkas sebagai berikut:

Model	Pola ACF	Pola PACF
<b>AR(p)</b>	Menurun secara eksponensial pada gelombang sinus	Ada tiang pancang sampai lag ke- $p$ , lalu terpotong setelah lag ke- $p$
<b>MA(q)</b>	Terpotong setelah lag ke- $q$ , ada tiang pancang yang jelas sampai lag ke- $q$	Menurun secara eksponensial
<b>ARMA(p, q)</b>	Menurun secara eksponensial, setelah lag ke $(q - p)$	Menurun secara eksponensial, setelah lag ke $(p - q)$

### 2.5.3. Extended Autocorrelation Function (EACF)

Model ARMA tidak dapat diidentifikasi menggunakan menggunakan ACF dan PACF sehingga diperlukan metode lain yaitu EACF yang merupakan perluasan dari fungsi autokorelasi parsial. EACF berguna untuk menentukan orde  $p$  dan  $q$  dari ARMA yang didefinisikan dalam persamaan berikut.

$$W_{t,k,j} = X_t - \tilde{\phi}_1 X_{t-1} - \tilde{\phi}_2 X_{t-2} - \cdots - \tilde{\phi}_k X_{t-k}$$

Persamaan di atas merupakan persamaan untuk residual autoregresif yang didefinisikan dengan koefisien AR yang diestimasi secara iteratif dengan asumsi orde AR sebesar  $k$  dan orde MA sebesar  $j$ . EACF dapat disajikan dalam tabel yang berisi simbol 0 dan x. Simbol x diberikan pada baris ke- $k$  dan kolom ke- $j$  jika autokorelasi sampel  $W_{t,k,j}$  untuk lag  $j + 1$  berbeda secara signifikan dengan

nol. Sebaliknya, symbol 0 diberikan pada baris ke- $k$  dan kolom ke- $j$  jika autokorelasi sampel  $W_{t,k,j}$  untuk lag  $j + 1$  tidak berbeda secara signifikan dengan nol.

#### 2.5.4. Akaike Information Criterion (AIC)

AIC dapat digunakan sebagai salah satu kriteria pemilihan model yang terbaik. Model yang dipilih adalah model yang meminimumkan nilai:

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

di mana  $k = p + q$  jika model tidak mengandung suku konstanta dan  $k = p + q + 1$  jika model mengandung suku konstanta.

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

Kriteria lain dalam pemilihan model yang terbaik selain AIC adalah BIC.

Model yang dipilih adalah model yang meminimumkan nilai:

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

di mana  $k$  adalah jumlah parameter bebas yang diestimasi.

### 2.6. Estimasi Parameter

Jika sudah didapatkan model ARIMA( $p, d, q$ ), maka langkah selanjutnya kita perlu mengestimasi parameter dari model tersebut. Metode yang biasa digunakan antara lain sebagai berikut.

#### 1. Metode Momen

Metode Momen adalah metode untuk mengestimasi parameter dengan menyamakan sampel momen dengan teori momen serta menyelesaikan persamaan tersebut untuk memperoleh parameter yang tidak diketahui. Metode ini akan menghasilkan parameter yang cukup baik untuk model AR( $p$ ), namun kurang baik untuk model MA( $q$ ).

#### 2. Metode *Least-Square*

Metode *Least-Square* adalah metode untuk mengestimasi parameter dengan meminimumkan jumlah kuadrat *error*.

#### 3. Metode *Maximum Likelihood*

Metode *Maximum Likelihood* adalah metode untuk mengestimasi parameter dengan memaksimalkan fungsi *likelihood*, yaitu pdf bersama dari sampel random yang menghasilkan titik *estimator* oleh distribusi dari sampel random tersebut.

## 2.7. *Diagnostic Checking*

Diagnostic checking adalah tahapan penting dalam evaluasi model yang melibatkan pengujian model estimasi terhadap data aktual. Pada tahap ini, kita melakukan analisis terhadap residual, yaitu selisih antara nilai prediksi yang dihasilkan oleh model dengan nilai sebenarnya dari data.

Pertama, kita melakukan uji independensi terhadap residual. Uji ini bertujuan untuk memastikan bahwa residual tidak memiliki korelasi atau pola tertentu yang masih ada setelah model diterapkan. Hal ini penting karena jika terdapat korelasi atau pola pada residual, maka itu berarti model belum mampu menjelaskan semua variabilitas dalam data dengan baik. Uji independensi biasanya melibatkan analisis autokorelasi pada residual, seperti uji Durbin-Watson atau uji Ljung-Box.

Selanjutnya, kita melakukan uji normalitas terhadap residual. Uji ini berguna untuk memverifikasi apakah residual mengikuti distribusi normal. Jika residual terdistribusi secara normal, maka asumsi dasar dari model kita terpenuhi. Beberapa uji yang umum digunakan adalah uji Jarque-Bera dan uji Shapiro-Wilk. Jika hasil uji menunjukkan bahwa residual tidak memenuhi asumsi normalitas, maka perlu dilakukan transformasi atau pertimbangan lain terhadap model yang digunakan.

Melalui tahapan diagnostic checking ini, kita dapat memperoleh pemahaman yang lebih mendalam tentang performa model yang kita gunakan. Jika hasil pengujian menunjukkan adanya masalah pada residual, seperti adanya pola korelasi atau ketidaknormalan, maka perlu dilakukan penyesuaian atau pemilihan model yang lebih sesuai. Dengan melakukan evaluasi yang cermat terhadap model estimasi, kita dapat meningkatkan akurasi dan keandalan model dalam meramalkan atau menggambarkan pola data sebenarnya.

### 2.7.1. *Uji Independensi Ljung-Box*

Uji ini dilakukan untuk melihat independensi antar residual. Langkah-langkah dalam pengujian ini adalah sebagai berikut.

- Hipotesis  
 $H_0$ : Tidak ada korelasi antar residual  
 $H_i$ : Terdapat korelasi antar residual
- Taraf signifikansi  
 $\alpha = 0,05$

- Statistik Uji

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

- Aturan keputusan

$H_0$  ditolak apabila  $Q > \chi_{\alpha, m}^2$  atau  $p - value < \alpha$

- Kesimpulan

Jika  $H_0$  ditolak maka dapat disimpulkan bahwa terdapat korelasi antar residual (residual tidak independen) sehingga model yang diestimasi kurang sesuai.

### 2.7.2. Uji Normalitas Jarque-Bera

Uji ini dilakukan untuk melihat apakah residual berdistribusi normal. Langkah-langkah dalam pengujian ini adalah sebagai berikut.

- Hipotesis

$H_0$ : Residual data berdistribusi normal

$H_t$ : Residual data tidak berdistribusi normal

Taraf signifikansi

$\alpha = 0,05$

- Statistik Uji

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

di mana  $K$  adalah kurtosis,  $K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}$  dan  $S$  adalah *skewness*,

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{3}{2}}}$$

- Aturan keputusan

$H_0$  ditolak apabila  $JB > \chi_{\alpha, 2}^2$  atau  $p - value < \alpha$

- Kesimpulan

Jika  $H_0$  ditolak maka dapat disimpulkan bahwa residual data tidak berdistribusi normal sehingga model yang diestimasi kurang sesuai.

### 2.7.3. *Overfitting Model*

*Overfitting* adalah salah satu langkah dalam analisis time series yang dilakukan untuk menguji apakah model yang dipilih benar-benar merupakan model yang paling baik untuk menggambarkan pola dalam data. Pada tahap ini, kita mencoba menambahkan parameter tambahan pada model dengan meningkatkan orde baik pada model autoregressive (AR) maupun model moving average (MA).

Pada model AR, orde mengacu pada jumlah lag atau periode waktu sebelumnya yang digunakan untuk memprediksi nilai pada periode berikutnya. Pada model MA, orde mengacu pada jumlah lag atau periode waktu sebelumnya yang digunakan dalam perhitungan rata-rata bergerak.

Dalam proses *overfitting*, penambahan parameter dilakukan secara bertahap, artinya kita meningkatkan orde satu suku (misalnya AR) sementara mempertahankan orde suku lainnya (misalnya MA). Setelah itu, kita mengevaluasi model baru yang telah ditambahkan parameter tersebut untuk melihat apakah menghasilkan peningkatan yang signifikan dalam kualitas perkiraan.

Namun, perlu diingat bahwa *overfitting* juga bisa berpotensi menimbulkan masalah. Jika kita terlalu memaksakan penambahan parameter atau orde pada model, maka model bisa menjadi terlalu kompleks dan cenderung mengikuti pola-pola acak dalam data training yang tidak akan terefleksikan dengan baik dalam data masa depan. Oleh karena itu, penting untuk melakukannya dengan hati-hati dan menggunakan teknik evaluasi yang tepat untuk memastikan bahwa penambahan parameter atau orde pada model benar-benar meningkatkan kualitas perkiraan secara signifikan.

### 2.7.4. *Cross Validation*

Pada tahap ini, data aktual akan dibagi menjadi dua bagian, yaitu data untuk membangun model dan data untuk menguji model. Tujuan utamanya adalah untuk memastikan bahwa data aktual masuk dalam interval kepercayaan dari nilai prediksi. Jika data aktual tidak masuk dalam interval kepercayaan tersebut, dapat disimpulkan bahwa model yang digunakan kurang sesuai dan perlu diperbaiki.

## 2.8. *Forecasting*

Tujuan utama dalam pembuatan model runtun waktu adalah melakukan peramalan terhadap nilai runtun tersebut untuk periode waktu yang akan datang. Selain itu, penting juga untuk mencapai presisi yang tinggi dalam melakukan peramalan tersebut. Dalam asumsi ini, diasumsikan bahwa model yang digunakan sudah diketahui dengan pasti, termasuk nilai-nilai estimasi parameter yang digunakan dalam model tersebut. Peramalan model ARMA(p,q) dirumuskan sebagai berikut.

$$\begin{aligned}
 X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \\
 \hat{X}_t(l) &= \phi_1 \hat{X}_t(l-1) + \phi_2 \hat{X}_t(l-2) + \dots + \phi_p \hat{X}_t(l-p) + \theta_0 - \theta_1 E(\varepsilon_{t+l-1} | X_1, X_2, \dots, X_t) \\
 &\quad - \theta_2 E(\varepsilon_{t+l-2} | X_1, X_2, \dots, X_t) - \dots - \theta_q E(\varepsilon_{t+l-q} | X_1, X_2, \dots, X_t)
 \end{aligned}$$

Untuk model ARIMA(p,d,q);

$$\begin{aligned}
 X_t &= \psi_1 X_{t-1} + \psi_2 X_{t-2} + \dots + \psi_p X_{t-p} + \psi_{p+1} X_{t-p-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots \\
 &\quad - \theta_q \varepsilon_{t-q} \\
 \hat{X}_t(l) &= \phi_1 \hat{X}_t(l-1) + \phi_2 \hat{X}_t(l-2) + \dots + \phi_p \hat{X}_t(l-p) + \theta_0 - \theta_1 E(\varepsilon_{t+l-1} | X_1, X_2, \dots, X_t) \\
 &\quad - \theta_2 E(\varepsilon_{t+l-2} | X_1, X_2, \dots, X_t) - \dots - \theta_q E(\varepsilon_{t+l-q} | X_1, X_2, \dots, X_t)
 \end{aligned}$$

### BAB III

#### METODE PENELITIAN

##### 3.1. Variabel Penelitian

Variabel yang digunakan pada penelitian ini adalah temperatur harian Kota Warsaw periode 1 Juni 2022 – 31 Desember 2022.

##### 3.2. Data

Data yang digunakan dalam penelitian ini merupakan data kuantitatif yang berupa numerik dan dapat diukur atau dihitung secara langsung. Data yang digunakan dalam penelitian ini berisi temperatur harian Kota Warsaw pada periode 1 Juni 2022 – 31 Desember 2022 yang terdiri dari 213 observasi. Berikut cuplikan dari data yang digunakan.

Tanggal	Temperatur
01/06/2022	15.8
02/06/2022	14.7
03/06/2022	16.5
04/06/2022	18.6
05/06/2022	18.9
06/06/2022	20.2
07/06/2022	21.3
08/06/2022	18.1
09/06/2022	21.4
10/06/2022	19.5

##### 3.3. Metode Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan data sekunder yaitu data yang diperoleh dari sumber-sumber yang sudah ada sebelumnya. Pada penelitian ini, data yang digunakan bersumber dari *website* bernama Kaggle. Kaggle merupakan salah satu *platform* yang menyediakan berbagai *dataset* publik yang dapat digunakan oleh para praktisi *data science* untuk melakukan eksplorasi, analisis, dan pengembangan model.

### 3.4. Metode Analisis Data

Pada penelitian ini, metode analisis yang digunakan adalah analisis runtun waktu. Data yang telah tersedia dari periode waktu sebelumnya diolah dan kemudian dilakukan peramalan (*forecasting*) menggunakan metode ARIMA (*Autoregressive Integrated Moving Average*). Hasil dari analisis ini adalah prediksi temperatur harian Kota Warsaw untuk tiga periode ke depan yaitu 1 Januari 2023 – 3 Januari 2023. Proses analisis data dilakukan dengan bantuan *software* RStudio. Langkah-langkah analisis yang dilakukan dijabarkan sebagai berikut.

1. Melakukan uji stasioner pada data melalui plot dan uji Augmented Dickey Fuller.
2. Melakukan transformasi (*differencing*) pada data jika data tidak stasioner.
3. Melakukan spesifikasi model dengan menduga model ARMA(p,q) dari grafik ACF, PACF, ataupun EACF (jika diperlukan) dengan acuan sebagai berikut.
  - Jika pola ACF menyusut secara eksponensial atau pola berbentuk gelombang sinusoidal yang tidak begitu jelas dan pada PACF terdapat tiang pancang hingga lag p, maka model yang tepat adalah AR(p).
  - Jika pada ACF terdapat tiang pancang yang jelas hingga lag p dan pola PACF menyusut secara eksponensial maka model yang tepat adalah MA(q).
  - Jika pola ACF dan PACF menyusut secara eksponensial maka model yang tepat adalah ARMA(p,q).
4. Jika menggunakan model EACF dan terdapat beberapa kandidat model, penentuan model yang tepat dilakukan dengan langkah sebagai berikut.
  - 1) Mencari tahu bentuk tabel EACF
  - 2) Menduga kandidat model dari tabel EACF. Secara teoritis, tabel EACF untuk model ARMA(p,q) memiliki pola segitiga dengan angka "0", di mana angka "0" terletak di bagian kiri atas tabel pada baris ke-p dan kolom ke-q. Namun, dalam praktiknya, EACF sampel tidak selalu menunjukkan pola yang jelas.
  - 3) Membandingkan nilai AIC dan BIC dari kandidat model. Model yang dipilih adalah model yang memiliki nilai AIC dan BIC terkecil.
5. Mengestimasi parameter untuk model terbaik menggunakan metode momen, *least square*, dan *maximum likelihood*.
6. Mengevaluasi kecocokan model dengan melakukan *model diagnostic* menggunakan dua pendekatan, yaitu analisis residual dan *overfitting*. Jika hasil analisis



menunjukkan bahwa model tidak memadai, maka akan dilakukan penyesuaian atau spesifikasi ulang pada model tersebut.

7. Melakukan *forecasting* (peramalan) untuk beberapa periode selanjutnya.

## BAB IV

### ANALISIS DATA DAN PEMBAHASAN

#### 4.1. *Preprocessing Data*

Sebelum mengubah data ke bentuk *time-series*, akan dilakukan *preprocessing* terlebih dahulu untuk memastikan tidak ada masalah yang dapat mengganggu pemrosesan data nantinya. *Preprocessing* dilakukan dengan mengambil data temperatur dari *dataset*, kemudian mengecek *missing values* dan *outlier* dari data.

Sebelumnya, berikut adalah *packages* dan *library* yang akan digunakan dalam analisis data.

```
install.packages("tseries")
install.packages("adf")
install.packages("forecast")
install.packages("lmtest")
library(zoo)
library(tseries)
library(forecast)
library(lmtest)
library(TSA)
library(car)
```

- Mengambil data temperatur dari *dataset*

Pertama, akan dilihat *datasetnya* terlebih dahulu(enam data pertama)

```
> head(temp0)
```

	STATION	NAME	LATITUDE	LONGITUDE	ELEVATION	DATE	PRCP	SNWD
1	PLM00012375	OKECIE, PL	52.166	20.967	110.3	1993-01-01	0	10
2	PLM00012375	OKECIE, PL	52.166	20.967	110.3	1993-01-02	NA	10
3	PLM00012375	OKECIE, PL	52.166	20.967	110.3	1993-01-03	0	10
4	PLM00012375	OKECIE, PL	52.166	20.967	110.3	1993-01-04	0	10
5	PLM00012375	OKECIE, PL	52.166	20.967	110.3	1993-01-05	0	10
6	PLM00012375	OKECIE, PL	52.166	20.967	110.3	1993-01-06	NA	10

	TAVG	TMAX	TMIN
1	-8.3	NA	NA
2	-14.9	NA	NA
3	-13.6	-9.7	NA
4	-10.5	-6.5	-13.3
5	-12.0	-8.9	-14.1
6	-2.3	NA	-12.9

Karena data yang akan dianalisis adalah data temperatur harian, akan dibuat *data frame* baru yang memuat data “DATE” dan “TAVG” saja, kemudian ubah nama kolomnya menjadi “Tanggal” dan “Temperatur”. Selain itu, akan diubah format data “Tanggal” dengan perintah `as.Date()` untuk memudahkan pengubahan *dataframe* ke *time-series* nantinya. Hasilnya sebagai berikut

```

> temp <- data.frame(temp0$DATE, temp0$TAVG)
> colnames(temp) <- c("Tanggal", "Temperatur")
> temp$Tanggal <- as.Date(temp$Tanggal)
> head(temp)
  Tanggal Temperatur
1 1993-01-01      -8.3
2 1993-01-02     -14.9
3 1993-01-03     -13.6
4 1993-01-04     -10.5
5 1993-01-05     -12.0
6 1993-01-06      -2.3

```

Selanjutnya, akan diambil data temperatur di bulan Juni sampai Desember 2022 sebagai berikut.

```

> #mengambil data temperatur di bulan juni-desember 2022
> temp <- temp[temp[["Tanggal"]] >= "2022-06-01",]
> rownames(temp) <- 1:nrow(temp)
> head(temp)
  Tanggal Temperatur
1 2022-06-01      15.8
2 2022-06-02      14.7
3 2022-06-03      16.5
4 2022-06-04      18.6
5 2022-06-05      18.9
6 2022-06-06      20.2

```

- Mengecek *missing values*

Kemudian, akan dicek apakah terdapat *missing values* pada data.

```

> #mengecek missing values
> sum(is.na(temp$Tanggal)) ; sum(is.na(temp$Temperatur))
[1] 0
[1] 0

```

Terlihat bahwa *output* menghasilkan nilai 0 yang berarti tidak terdapat *missing values* pada data sehingga tidak perlu dilakukan imputasi data.

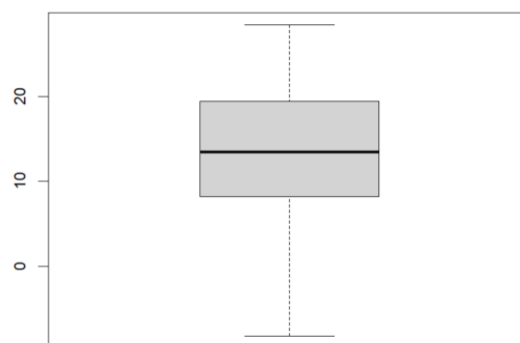
- Mengecek *Outlier*

Selanjutnya, akan dicek apakah terdapat *outlier* pada data temperatur menggunakan visualisasi boxplot

```

> #mengecek outlier
> boxplot(temp$Temperatur)

```

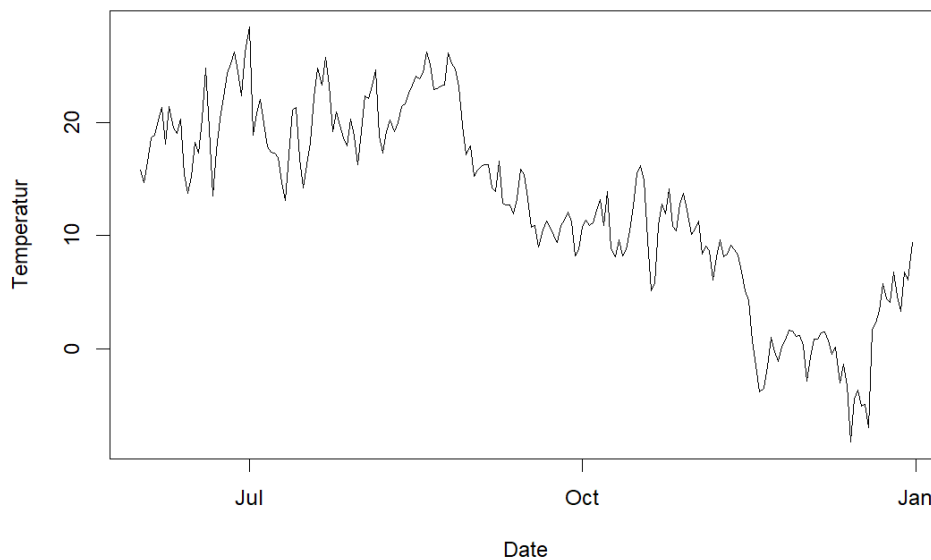


Terlihat bahwa tidak terdapat titik yang berada di luar boxplot sehingga tidak terdapat *outlier* pada data temperatur.

#### 4.2. Membentuk Data *Time-Series*

Setelah dilakukan preprocessing, akan dilakukan pembentukan data time-series dan akan dilihat plot-nya sebagai berikut

```
> #mengubah data ke bentuk time series
> temp <- read.zoo(temp)
> head(temp)
2022-06-01 2022-06-02 2022-06-03 2022-06-04 2022-06-05 2022-06-06
      15.8      14.7      16.5      18.6      18.9      20.2
> #plot time seriesnya
> plot(temp, xlab="Date", ylab="Temperatur")
```



Dari hasil *plot time-series* terlihat persebaran data temperatur dari awal bulan Juni sampai akhir Desember. Terdapat dugaan bahwa data masih belum stasioner karena terdapat tren pada data, yaitu temperatur cenderung menurun saat mendekati bulan Desember.

#### 4.3. Uji Stasioneritas Data

Akan dilakukan uji stasioner pada data time-series menggunakan uji Augmented Dickey-Fuller(ADF).

- Hipotesis

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

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

- Taraf signifikansi

$\alpha = 0,05$

- Statistik Uji

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

Menggunakan bantuan R, didapatkan hasil dari uji ADF sebagai berikut.

```
> #uji stasioneritas
> adf.test(temp)

Augmented Dickey-Fuller Test

data: temp
Dickey-Fuller = -2.6093, Lag order = 5, p-value =
0.3205
alternative hypothesis: stationary
```

- Aturan keputusan

$H_0$  ditolak apabila  $|\tau| > |DF_{tabel}|$  (tabel uji-t) atau  $p - value < \alpha$

- Kesimpulan

Karena  $p - value = 0.3205 > 0.05$  maka  $H_0$  tidak ditolak yang berarti data tidak stasioner sehingga perlu dilakukan *differencing* pada data.

Karena data tidak stasioner, akan dilakukan uji stasioner ulang pada data *time-series* yang telah di-*differencing* satu kali menggunakan uji Augmented Dickey-Fuller(ADF).

- Hipotesis

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

$H_i: \pi \neq 0$  (data stasioner)

- Taraf signifikansi

$\alpha = 0,05$

- Statistik Uji

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

Menggunakan bantuan R, akan dilakukan *differencing* satu kali pada data dan dilakukan uji ADF sebagai berikut.

```
> #differencing satu kali
> difftemp <- diff(temp,differences = 1)
> adf.test (difftemp)

Augmented Dickey-Fuller Test

data: difftemp
Dickey-Fuller = -7.5745, Lag order = 5, p-value =
0.01
alternative hypothesis: stationary
```

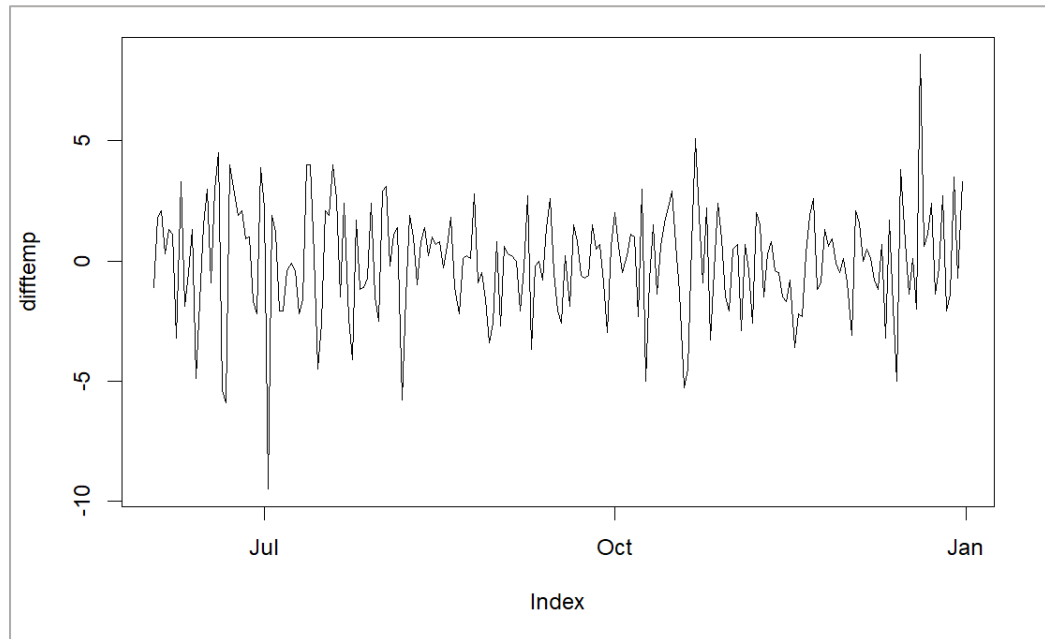
- Aturan keputusan

$H_0$  ditolak apabila  $|\tau| > |DF_{tabel}|$  (tabel uji-t) atau  $p - value < \alpha$

- Kesimpulan

Karena  $p - value = 0.01 < 0.05$  maka  $H_0$  ditolak yang berarti data stasioner.

Hal ini juga dapat dilihat dari hasil plot *differencing* yang sudah tidak terdapat tren.



#### 4.4. Spesifikasi Model

Karena data sudah stasioner setelah di-*differencing* satu kali, selanjutnya akan dilakukan spesifikasi model, yaitu mencari model ARIMA(p,1,q) yang terbaik untuk data yang digunakan.

##### 4.4.1. Identifikasi Model

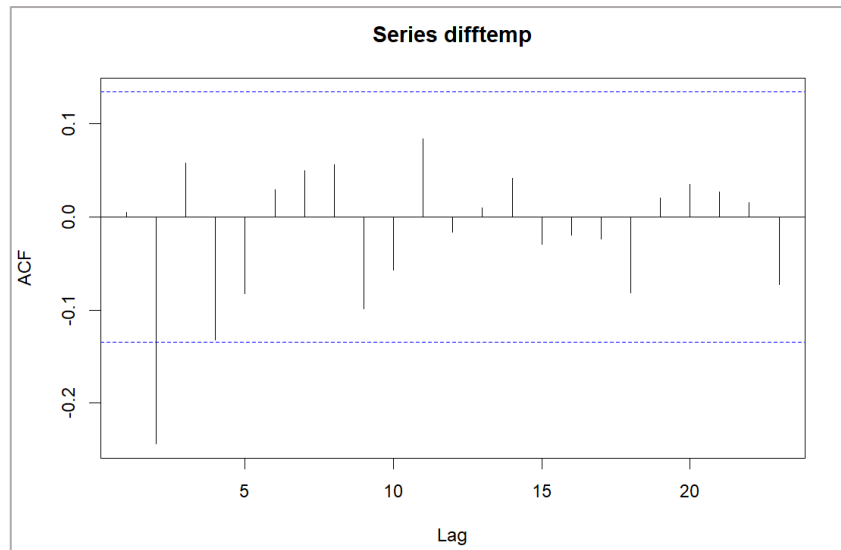
Pertama, akan diidentifikasi model dari data dengan melihat grafik ACF, PACF, dan tabel EACF untuk mendapatkan order p dan q dari model.

*Input:*

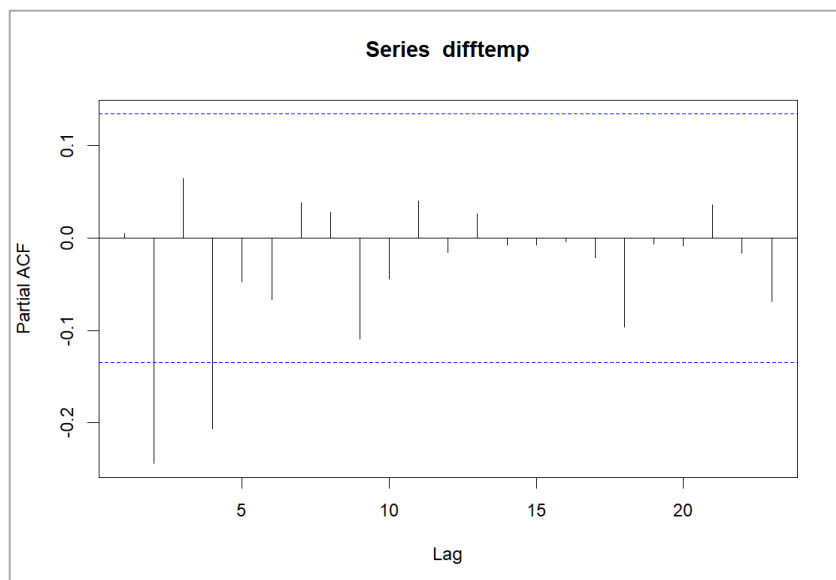
```
#spesifikasi model
pacf(difftemp)
acf(difftemp)
eacf(difftemp)
```

*Output:*

- Grafik ACF



- Grafik PACF



- Tabel EACF

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

Dari ketiga hasil di atas, dapat dilihat bahwa grafik ACF dan PACF tidak memberikan informasi yang cukup mengenai model yang akan diajukan.

Maka, perlu dilihat tabel EACF untuk mengetahui lebih lanjut model yang akan diajukan. Berdasarkan tabel EACF, kami menduga terdapat 4 model ARIMA (p,1,q) yang cukup baik untuk menggambarkan data, kandidat model tersebut antara lain

- Model 1: ARIMA (1, 1, 0)
- Model 2: ARIMA (0, 1, 2)
- Model 3: ARIMA (1, 1, 2)
- Model 4: ARIMA (2, 1, 2)

#### 4.4.2. Membandingkan Model

Dari 4 model yang telah diajukan sebelumnya, akan dicari model yang terbaik dari keempatnya. Untuk membandingkannya, akan dilihat nilai AIC dan BIC dari setiap model. Model dengan nilai AIC dan BIC yang paling minimum akan dipilih menjadi model terbaik. Menggunakan bantuan R, didapatkan hasil sebagai berikut

```
> #model yang diajukan
> model1 <- Arima(temp,order=c(1,1,0))
> model2 <- Arima(temp,order=c(0,1,2))
> model3 <- Arima(temp,order=c(1,1,2))
> model4 <- Arima(temp,order=c(2,1,2))

data.frame(AIC(model1),AIC(model2),AIC(model3),AIC(model4),BIC(model1),BIC(model2),BIC(model3),BIC(model4))
AIC.model1. AIC.model2. AIC.model3. AIC.model4. BIC.model1. BIC.model2. BIC.model3. BIC.model4.
      961.9405      943.75      945.6289      945.559      968.663      953.8339      959.0741      962.3654
```

Dari hasil di atas, terlihat bahwa nilai AIC dan BIC minimum dimiliki oleh model 2, yaitu AIC = 943.75 dan BIC = 953.8339 sehingga model terbaik untuk data adalah ARIMA(0,1,2).

#### 4.5. Estimasi Parameter

Menggunakan bantuan RStudio, didapat estimasi parameter model ARIMA(0,1,2) dari data, yaitu sebagai berikut.

```
> #estimasi parameter
> model <- Arima(temp,order=c(0,1,2))
> model
Series: temp
ARIMA(0,1,2)

Coefficients:
      ma1      ma2
      0.0384 -0.3590
s.e.    0.0671  0.0715

sigma^2 = 4.82: log likelihood = -468.87
AIC=943.75 AICc=943.86 BIC=953.83
```



Sehingga didapatkan model data *time-series* berikut.

$$\begin{aligned} W_t &= e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \\ &= e_t - 0.0384e_{t-1} - 0.359e_{t-2} \end{aligned}$$

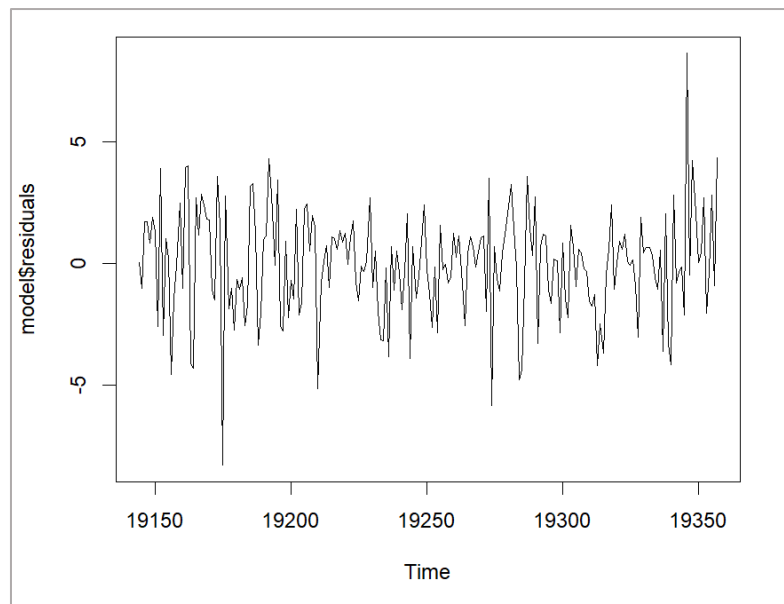
#### 4.6. Diagnosis Model

Setelah mendapatkan estimasi parameter dari model, langkah selanjutnya adalah melakukan diagnosis model yang terdiri dari analisis residual dan *overfitting*.

##### 4.6.1. Analisis Residual

Pada langkah analisis residual, akan dilakukan uji stasioneritas, independensi, dan kenormalan residual dari model. Sebelum itu, berikut adalah *plot* dari residual model

Residual adalah selisih antara data hasil ramalan dengan data sebenarnya. Residual akan bernilai kecil jika selisih antara data hasil ramalan dengan data sebenarnya tidak berbeda terlalu jauh. Jika model dikatakan baik dan estimasi parameternya mendekati nilai sebenarnya, maka residualnya akan memiliki properti *white noise*. Berikut merupakan *plot* dari residual model data.



Keterangan “Time”:

19150 = 7 Juni 2022

19200 = 27 Juli 2022

19250 = 15 September 2022

19300 = 4 November 2022

#### 4.6.1.1. Uji Stasioneritas Residual

Akan digunakan uji Augmented Dickey-Fuller (ADF) untuk mengecek kestasioneran residual dari model.

- Hipotesis

$H_0: \pi = 0$  (residual model tidak stasioner)

$H_1: \pi \neq 0$  (residual model stasioner)

- Taraf signifikansi

$\alpha = 0,05$

- Statistik Uji

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

Menggunakan bantuan R, didapatkan hasil dari uji ADF sebagai berikut.

```
> #uji stasioner residual (pvalue < 0.05)
> adf.test(model$residuals)
```

Augmented Dickey-Fuller Test

```
data: model$residuals
Dickey-Fuller = -6.2795, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

- Aturan keputusan

$H_0$  ditolak apabila  $|\tau| > |DF_{tabel}|$  (tabel uji-t) atau  $p - value < \alpha$

- Kesimpulan

Karena  $p - value = 0.01 < 0.05$  maka  $H_0$  ditolak yang berarti residual model stasioner.

#### 4.6.1.2. Uji Independensi Residual

Akan digunakan uji Ljung-Box untuk menguji asumsi independensi residual dari model.

- Hipotesis

$H_0$ : Semua  $\rho_k = 0$  (tidak terdapat korelasi antar residual)

$H_1$ : Setidaknya terdapat satu  $\rho_k \neq 0$  (terdapat korelasi antar residual)

- Taraf signifikansi

$\alpha = 0,05$

- Statistik Uji

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

Menggunakan bantuan R, didapatkan hasil dari uji Ljung-Box sebagai berikut.

```
> #uji independensi residual ljung-box
> checkresiduals(model$residuals)
```

Ljung-Box test

```
data: Residuals
Q* = 6.9187, df = 10, p-value = 0.7331
```

```
Model df: 0. Total lags used: 10
```

- Aturan keputusan

$H_0$  ditolak apabila  $Q > \chi_{\alpha, m}^2$  atau  $p - value < \alpha$

- Kesimpulan

Karena  $p - value = 0.7331 > 0.05$  maka  $H_0$  tidak ditolak yang berarti asumsi independensi terpenuhi atau tidak terdapat korelasi residual antar-lag.

#### 4.6.1.3. Uji Normalitas Residual

Akan digunakan uji Shapiro-Wilk dan uji Jarque-Bera untuk menguji asumsi kenormalan residual.

- Hipotesis

$H_0$ : Residual data berdistribusi normal

$H_1$ : Residual data tidak berdistribusi normal

- Taraf signifikansi

$\alpha = 0,05$

- Statistik Uji

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

$$\text{Dengan } K \text{ adalah kurtosis} = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}$$

$$\text{dan } S \text{ adalah skewness} = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{3}{2}}}$$

Menggunakan bantuan R, didapatkan hasil dari uji Shapiro-Wilk dan Jarque-Bera sebagai berikut.

```

> #uji normalitas residual
> shapiro.test(model$residuals)

      Shapiro-Wilk normality test

data:  model$residuals
W = 0.98522, p-value = 0.02493

> jarque.bera.test(model$residuals)

      Jarque Bera Test

data:  model$residuals
X-squared = 14.829, df = 2, p-value = 0.0006024

```

- Aturan keputusan

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

- Kesimpulan
- Karena  $p - value = 0.02493 < 0.05$  untuk uji Shapiro-Wilk dan  $p - value = 0.0006024$  untuk uji Shapiro-Wilk maka  $H_0$  ditolak yang berarti asumsi kenormalan residual tidak terpenuhi.

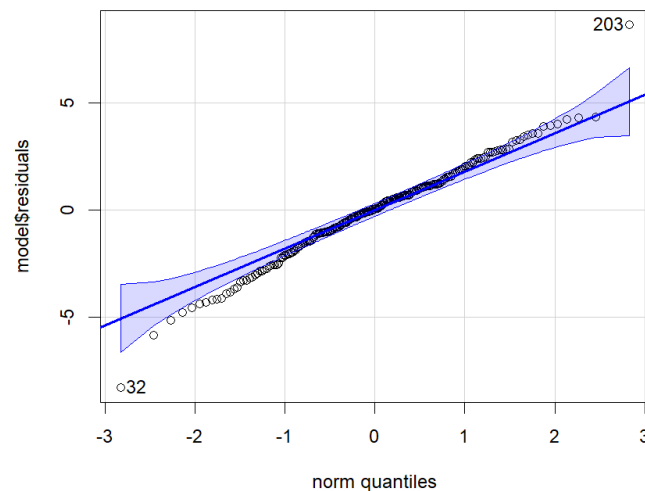
Berdasarkan uji Jarque-Bera dan uji Shapiro-Wilk, kita menemukan bahwa residual (selisih antara nilai aktual dan nilai yang diprediksi oleh model) tidak memenuhi asumsi normalitas. Asumsi normalitas mengasumsikan bahwa data harus mengikuti distribusi normal. Namun, dalam konteks analisis time series, ketidaknormalan pada residual tidak selalu menjadi masalah yang signifikan.

Dalam beberapa kasus, pelanggaran terhadap asumsi normalitas pada residual tidak menjadi masalah karena kita dapat mengandalkan Central Limit Theorem (CLT). CLT menyatakan bahwa ketika kita memiliki banyak data yang cukup, distribusi hasil yang dihasilkan oleh variabel acak cenderung mendekati distribusi normal. Dalam konteks time series, karena kita sering kali memiliki jumlah pengamatan yang besar, kita dapat mengasumsikan bahwa CLT berlaku.

Oleh karena itu, meskipun hasil uji Jarque-Bera dan uji Shapiro-Wilk menunjukkan ketidaknormalan pada residual, kita masih dapat mengandalkan analisis time series dan interpretasi hasilnya. Namun, penting untuk diingat bahwa ketidaknormalan pada residual dapat mempengaruhi interpretasi beberapa statistik atau uji hipotesis tertentu.

yang membutuhkan asumsi normalitas. Dalam hal ini, ada alternatif metode atau pendekatan yang dapat digunakan untuk mengatasi ketidaknormalan tersebut, seperti melakukan transformasi data atau menggunakan metode yang lebih robust terhadap asumsi distribusi. Untuk lebih jelasnya perhatikan Q-Q plot dari residual model berikut

```
> qqPlot(model$residuals)
```



Berdasarkan output dari Q-Q plot terlihat bahwa residual dari model mendekati distribusi normal.

#### 4.6.2. *Overfitting* Model

Pada tahap ini, akan dilakukan analisis *overfitting* pada model ARIMA (1,1,2), dimana terjadi penambahan orde AR(p) sebanyak satu dan model ARIMA (0,1,3) dimana terjadi penambahan orde MA(q) sebanyak satu dari model yang telah diajukan sebelumnya yaitu ARIMA (0,1,2).

- *Overfitting* Model ARIMA (1,1,2)

Menggunakan bantuan R, didapatkan hasil sebagai berikut

```

> #overfitting model
> #menambahkan order AR
> overfit1 <- Arima(temp, order=c(1,1,2))
> model ;overfit1
Series: temp
ARIMA(0,1,2)

Coefficients:
      ma1      ma2
    0.0384 -0.3590
s.e.  0.0671  0.0715

sigma^2 = 4.82: log likelihood = -468.87
AIC=943.75 AICc=943.86 BIC=953.83
Series: temp
ARIMA(1,1,2)

Coefficients:
      ar1      ma1      ma2
    -0.0917  0.1247 -0.3464
s.e.   0.2608  0.2552  0.0845

sigma^2 = 4.841: log likelihood = -468.81
AIC=945.63 AICc=945.82 BIC=959.07

```

Dari hasil di atas, terlihat bahwa nilai estimasi parameter  $\theta_1$  dan  $\theta_2$  dari model ARIMA (1,1,2) tidak berubah signifikan. Begitu pun dengan nilai log likelihood, AIC, dan BIC yang tidak berubah signifikan.

Untuk memastikan lebih lanjut, akan dilakukan uji-t terhadap parameter  $\Phi_1$  untuk memastikan apakah penambahan orde AR diperlukan atau tidak.

- Hipotesis:

$$H_0: \Phi_1 = 0$$

$$H_1: \Phi_1 \neq 0$$

- Taraf Signifikansi:

$$\alpha = 0.05$$

- Statistik Uji

$$t = \frac{\hat{\Phi}_1}{SE} = \frac{-0.0917}{0.2608} = -0.35161$$

Akan dihitung nilai p-valuenya menggunakan bantuan R.

```

> #uji t overfitting ARIMA(1,1,2)
> stat_uji_1 <- overfit1$coef[['ar1']]/0.2608
> stat_uji_1
[1] -0.3515494
> df <- length(temp)-1
> df
[1] 213
> daerah_kritis <- qt(0.025, df)
> daerah_kritis
[1] -1.971164
> stat_uji_1;daerah_kritis
[1] -0.3515494
[1] -1.971164
> 2*(pt(stat_uji_1, df))
[1] 0.7255238

```

- Aturan Keputusan  
 $H_0$  ditolak jika  $p - value < \alpha$
- Kesimpulan
- Karena  $p\text{-value} = 0.7255238 > 0.05$  maka  $H_0$  tidak ditolak yang berarti parameter  $\phi_1$  tidak signifikan terhadap 0. Dengan demikian, penambahan orde AR tidak diperlukan.
- *Overfitting* Model ARIMA (0,1,3)

Menggunakan bantuan R, didapatkan hasil sebagai berikut

```
> #menambahkan order MA
> overfit2 <- Arima(temp, order=c(0,1,3))
> model;overfit2
Series: temp
ARIMA(0,1,2)

Coefficients:
      ma1      ma2
    0.0384  -0.3590
s.e.  0.0671   0.0715

sigma^2 = 4.82:  log likelihood = -468.87
AIC=943.75  AICC=943.86  BIC=953.83
Series: temp
ARIMA(0,1,3)

Coefficients:
      ma1      ma2      ma3
    0.0359  -0.3554   0.0157
s.e.  0.0680   0.0735   0.0661

sigma^2 = 4.842:  log likelihood = -468.85
AIC=945.69  AICC=945.89  BIC=959.14
```

Dari hasil di atas, terlihat bahwa nilai estimasi parameter  $\theta_1$  dan  $\theta_2$  dari model ARIMA (0,1,3) tidak berubah signifikan. Begitu pun dengan nilai log likelihood, AIC, dan BIC yang tidak berubah signifikan.

Untuk memastikan lebih lanjut, akan dilakukan uji-t terhadap parameter  $\theta_3$  untuk memastikan apakah penambahan orde MA diperlukan atau tidak.

- Hipotesis:  
 $H_0: \theta_3 = 0$   
 $H_1: \theta_3 \neq 0$
- Taraf Signifikansi:  
 $\alpha = 0.05$
- Statistik Uji

$$t = \frac{\hat{\theta}_3}{SE} = \frac{0.0157}{0.0661} = 0.237519$$

Akan dihitung nilai p-valuenya menggunakan bantuan R

```
> #uji t overfitting ARIMA(0,1,3)
> stat_uji_2 <- overfit2$coef[['ma3']]/0.0661
> stat_uji_2
[1] 0.2382283
> df <- length(temp)-1
> df
[1] 213
> daerah_kritis <- qt(0.025, df)
> daerah_kritis
[1] -1.971164
> stat_uji_2;daerah_kritis
[1] 0.2382283
[1] -1.971164
> 2*(pt(stat_uji_2, df))
[1] 1.188067
```

- Aturan Keputusan

$H_0$  ditolak jika  $p - value < \alpha$

- Kesimpulan

Karena  $p - value = 1.188067 > 0.05$  maka  $H_0$  tidak ditolak yang berarti parameter  $\theta_3$  tidak signifikan terhadap 0. Dengan demikian, penambahan orde AR tidak diperlukan.

Dengan demikian, dapat disimpulkan bahwa penambahan orde AR ataupun MA tidak diperlukan. Sehingga model ARIMA (0,1,2) merupakan model yang terbaik bagi data.

#### 4.7. Forecasting

*Forecasting* adalah proses memprediksi atau memperkirakan nilai-nilai di masa depan berdasarkan data masa lalu. Hal ini dilakukan dengan menggunakan berbagai metode dan teknik analisis untuk mengidentifikasi pola, tren, dan hubungan dalam data historis guna membuat perkiraan yang akurat.

*Time series analysis* merupakan salah satu metode yang digunakan dalam forecasting. Ini melibatkan analisis data yang terorganisir dalam urutan waktu. Dalam time series analysis, kita memperhatikan pola dan perilaku data seiring waktu, seperti tren, musiman, atau fluktuasi acak. Dengan memahami karakteristik *time series* dan menggunakan model statistik yang sesuai, kita dapat membuat perkiraan yang baik tentang nilai-nilai di masa depan.



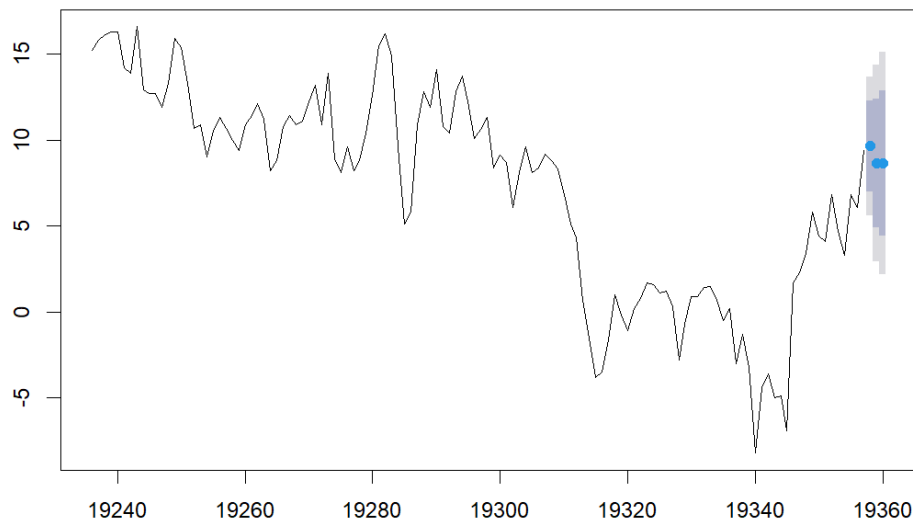
Sebelum kita melakukan peramalan untuk temperatur harian di Warsaw selama 3 hari ke depan, akan dilakukan validasi model terlebih dahulu dengan membagi data menjadi 2 bagian, yaitu data training dan data testing. Dalam hal ini, data training adalah data temperatur harian di Warsaw dari bulan Juni sampai Agustus 2022 yang dimodelkan dengan model terbaik yang sudah kita dapatkan yaitu ARIMA(0,1,2). Sementara itu, data testingnya adalah data temperatur harian di Warsaw dari bulan September sampai Desember 2022. Melalui validasi ini, kita ingin melihat seberapa baik model yang sudah kita tentukan dalam meramal temperatur harian dengan membandingkan kondisi data sebenarnya dan hasil ramalannya.

Berikut akan ditampilkan *output* dari hasil ramalan dan *plot*-nya oleh software R:

```
> #forecasting
> #data testing
> test <- window(temp, start=as.Date(as.character("2022-06-01")),end=as.Date(as.character("2022-08-31")))
> #data training
> train <- window(temp, start=as.Date(as.character("2022-09-01")),end=as.Date(as.character("2022-12-31")))
> trainmodel <- Arima(train, order = c(0, 1, 2))
> #forecasting data training untuk 3 periode ke depan
> testfc <- forecast(trainmodel, h = 3) ; testfc
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
19358	9.636923	7.001922	12.27192	5.607037	13.66681
19359	8.647049	4.913062	12.38104	2.936409	14.35769
19360	8.647049	4.411193	12.88290	2.168867	15.12523

**Forecasts from ARIMA(0,1,2)**



Keterangan sumbu X:

19240 = 9 September 2022

19260 = 25 September 2022

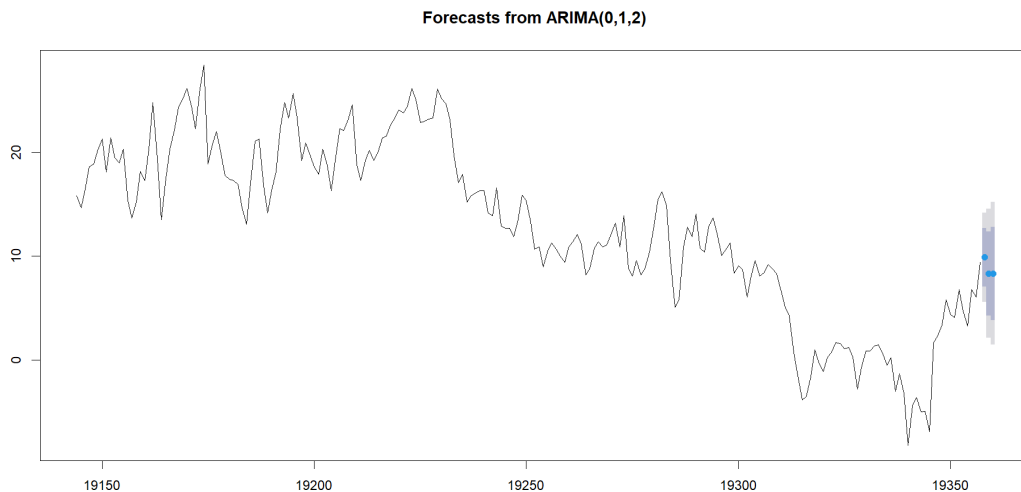
19280 = 15 Oktober 2022

19300 = 4 November 2022  
 19320 = 24 November 2022  
 19340 = 14 Desember 2022  
 19360 = 3 Januari 2023

Kemudian, berikut adalah hasil peramalan data sebenarnya beserta *plot*-nya:

```
> #Forecasting (Periode = 3)
> fc <- forecast(model, h = 3)
> plot(fc)
> fc
```

	Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
19358		9.901904	7.088248	12.71556	5.598788	14.20502
19359		8.345926	4.289597	12.40225	2.142306	14.54955
19360		8.345926	3.861725	12.83013	1.487932	15.20392



Keterangan sumbu X:

19150 = 7 Juni 2022  
 19200 = 27 Juli 2022  
 19250 = 15 September 2022  
 19300 = 4 November 2022  
 19350 = 24 Desember 2022

Dari output di atas, terlihat bahwa hasil ramalan temperatur tidak menunjukkan perbedaan yang signifikan dengan data aktualnya. Kita juga bisa lihat bahwa data aktual masih berada di antara interval forecasting yang kita ramal baik, dengan yang tingkat kepercayaannya sebesar 95%. Oleh karena itu, forecasting yang akan kita lakukan dapat dipercaya karena telah dimiliki akurasi yang baik.

Berdasarkan analisis di atas, didapatkan peramalan temperatur harian di Warsaw untuk 3 periode mendatang, yaitu tanggal 1 Januari 2023, 2 Januari 2023, dan 3 Januari 2023. Dengan menggunakan software R, didapat hasil sebagai berikut:

Periode Ramalan	Bulan	Hasil Peramalan
1	1 Januari 2023	9.901904
2	2 Januari 2023	8.345926
3	3 Januari 2023	8.345926

## BAB V

### PENUTUP

#### 5.1. Kesimpulan

Berdasarkan hasil analisis data runtun waktu temperatur harian di Kota Warsaw periode 1 Juni 2023 sampai 31 Desember 2022, didapat kesimpulan sebagai berikut.

1. Data tersebut merupakan data yang stasioner setelah dilakukan 1 kali *differencing*. Dalam kasus ini, differencing diperlukan untuk mencapai stasioneritas data.
2. Model terbaik yang didapatkan setelah spesifikasi model adalah ARIMA(0, 1, 2) dengan koefisien  $\theta_1 = 0.0384$  dan  $\theta_2 = -0.359$ .

$$W_t = e_t - 0.0384e_{t-1} - 0.359e_{t-2}$$

3. Model tersebut memenuhi asumsi residual yang stasioner, independen, dan distribusi normal. Hal ini menunjukkan bahwa model tersebut mampu menjelaskan variasi dalam data dengan baik.
4. Hasil dari *overfitting* menunjukkan bahwa penambahan orde AR dan MA tidak diperlukan sehingga model ARIMA(0,1,2) yang dipilih merupakan model terbaik.
5. Dari hasil validasi model, dapat ditunjukkan bahwa data aktual masuk ke interval kepercayaan 95% dari ramalan sehingga model cukup baik untuk meramalkan temperatur Warsaw tiga hari ke depan, yaitu tanggal 1 Januari 2023, 2 Januari 2023, dan 3 Januari 2023.
6. Hasil prediksi temperatur Warsaw pada tanggal 1 Januari 2023 adalah 9.901904 derajat Celsius, pada tanggal 2 Januari 2023 adalah 8.345926 derajat Celsius, dan pada tanggal 3 Januari 2023 adalah 8.345926 derajat Celsius.
7. Dengan demikian, analisis data *time series* ini memberikan pemahaman yang lebih baik tentang pola dan peramalan temperatur harian di Kota Warsaw, serta memberikan kesimpulan mengenai model terbaik yang dapat digunakan untuk meramalkan temperatur di masa mendatang.

## DAFTAR PUSTAKA

- Cryer, Jonathan D., Kung-Sik Chan. (2008). *Time Series Analysis With Application in R Second Edition*. Iowa : Springer
- Gupta, S. (2022). *Pre-processing of Time Series Data*. Diakses pada 14 Juni 2023 dari <https://medium.com/enjoy-algorithm/pre-processing-of-time-series-data-c50f8a3e7a98>
- Kowalczyk, M. (2023). *Warsaw Daily Weather Dataset*. Diakses pada 4 Juni 2023 dari <https://www.kaggle.com/datasets/mateuszk013/warsaw-daily-weather>
- Susantri, L., Hasanah, P., Winarni. 2020. *Peramalan Suhu Udara dan Dampaknya Terhadap Konsumsi Energi Listrik di Kalimantan Timur*. Barekeng: Jurnal Ilmu Matematika dan Terapan, 14(3), 397-410.