



UNIVERSITAS INDONESIA

Analisis Runtun Waktu pada Penumpang Pesawat di Changi Airport

Selama Tahun 2009 - 2019

PROJECT UAS METODE PERAMALAN

MATA KULIAH : Metode Peramalan
DOSEN : Dr. Yekti Widyaningsih P, M.Si.

Anggota Kelompok:

Kirono Dwi Saputro	2106656365
Luthfi Athallah Herdita Wiryaman	2206826980
Zavier Raiyan Dana	2206829875

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

PROGRAM STUDI STATISTIKA

DEPOK

JUNI 2024

DAFTAR ISI

DAFTAR ISI	0
Bagian 1. Pendahuluan	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan Penelitian	1
1.4 Manfaat Penelitian	2
Bagian 2. Landasan Teori	3
2.1 Time Series Model	3
2.1.1 Model ARIMA (Autoregressive Integrated Moving Average)	3
2.1.2 Model SARIMA (Seasonal ARIMA)	3
2.2 Uji Stasioneritas	3
2.2.1 Uji ADF (Augmented Dickey-Fuller)	3
2.2.2 Uji KPSS (Kwiatkowski-Phillips-Schmidt-Shin)	3
2.3 Dekomposisi Runtun Waktu	4
2.4 Evaluasi Model	4
2.4.1 AIC (Akaike Information Criterion):	4
2.4.2 BIC (Bayesian Information Criterion):	4
2.4.3 Residual Analysis	4
Bagian 3. Metode Penelitian	5
3.1 Data Penelitian	5
3.1.1 Sumber Data	5
3.1.2 Ukuran Data	5
3.1.3 Metadata	5
3.2 Tahapan Analisis Data	5
Bagian 4. Pre-Processing	6
4.1 Konversi Format Kolom	6
4.2 Pengecekan Missing Value	6
4.3 Visualisasi Data	6
4.3.1 Plot Time Series	6
4.3.2 Box Plot	7
4.3.3 Time Plot $\ln(Y_t)$	7
4.4 Dekomposisi Time Series	7
4.5 Data Splitting	8
4.6 Uji Normalitas	9
4.7 Uji Stasioneritas	9
4.7.1 Uji Dickey-Fuller	10
4.7.2 Uji Kwiatkowski-Phillips-Schmidt-Shin (KPSS)	10
Bagian 5. Hasil dan Pembahasan	12
5.1 Spesifikasi Model	12

5.1.1 (ACF) Autocorrelation Function dan (PACF) Partial Autocorrelation Function	12
5.1.2 (EACF) Extended Autocorrelation Function	13
5.2 Perbandingan Model	13
5.3 Perbandingan AIC dan BIC	14
5.4 Estimasi Parameter	15
5.4.1 Transformasi Box-Cox	15
5.4.2 Uji Signifikansi Estimasi Parameter	16
5.5 Model Diagnostik	16
5.5.1 Uji Independensi (Ljung-Box Test)	17
5.5.2 Uji Normalitas	18
5.6 Forecasting	19
Bagian 6. Penutup	21
6.1 Kesimpulan	21
6.2 Saran dan Rekomendasi	22
DAFTAR PUSTAKA	23
LAMPIRAN	24
Link Google Drive untuk Code R	24

Bagian 1. Pendahuluan

1.1 Latar Belakang

Bandara Changi di Singapura merupakan salah satu hub transportasi udara tersibuk dan terbaik di dunia. Bandara ini tidak hanya melayani penerbangan internasional yang menghubungkan Asia dengan berbagai belahan dunia, tetapi juga berfungsi sebagai pusat transit utama bagi penumpang yang melakukan perjalanan antar benua. Dari tahun 2009 hingga 2019, jumlah penumpang yang melewati Bandara Changi terus meningkat, mencerminkan peningkatan mobilitas global dan peran strategis Singapura dalam jaringan penerbangan internasional.

Namun, seperti halnya bandara besar lainnya, Bandara Changi menghadapi tantangan dalam mengelola fluktuasi jumlah penumpang yang dipengaruhi oleh berbagai faktor seperti musiman, ekonomi global, dan perubahan dalam industri penerbangan. Oleh karena itu, analisis data time series sangat penting untuk memahami pola historis dan memprediksi tren masa depan jumlah penumpang. Dengan pemodelan yang tepat, kami dapat memberikan wawasan yang berharga untuk pengambilan keputusan strategis dan perencanaan operasional bandara.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah yang diangkat dalam penelitian ini adalah:

1. Bagaimana pola jumlah penumpang pesawat bulanan di Bandara Changi dari tahun 2009 hingga 2019?
2. Apakah terdapat komponen musiman yang signifikan dalam data jumlah penumpang pesawat bulanan di Bandara Changi?
3. Model SARIMA apa yang paling cocok untuk memodelkan dan memprediksi jumlah penumpang pesawat bulanan di Bandara Changi?
4. Bagaimana akurasi model SARIMA yang telah dipilih dalam memprediksi jumlah penumpang pesawat pada periode tertentu di Bandara Changi?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengidentifikasi dan menganalisis pola jumlah penumpang pesawat bulanan di Bandara Changi dari tahun 2009 hingga 2019.

2. Menemukan komponen musiman yang signifikan dalam data jumlah penumpang pesawat bulanan di Bandara Changi.
3. Memilih dan mengevaluasi model SARIMA yang paling cocok untuk memodelkan data jumlah penumpang pesawat bulanan di Bandara Changi.
4. Mengevaluasi akurasi model SARIMA yang telah dipilih dalam memprediksi jumlah penumpang pesawat pada periode tertentu dan memberikan rekomendasi berdasarkan hasil analisis tersebut.

1.4 Manfaat Penelitian

1. Bagi Manajemen Bandara Changi:
 - a. Memberikan wawasan yang lebih baik tentang tren dan pola jumlah penumpang pesawat, sehingga dapat digunakan untuk perencanaan operasional yang lebih efektif dan efisien.
 - b. Memungkinkan manajemen untuk mengantisipasi dan mengelola fluktuasi jumlah penumpang, terutama pada periode puncak dan low season.
 - c. Membantu dalam pengambilan keputusan strategis terkait pengembangan infrastruktur dan layanan bandara berdasarkan prediksi yang akurat.
2. Bagi Industri Penerbangan:
 - a. Menyediakan informasi yang berguna bagi maskapai penerbangan untuk merencanakan jadwal penerbangan dan kapasitas pesawat sesuai dengan tren permintaan penumpang.
 - b. Mendukung perencanaan strategis dalam hal penetapan harga tiket dan penawaran promosi berdasarkan pola musiman penumpang.
3. Bagi Peneliti dan Akademisi:
 - a. Menambah literatur dan referensi dalam bidang analisis time series dan pemodelan SARIMA, khususnya dalam konteks transportasi udara.
 - b. Memberikan contoh aplikasi nyata dari metode time series dalam memecahkan masalah praktis di industri penerbangan

Bagian 2. Landasan Teori

2.1 Time Series Model

Model time series digunakan untuk menangkap pola dalam data dan memprediksi nilai masa depan. Beberapa model yang umum digunakan dalam analisis time series meliputi:

2.1.1 Model ARIMA (Autoregressive Integrated Moving Average)

Kombinasi dari model autoregressive (AR) dan moving average (MA) yang juga memperhitungkan perbedaan stasioneritas dalam data. Model ARIMA(p,d,q) dapat dituliskan sebagai:

$$\Delta^d Y_t = c + \phi_1 \Delta^d Y_{t-1} + \dots + \phi_p \Delta^d Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

2.1.2 Model SARIMA (Seasonal ARIMA)

Perpanjangan dari ARIMA yang memperhitungkan komponen musiman dalam data. Model SARIMA(p,d,q)(P,D,Q)[s] dapat dituliskan sebagai:

$$\Phi_P(B^s) \Delta_s^D \phi_p(B) \Delta^d Y_t = \theta_q(B) \Theta_Q(B^s) \epsilon_t$$

2.2 Uji Stasioneritas

Uji stasioneritas digunakan untuk menentukan apakah deret waktu memiliki mean, varians, dan autokorelasi yang konstan sepanjang waktu. Beberapa uji yang umum digunakan meliputi:

2.2.1 Uji ADF (Augmented Dickey-Fuller)

Menguji keberadaan unit root dalam data. Hipotesis nol adalah bahwa data memiliki unit root (tidak stasioner). Persamaan ADF dapat dituliskan sebagai:

$$\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \delta_1 \Delta Y_{t-1} + \dots + \delta_k \Delta Y_{t-k} + \epsilon_t$$

2.2.2 Uji KPSS (Kwiatkowski-Phillips-Schmidt-Shin)

Menguji hipotesis bahwa data adalah stasioner di sekitar mean atau trend linier. Hipotesis nol adalah bahwa data adalah stasioner. Persamaan KPSS dapat dituliskan sebagai:

$$Y_t = \mu + \beta t + \epsilon_t$$

2.3 Dekomposisi Runtun Waktu

Dekomposisi time series adalah teknik untuk memisahkan komponen trend, seasonal, dan irregular dari data. Metode ini membantu dalam memahami struktur data dan meningkatkan akurasi model prediksi. Dekomposisi dapat dilakukan menggunakan pendekatan aditif atau multiplikatif tergantung pada sifat data:

1. Pendekatan Aditif:

$$Y_t = T_t + S_t + E_t$$

2. Pendekatan Multiplikatif:

$$Y_t = T_t \times S_t \times E_t$$

2.4 Evaluasi Model

Evaluasi model time series dilakukan untuk mengukur kinerja dan akurasi model prediksi. Beberapa metrik yang sering digunakan meliputi:

2.4.1 AIC (Akaike Information Criterion):

AIC dikembangkan oleh Hirotugu Akaike dan merupakan kriteria berbasis informasi yang mengukur seberapa baik model statistik dalam memprediksi data. AIC didasarkan pada teori entropi informasi.

$$AIC = 2k - 2 \ln(L)$$

2.4.2 BIC (Bayesian Information Criterion):

BIC, juga dikenal sebagai Schwarz Criterion, dikembangkan oleh Gideon Schwarz. BIC adalah kriteria seleksi model yang didasarkan pada pendekatan Bayesian dan menghitung probabilitas model.

$$BIC = k \ln(n) - 2 \ln(L)$$

2.4.3 Residual Analysis

Mengamati pola dalam sisa-sisa prediksi untuk memastikan bahwa mereka tidak menunjukkan struktur yang dapat diprediksi. Uji Ljung-Box digunakan untuk memeriksa autokorelasi sisa-sisa

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}$$

Bagian 3. Metode Penelitian

3.1 Data Penelitian

Bandara Changi Singapura melayani lebih dari 100 maskapai penerbangan yang terbang ke sekitar 380 kota di sekitar 100 negara dan wilayah di seluruh dunia. Setiap minggu, sekitar 7.400 penerbangan mendarat atau berangkat dari Changi, dengan lebih dari 65,6 juta penumpang melewati bandara ini setiap tahunnya. Dataset ini terdiri dari penumpang yang melewati Bandara Changi Singapura dari tahun 2009 hingga 2019. Dataset ini terdiri dari dua kolom yaitu: Bulan/Tahun dan Jumlah penumpang untuk Bulan/Tahun tersebut.

3.1.1 Sumber Data

[Air Passengers @SGP Changi Airport for past 10yrs \(kaggle.com\)](https://www.kaggle.com/airpassengers)

3.1.2 Ukuran Data

Dataset ini terdiri dari 2 variabel/kolom dengan 130 observasi/baris

3.1.3 Metadata

No	Pengukuran	Tipe	Keterangan	Predikat Variabel
1.	month	object	Bulan-Tahun - Ini mewakili bulan dalam tahun tersebut.	Digunakan
2.	value	integer	Ini mewakili jumlah total penumpang untuk bulan-tahun tersebut.	Digunakan

3.2 Tahapan Analisis Data

Berikut adalah tahapan proses yang dilakukan:

1. Memuat dan Memeriksa Data
2. Pre-Processing Data
3. Pembagian Data menjadi Train dan Test Set
4. Dekomposisi Time Series
5. Uji Stasioneritas
6. Penentuan Orde Model ARIMA
7. Pembuatan Model ARIMA
8. Pemilihan Model Terbaik Berdasarkan AIC dan BIC
9. Forecasting

Bagian 4. Pre-Processing

4.1 Konversi Format Kolom

```
# Konversi kolom 'month' menjadi objek 'yearmon'  
data <- data %>%  
  mutate(month = as.yearmon(month, "%Y-%m"))  
data
```

Mengubah kolom 'month' menjadi format 'yearmon' agar sesuai untuk analisis time series.

4.2 Pengecekan *Missing Value*

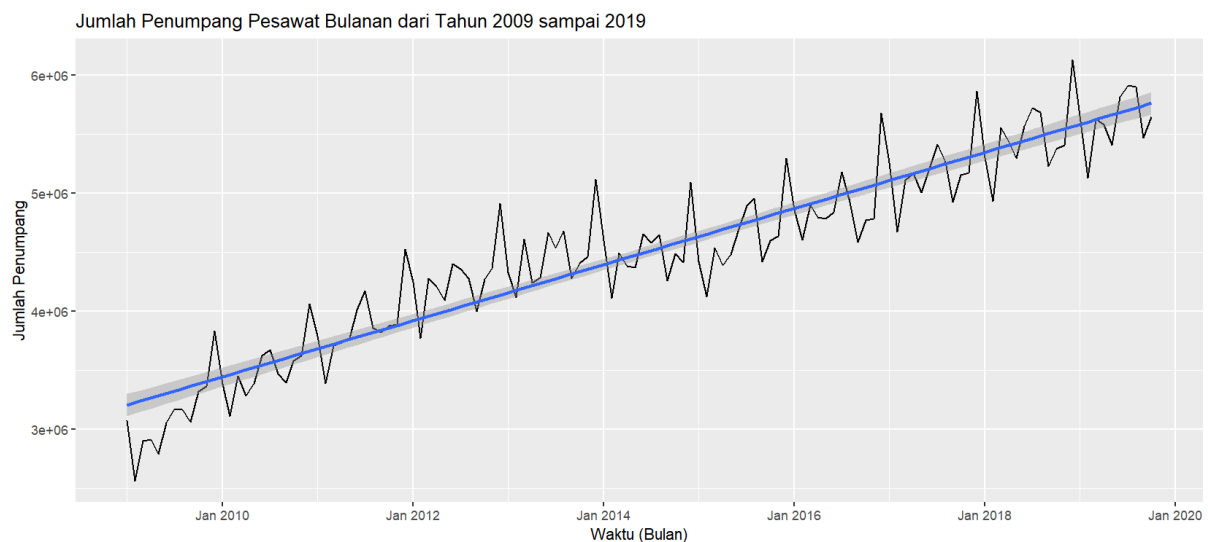
```
> sapply(df, function(x) sum(is.na(x)))  
month passenger  
0 0
```

Terlihat bahwa data kami tidak memiliki missing value.

4.3 Visualisasi Data

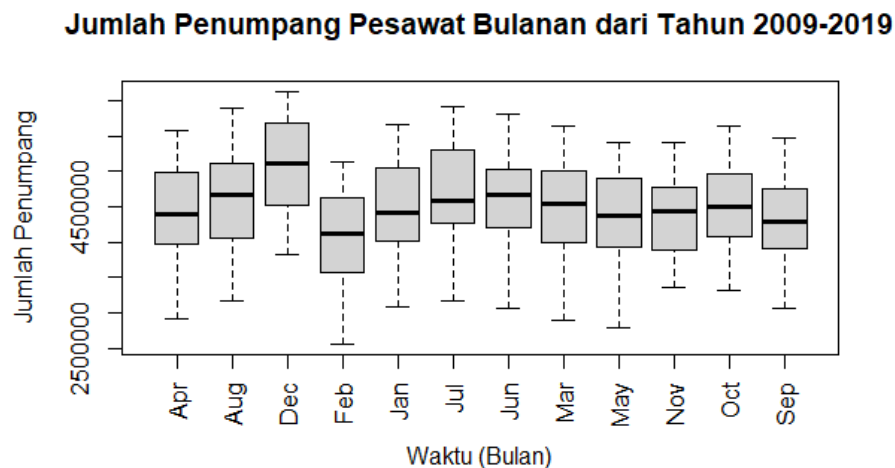
Membuat plot garis dan boxplot untuk memvisualisasikan tren dan pola musiman jumlah penumpang.

4.3.1 Plot Time Series



Plot time series di atas menunjukkan bahwa jumlah penumpang Bandara Changi menunjukkan kecenderungan kenaikan yang signifikan dari tahun 2009 hingga 2019. Dari time plot tersebut, terlihat adanya fluktuasi dalam *mean* dan variansi yang berubah seiring waktu, menunjukkan adanya indikasi ketidakstasioneran dalam data. Selain itu, terdapat pola titik tajam ke atas yang terlihat setiap tahunnya, menandakan adanya fluktuasi musiman atau seasonality dalam jumlah penumpang.

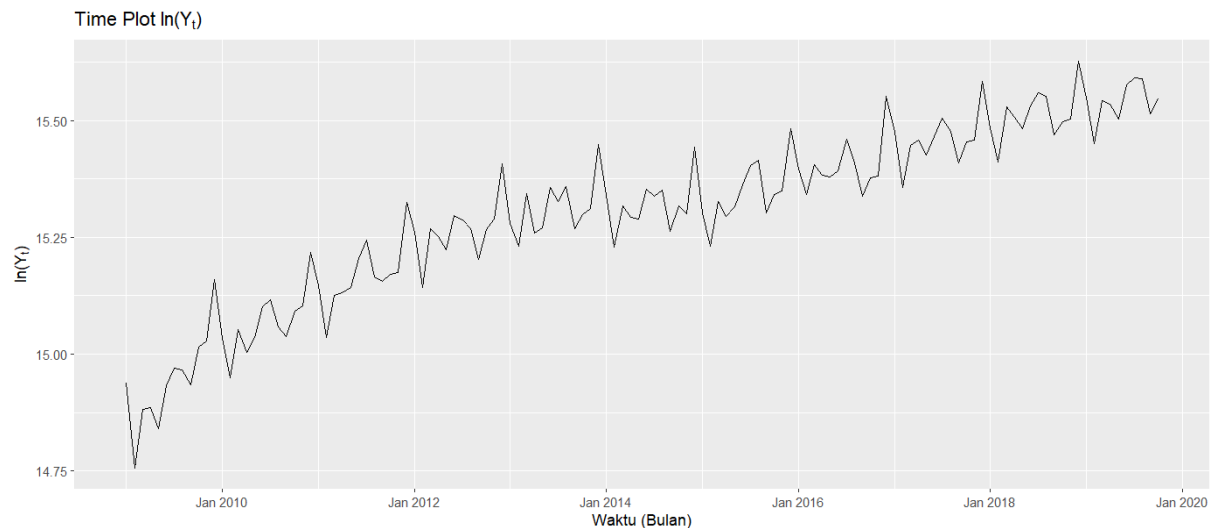
4.3.2 Box Plot



Dari boxplot tersebut dapat dilihat bahwa Bulan Desember memiliki jumlah penumpang terbanyak dari bulan bulan lainnya

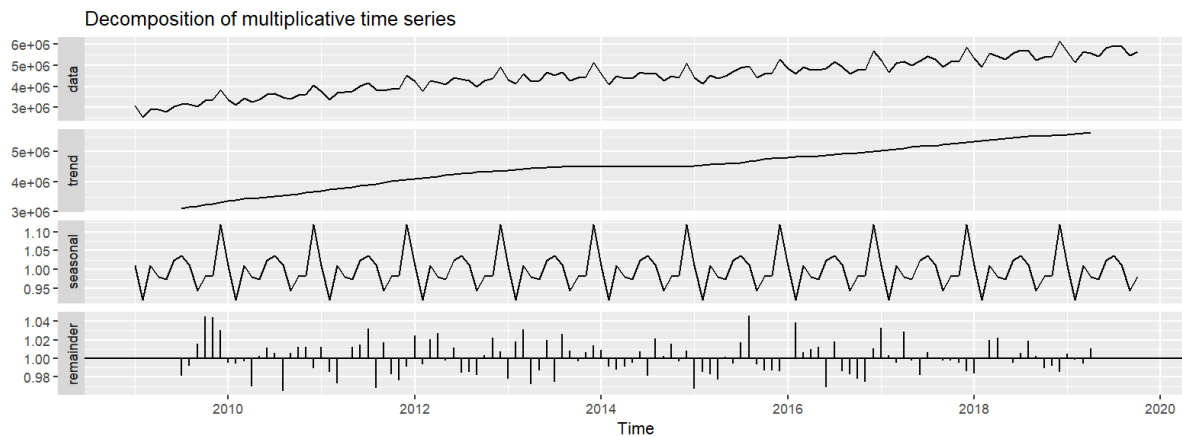
4.3.3 Time Plot $\ln(Y_t)$

Karena runtun waktu Y_t variansinya meningkat seiring bertambahnya waktu, akan dibuat time plot dari transformasi Y_t yaitu $Z_t = \ln(Y_t)$ untuk mengecek variansi menjadi lebih konstan atau tidak.



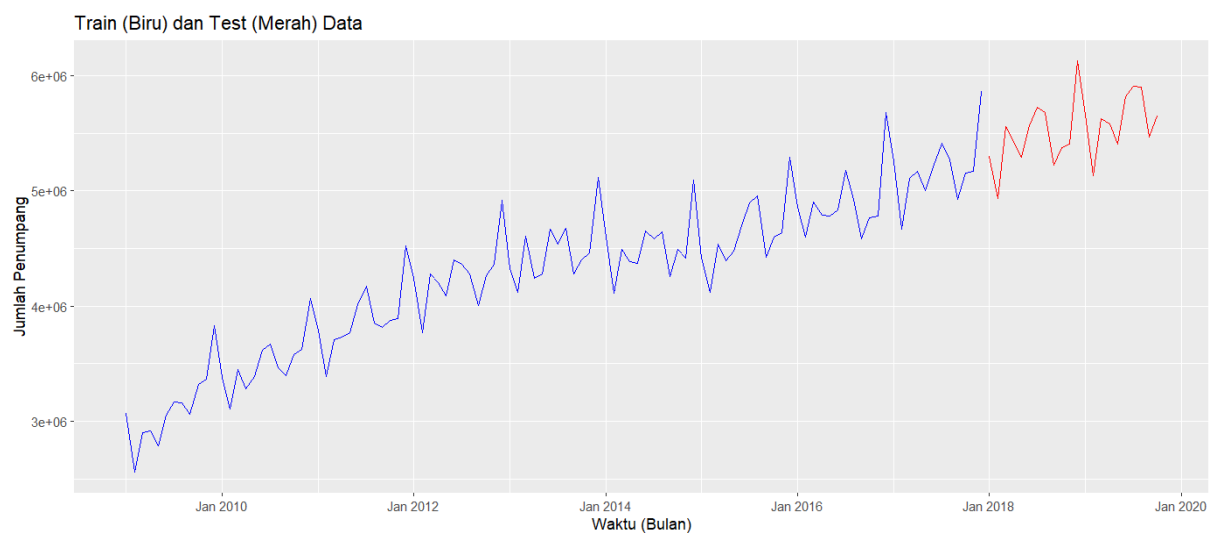
Dapat dilihat bahwa variansi dari runtun waktu $\ln(Y_t)$ lebih konstan dibandingkan dengan runtun waktu asli Y_t .

4.4 Dekomposisi Time Series



Dari plot tersebut dapat dilihat bahwa terdapat pola trend dan musiman yang kuat.

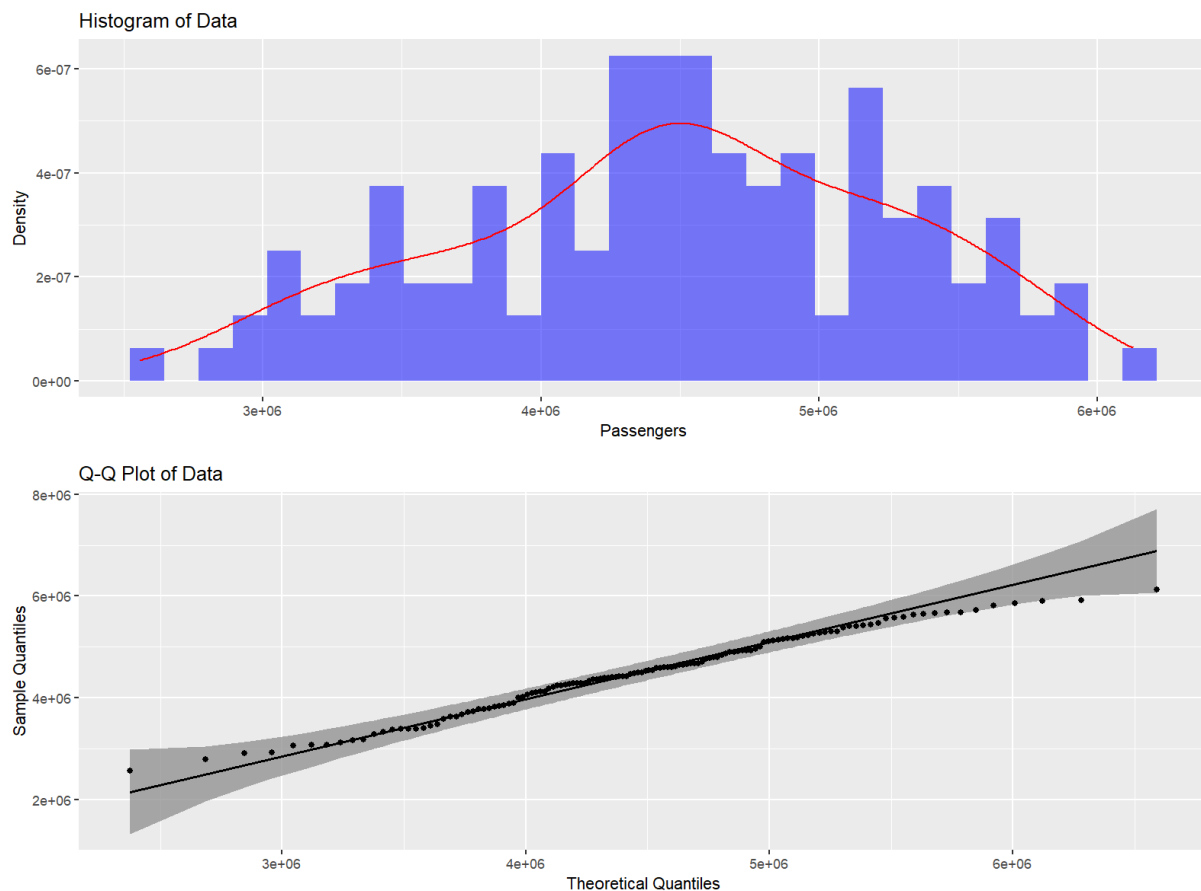
4.5 Data Splitting



Untuk mengukur kinerja model secara objektif, menghindari overfitting, dan mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya, akan dilakukan data splitting. Data akan dibagi menjadi train data dan test data. Untuk data train akan diambil data dari tahun 2009-2017. Adapun untuk data test akan diambil data dari tahun 2018-2019.

4.6 Uji Normalitas

Melalui visualisasi dan uji Shapiro-Wilk akan dilihat apakah data berdistribusi normal atau tidak.



Shapiro-Wilk normality test

```
data: as.numeric(ts_data)
W = 0.98447, p-value = 0.1452
```

Anderson-Darling normality test

```
data: as.numeric(ts_data)
A = 0.47846, p-value = 0.2319
```

Hasil dari uji Shapiro-Wilk dan Anderson-Darling akan memberikan nilai p. Jika nilai p lebih besar dari tingkat signifikansi 0.05, kami gagal menolak hipotesis nol dan menyimpulkan bahwa data mungkin berdistribusi normal. Jika nilai p lebih kecil dari tingkat signifikansi, kami menerima hipotesis nol dan menyimpulkan bahwa data tidak berdistribusi normal. Hasil uji kami menunjukkan $p\text{-value} > 0.05$, artinya kami gagal menolak hipotesis nol dan dapat disimpulkan bahwa data kami berdistribusi normal.

4.7 Uji Stasioneritas

Dari plot hasil visualisasi sebelumnya sekilas kami dapat melihat bahwa datanya tidak stasioner. Untuk memastikan apakah hal ini benar, kami akan melakukan beberapa uji.

4.7.1 Uji Dickey-Fuller

Disini kami akan melakukan uji Dickey-Fuller menggunakan program R.

Hipotesis:

H_0 : (data tidak stasioner/ mempunyai unit root)

H_1 : (data stasioner/ tidak mempunyai unit root)

Tingkat Signifikansi:

$\alpha = 0.05$

Statistik Uji:

```
> print(adf_test)
```

```
Augmented Dickey-Fuller Test
```

```
data: ts_data  
Dickey-Fuller = -3.4031, Lag order = 5, p-value = 0.05717  
alternative hypothesis: stationary
```

Dari hasil uji Dicky-Fuller di atas, p-value = 0.05717 > 0.05. Hal ini membuat kami akan menerima H_0 , sehingga benar bahwa data belum stasioner

4.7.2 Uji Kwiatkowski-Phillips-Schmidt-Shin (KPSS)

Pada bagian pre-processing telah didapat bahwa data tidak stasioner dengan Dickey-Fuller. Akan dilakukan kembali uji stasioneritas dengan Uji Kwiatkowski-Phillips-Schmidt-Shin (KPSS) sebagai perbandingan dari uji Dickey-Fuller. Uji dilakukan pada data train yang telah dipisah pada bagian data splitting.

Hipotesis:

H_0 : (data stasioner/ tidak mempunyai unit root)

H_1 : (data tidak stasioner/ mempunyai unit root)

Tingkat Signifikansi:

$\alpha = 0.05$

Statistik Uji:

```
> kpss.test(train_ts)  
  
KPSS Test for Level Stationarity  
  
data: train_ts  
KPSS Level = 2.1145, Truncation lag parameter = 4, p-value = 0.01
```

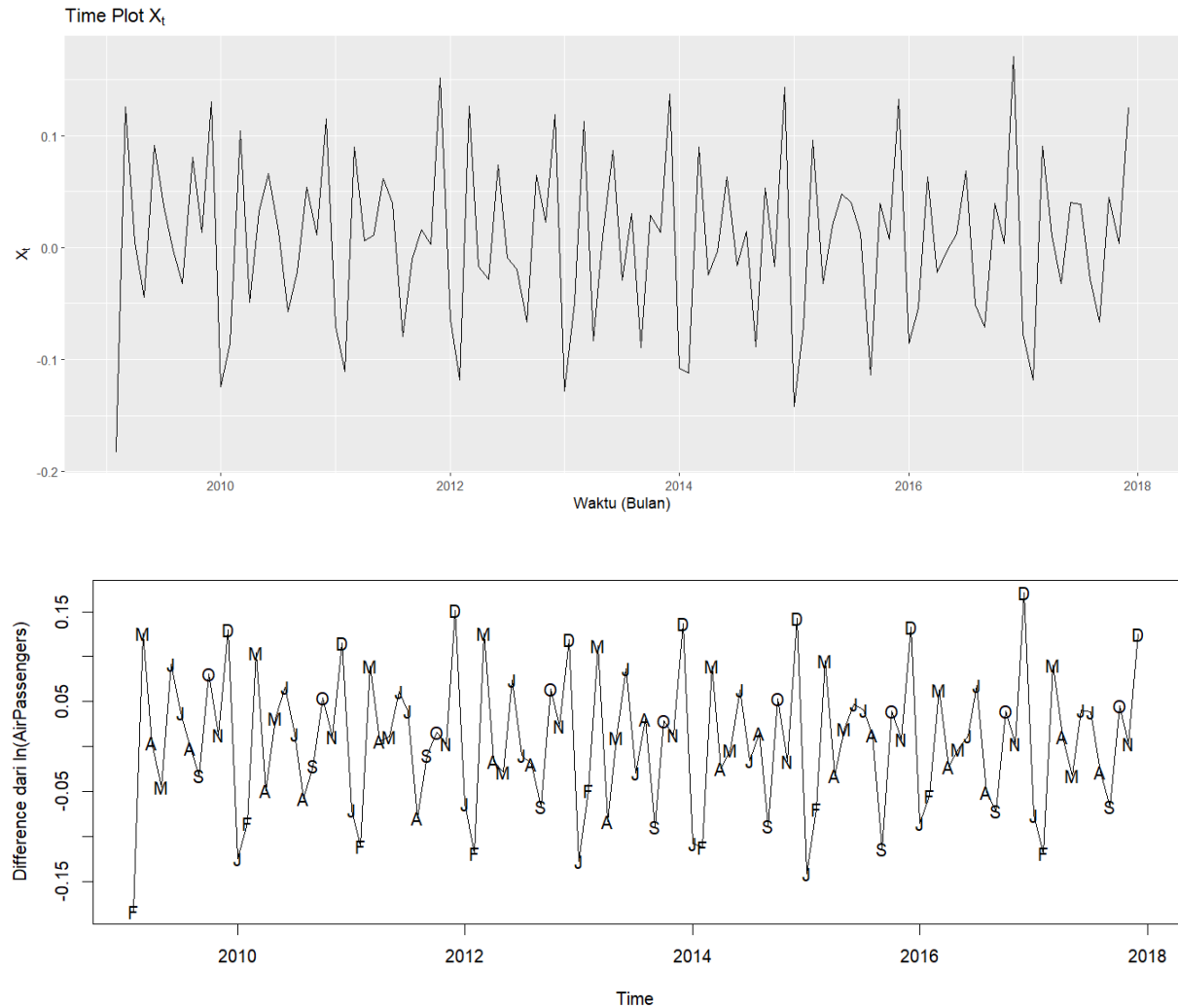
Dari hasil di atas didapat p-value = 0.01 < 0.05. Hal ini mengindikasikan bahwa kami akan menolak H_0 , sehingga benar bahwa data tidak stasioner. Selanjutnya akan dilakukan diferensiasi pertama

```
> kpss.test(diff(log(train_ts), lag = 1))

KPSS Test for Level Stationarity

data: diff(log(train_ts), lag = 1)
KPSS Level = 0.048156, Truncation lag parameter = 4, p-value = 0.1
```

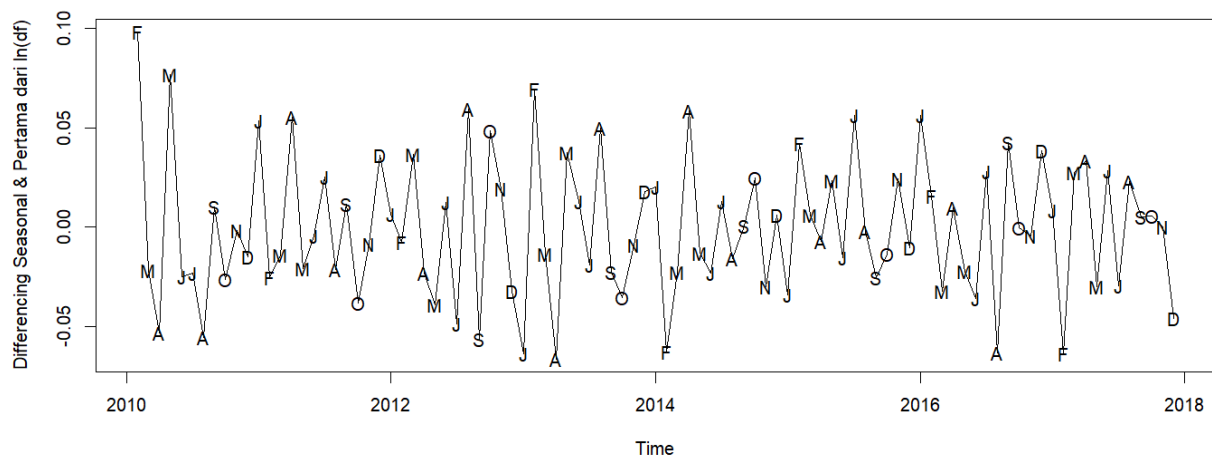
Dari hasil di atas didapat $p\text{-value} = 0.1 > 0.05$. Hal ini mengindikasikan bahwa kami tidak akan menolak H_0 , sehingga datanya sudah stasioner. Selanjutnya akan dilihat plot datanya



Bagian 5. Hasil dan Pembahasan

5.1 Spesifikasi Model

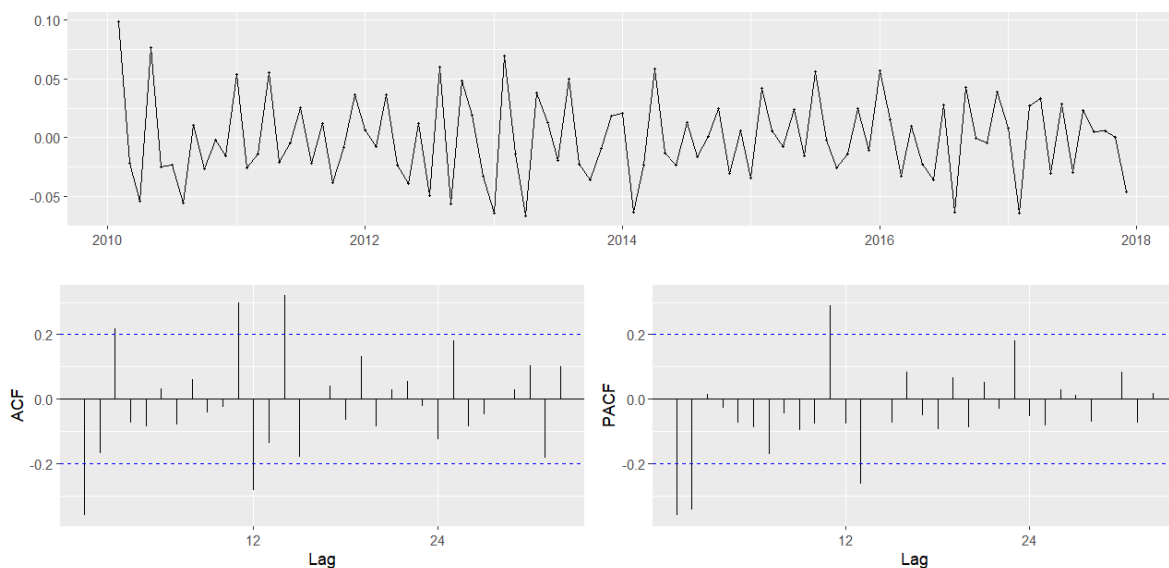
Dari plot sebelumnya kami dapat mendeteksi plot musiman. Bulan Desember dan Maret kebanyakan berada pada titik tertinggi, sedangkan Bulan September dan Februari kebanyakan berada pada titik terendah. Akan dibuat time plot runtun waktu differencing musiman dengan periode musiman $s = 12$ dari differencing pertama $\ln(Y_t)$ untuk mengetahui apakah pola musiman masih terdeteksi atau tidak.



Dapat dilihat bahwa sekarang pola musimannya sudah tidak jelas dibandingkan plot sebelumnya. Oleh karena itu, akan digabungkan differencing regular (nonseasonal) dan differencing seasonal untuk spesifikasi model selanjutnya.

5.1.1 (ACF) Autocorrelation Function dan (PACF) Partial Autocorrelation Function

Akan dibuat plot ACF dan PACF di mana $d = 1$, $D = 1$, dan $s = 12$.



5.1.2 (EACF) Extended Autocorrelation Function

Selanjutnya akan diplot EACF untuk menentukan order model SARIMA

```
> eacf(diff(diff(log(train_ts), lag = 12)))
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x o x o o o o o o o x x o x
1 x x o o o o o o o o o x o x
2 o o x o o o o o o o o o o o
3 x o o o o o o o o o o o o o
4 x o o o o o o o o o o o x o
5 x x o o x o o o o o o o o o
6 x x x o o o o o o o o o o o
7 o x x o o o o o o o o o x o
```

Dalam plot EACF, setiap sel mewakili nilai ACF pada lag tertentu. Pola dalam plot ini dapat membantu mengidentifikasi model ARIMA yang paling sesuai dengan data. Dalam kasus ini, berdasarkan plot EACF didapatkan pendugaan model untuk uji coba yaitu:

- SARIMA(2, 1, 0)(0, 1, 1)₁₂,
- SARIMA(0, 1, 0)(0, 1, 1)₁₂,
- SARIMA(3, 1, 0)(0, 1, 1)₁₂,
- SARIMA(4, 1, 0)(0, 1, 1)₁₂,
- SARIMA(5, 1, 0)(0, 1, 1)₁₂,

5.2 Perbandingan Model

```
> summary(model1)
Series: train_ts
ARIMA(2,1,0)(0,1,1)[12]
Box Cox transformation: lambda= 0

Coefficients:
      ar1      ar2      sma1
    -0.5530  -0.3128  -0.6093
s.e.   0.1022   0.1020   0.1298

sigma^2 = 0.0007796: log likelihood = 205.63
AIC=-403.26  AICC=-402.81  BIC=-393.04

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -15136.79 110367.7 84559.94 -0.3719493 1.971531 0.3126665 0.001655673

> summary(model2)
Series: train_ts
ARIMA(0,1,1)(0,1,1)[12]
Box Cox transformation: lambda= 0

Coefficients:
      ma1      sma1
    -0.5251  -0.5960
s.e.   0.0891   0.1332

sigma^2 = 0.0007994: log likelihood = 204.04
AIC=-402.08  AICC=-401.82  BIC=-394.42

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -16616.71 112178.7 86190.22 -0.4074624 2.013174 0.3186946 -0.046785
```



```

> summary(model3)
Series: train_ts
ARIMA(3,1,0)(0,1,1)[12]
Box Cox transformation: lambda= 0

Coefficients:
      ar1      ar2      ar3      sma1
    -0.5303  -0.2693  0.0812  -0.6221
s.e.   0.1060   0.1168  0.1069   0.1285

sigma^2 = 0.000781: log likelihood = 205.91
AIC=-401.83  AICC=-401.15  BIC=-389.06

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -14240.13 109980.3 84803.4 -0.3515243 1.972711 0.3135667 -0.01812974

> summary(model4)
Series: train_ts
ARIMA(4,1,0)(0,1,1)[12]
Box Cox transformation: lambda= 0

Coefficients:
      ar1      ar2      ar3      ar4      sma1
    -0.5317  -0.2638  0.0931  0.0231  -0.6230
s.e.   0.1062   0.1193  0.1201  0.1056   0.1278

sigma^2 = 0.0007892: log likelihood = 205.94
AIC=-399.88  AICC=-398.92  BIC=-384.55

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -13900.3 110002.8 84748.34 -0.3434422 1.970552 0.3133631 -0.01240721

> summary(model5)
Series: train_ts
ARIMA(5,1,0)(0,1,1)[12]
Box Cox transformation: lambda= 0

Coefficients:
      ar1      ar2      ar3      ar4      ar5      sma1
    -0.5289  -0.2535  0.0626  -0.0540  -0.1567  -0.6262
s.e.   0.1050   0.1186  0.1214  0.1165  0.1062  0.1214

sigma^2 = 0.0007804: log likelihood = 207.01
AIC=-400.02  AICC=-398.73  BIC=-382.14

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -16200.52 108942.3 83431.83 -0.3993963 1.939517 0.3084952 -0.02544633

```

5.3 Perbandingan AIC dan BIC

Untuk menentukan model terbaik akan dilihat AIC dan BIC dari setiap model. Model terbaik adalah model dengan AIC dan BIC terendah

```

> AIC(model1, model2, model3, model4, model5)
      df      AIC
model1  4 -403.2550
model2  3 -402.0792
model3  5 -401.8280
model4  6 -399.8760
model5  7 -400.0167
> BIC(model1, model2, model3, model4, model5)
      df      BIC
model1  4 -393.0395
model2  3 -394.4175
model3  5 -389.0587
model4  6 -384.5527
model5  7 -382.1396

```

Dari perbandingan di atas, model terbaiknya adalah model 1 SARIMA(2, 1, 0)(0, 1, 1)₁₂

5.4 Estimasi Parameter

Berikut summary dari model yang dipilih, SARIMA(2, 1, 0)(0, 1, 1)₁₂

```
> summary(model1)
Series: train_ts
ARIMA(2,1,0)(0,1,1)[12]
Box Cox transformation: lambda= 0

Coefficients:
      ar1      ar2      sma1
    -0.5530  -0.3128  -0.6093
s.e.    0.1022   0.1020   0.1298

sigma^2 = 0.0007796:  log likelihood = 205.63
AIC=-403.26  AICC=-402.81  BIC=-393.04

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -15136.79 110367.7 84559.94 -0.3719493 1.971531 0.3126665 0.001655673
```

Selanjutnya akan dilakukan estimasi parameter untuk model SARIMA(2, 1, 0)(0, 1, 1)₁₂

5.4.1 Transformasi Box-Cox

Transformasi Box-Cox dengan $\lambda=0$ ini setara dengan mengambil logaritma natural dari data. Jadi, kami mulai dengan data asli (Y_t) dan menerapkan transformasi log:

$$Z_t = \ln(Y_t)$$

Differencing Pertama dan Differencing Musiman

$$W_t = Z_t - Z_{t-1} = \ln(Y_t) - \ln(Y_{t-1})$$

$$V_t = W_t - W_{t-12} = (\ln(Y_t) - \ln(Y_{t-1})) - (\ln(Y_{t-12}) - \ln(Y_{t-13}))$$

Model ARIMA

Dari model SARIMA(2, 1, 0)(0, 1, 1)₁₂, maka kami peroleh:

- Untuk ARIMA (2, 1, 0)

$$V_t = \phi_1 V_{t-1} + \phi_2 V_{t-2} + \epsilon_t$$

- Untuk model musiman (0, 1, 1)₁₂

$$V_t = -\theta_{s1} \epsilon_{t-12} + \epsilon_t$$

Kombinasi Model

$$V_t = \phi_1 V_{t-1} + \phi_2 V_{t-2} - \theta_{s1} \epsilon_{t-12} + \epsilon_t$$

Selanjutnya kami substitusi nilai V_t , V_{t-1} , dan V_{t-2} , sehingga model matematis yang kami dapatkan yaitu

$$(\ln(Y_t) - \ln(Y_{t-1})) - (\ln(Y_{t-12}) - \ln(Y_{t-13})) = -0.5530((\ln(Y_{t-1}) - \ln(Y_{t-2})) - (\ln(Y_{t-11}) - \ln(Y_{t-12}))) + -0.3128((\ln(Y_{t-2}) - \ln(Y_{t-3})) - (\ln(Y_{t-10}) - \ln(Y_{t-11}))) - 0.6093\epsilon_{t-12} + \epsilon_t$$

5.4.2 Uji Signifikansi Estimasi Parameter

```
> coeftest(model1)

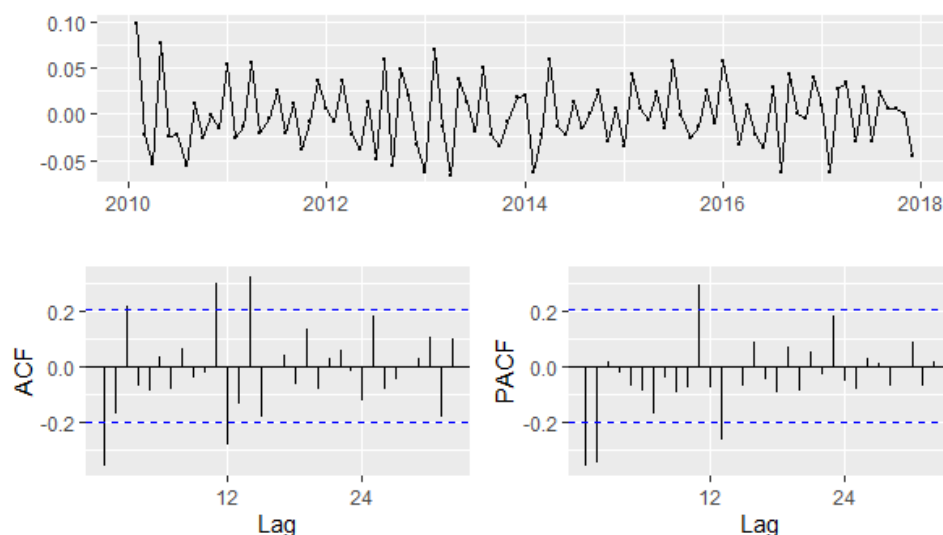
z test of coefficients:

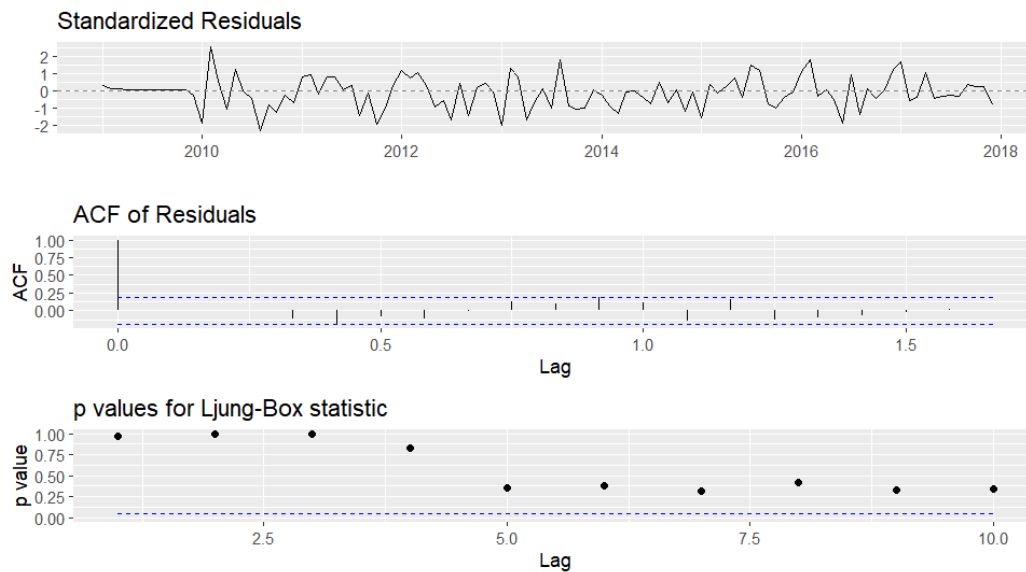
      Estimate Std. Error z value Pr(>|z|)
ar1  -0.55301    0.10218  -5.4119 6.235e-08 ***
ar2  -0.31279    0.10201  -3.0664 0.002166 **
sma1  -0.60927    0.12976  -4.6953 2.662e-06 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Semua koefisien (ar1, ar2, dan sma1) memiliki p-value yang kurang dari 0,05. yang berarti bahwa semua koefisien signifikan. Ini menunjukkan bahwa koefisien-koefisien tersebut secara statistik berbeda dari nol dan memberikan kontribusi yang signifikan terhadap model SARIMA.

5.5 Model Diagnostik

Akan dilakukan uji model diagnostik pada model SARIMA(2, 1, 0)(0, 1, 1)12





1. Plot Standardized Residual

Plot residual menunjukkan nilai residual dari model SARIMA dari waktu ke waktu. Residual ini seharusnya terlihat seperti white noise, artinya tidak ada pola yang terlihat dan residu tersebar acak di sekitar nol. Dari plot ini, residual tersebar secara acak dengan rata-rata mendekati nol, yang mengindikasikan model cukup baik dalam menangkap pola data.

2. ACF of Residuals

Plot ACF menunjukkan autokorelasi residual pada berbagai lag. Jika model SARIMA cocok dengan baik, nilai ACF dari residual seharusnya berada dalam batas signifikan (garis biru putus-putus). Pada plot ini, sebagian besar nilai ACF residual berada dalam batas signifikan, yang menunjukkan tidak ada autokorelasi signifikan yang tersisa dalam residual. Namun, ada beberapa nilai ACF pada lag tertentu yang mendekati batas signifikan, tetapi mereka masih dalam batas wajar untuk model ARIMA.

3. Partial Autocorrelation Function (PACF) of Residuals

Plot PACF menunjukkan autokorelasi parsial residual pada berbagai lag. Sama seperti ACF, jika model cocok dengan baik, nilai PACF dari residual seharusnya berada dalam batas signifikan. Pada plot ini, sebagian besar nilai PACF berada dalam batas signifikan, mengindikasikan bahwa model telah menangkap sebagian besar pola dalam data dan residual tidak memiliki autokorelasi parsial yang signifikan.

5.5.1 Uji Independensi (Ljung-Box Test)

Akan dilakukan uji independensi Ljung-Box Test menggunakan program R

Hipotesis:

H_0 : Residual independen

H_1 : Residual tidak independen

Tingkat Signifikansi:

$\alpha = 0.05$

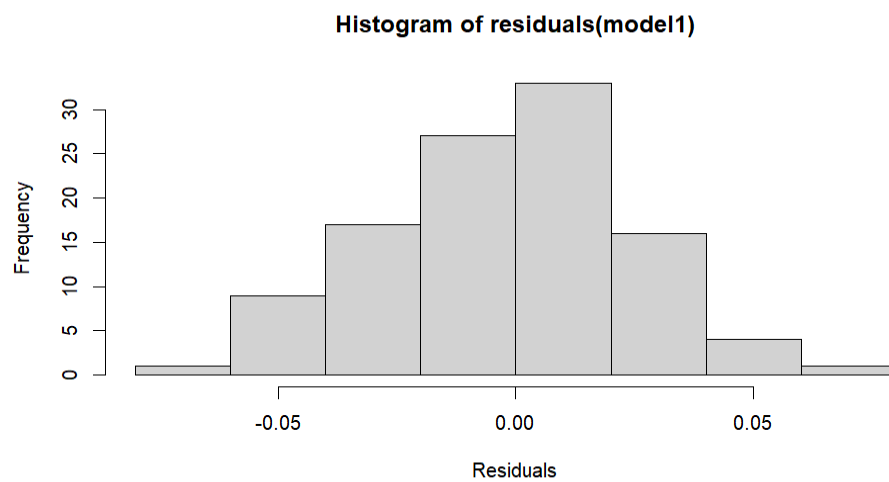
Statistik Uji:

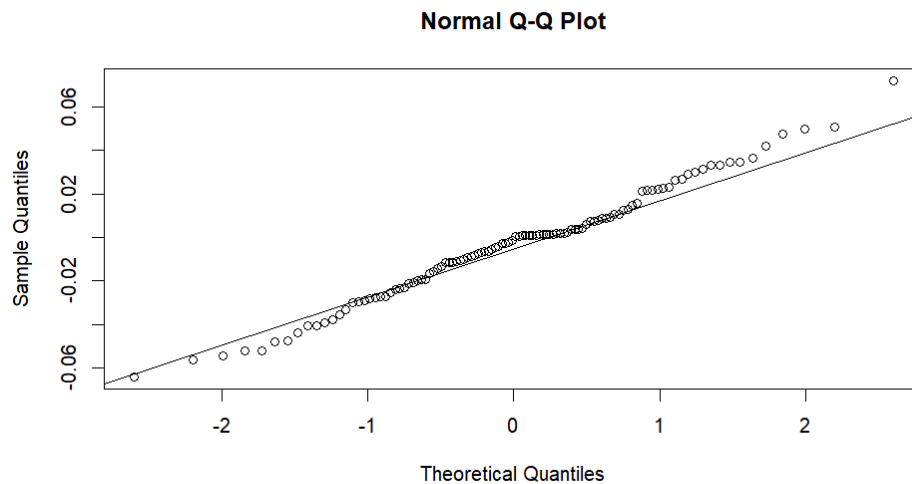
```
> Box.test(residuals(model1), lag = 12, type = "Ljung-Box")  
  
Box-Ljung test  
  
data: residuals(model1)  
X-squared = 16.712, df = 12, p-value = 0.1608
```

Berdasarkan hasil Uji Ljung-Box, kami dapat bahwa p-value nya $0,1608 > 0,05$. Hal ini mengindikasikan bahwa kami tidak menolak H_0 sehingga benar bahwa residual sudah independen

5.5.2 Uji Normalitas

Akan dilihat histogram dan Q-Q Plot dari residual





Dari kedua plot di atas, terlihat bahwa residual belum sepenuhnya memenuhi distribusi normal, terdapat sejumlah penyimpangan kecil dari kedua plot. Untuk memastikan apakah residual berdistribusi normal atau tidak akan dilakukan Uji Shapiro Wilk.

Hipotesis:

H_0 : Residual berdistribusi normal

H_1 : Residual tidak berdistribusi normal

Tingkat Signifikansi:

$\alpha = 0.05$

Statistik Uji:

```
> shapiro.test(residuals(model1))

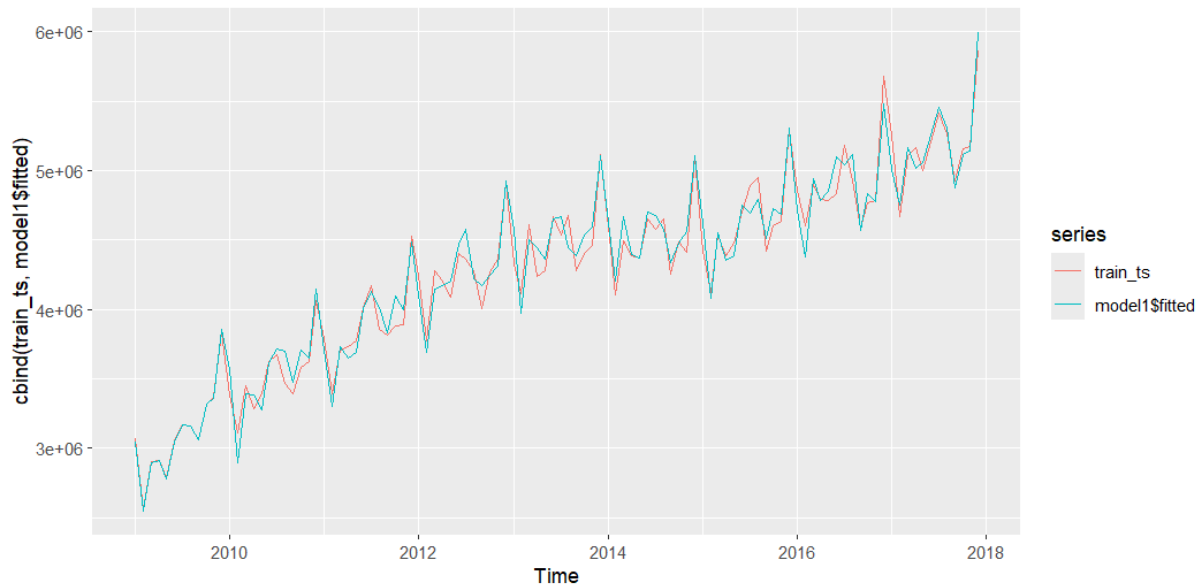
shapiro-wilk normality test

data: residuals(model1)
W = 0.99, p-value = 0.6116
```

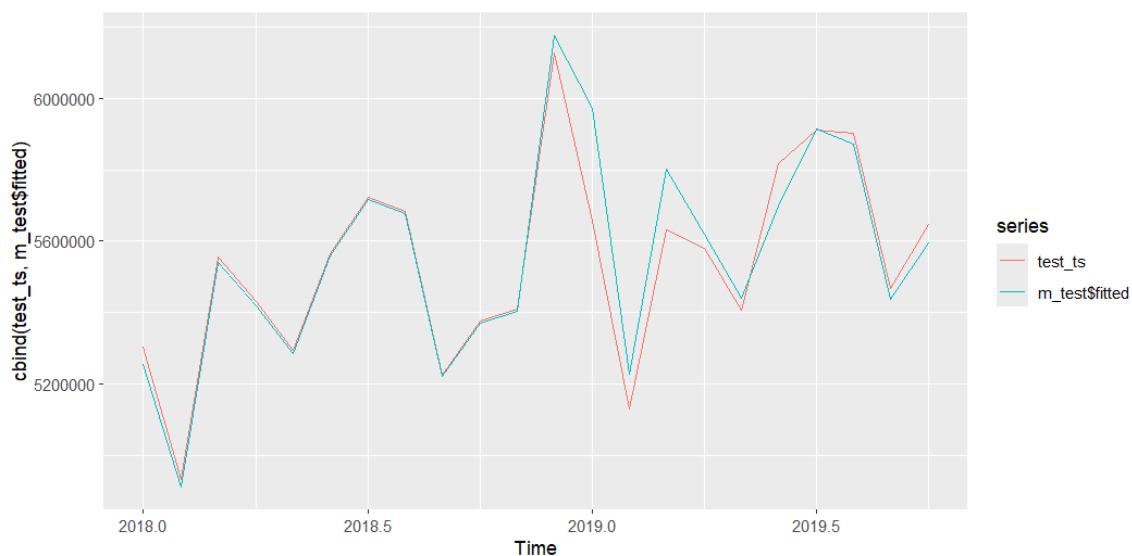
Dari hasil Uji Shapiro Wilk, kami dapat bahwa p-value nya $0,6116 > 0,05$. Hal ini mengindikasikan bahwa kami tidak menolak H_0 sehingga dapat kami simpulkan bahwa residual berdistribusi normal

Dari hasil-hasil uji model diagnostik di atas, didapat bahwa model SARIMA(2, 1, 0)(0, 1, 1)₁₂ sudah cukup sesuai dan selanjutnya akan dilakukan forecasting menggunakan model SARIMA(2, 1, 0)(0, 1, 1)₁₂.

5.6 Forecasting

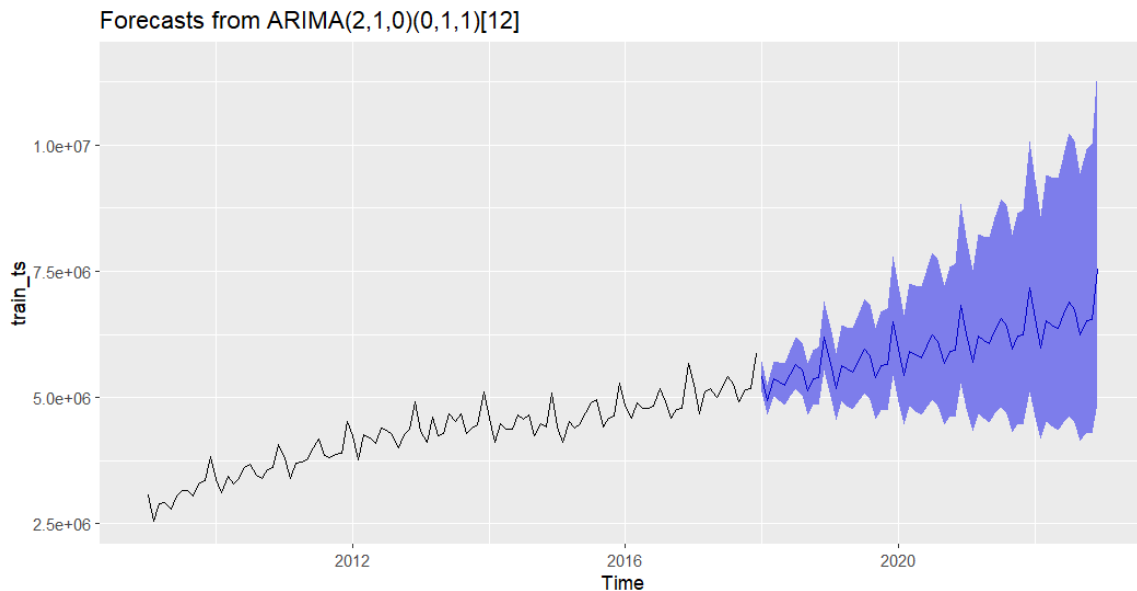


Hasil dari proses forecasting menggunakan model SARIMA(2, 1, 0)(0, 1, 1)[12] menunjukkan performa yang cukup baik. Berdasarkan metrik evaluasi, nilai Mean Absolute Percentage Error (MAPE) sebesar 8.68% menunjukkan bahwa model dapat melakukan prediksi jumlah penumpang dengan akurasi yang tinggi, yaitu rata-rata kesalahan prediksi kurang dari 10% dari nilai sebenarnya. Hal ini mengindikasikan bahwa model SARIMA yang telah dibangun cocok untuk digunakan dalam meramalkan jumlah penumpang pesawat di Bandara Changi pada periode waktu yang akan datang dengan tingkat kesalahan yang rendah.



Hasil dari plot ini menunjukkan bahwa data aktual (test set) dan nilai prediksi dari model SARIMA(2, 1, 0)(0, 1, 1)[12] saling berimpit secara baik. Hal ini mengindikasikan bahwa model yang telah dikembangkan mampu melakukan prediksi dengan baik terhadap

data yang tidak digunakan dalam proses pelatihan (test set). Kesesuaian antara data aktual dan prediksi menunjukkan bahwa model memiliki kemampuan untuk menangkap pola dan variabilitas dari data aktual dengan akurat, sehingga dapat diandalkan untuk melakukan forecasting jumlah penumpang pesawat di Bandara Changi pada periode waktu yang akan datang.



Hasil dari plot forecast menunjukkan prediksi jumlah penumpang pesawat di Bandara Changi untuk periode yang akan datang (60 bulan ke depan) berdasarkan model SARIMA(2, 1, 0)(0, 1, 1)[12]. Garis biru menunjukkan nilai prediksi sedangkan area biru menunjukkan rentang prediksi dengan level kepercayaan 95%.

Bagian 6. Penutup

6.1 Kesimpulan

Analisis time series pada jumlah penumpang pesawat di Changi Airport dari tahun 2009 hingga 2019 telah memberikan sejumlah wawasan mengenai pola dan karakteristik data. Proses analisis melibatkan beberapa langkah, termasuk eksplorasi data, dekomposisi, pemodelan, dan evaluasi model. Berikut adalah kesimpulan dari analisis yang telah dilakukan:

1. **Pola Musiman dan Tren:** Hasil dari analisis menunjukkan adanya pola musiman yang kuat dalam data penumpang pesawat. Peningkatan jumlah penumpang terlihat pada bulan-bulan tertentu, Maret dan Desember, setiap tahun, yang konsisten dengan periode liburan dan musim liburan akhir tahun.
2. **Transformasi dan Diferensiasi:** Transformasi dan diferensiasi digunakan untuk menstabilkan varians dan menghilangkan tren. Hal ini bertujuan agar identifikasi komponen musiman dan tren lebih akurat.
3. **Pemodelan:** Model SARIMA (2,1,0)(0,1,1)[12] dipilih sebagai model terbaik. Model ini berhasil menangkap pola musiman dan tren dalam data dengan baik. Berikut taksiran untuk model terpilih

$$\begin{aligned} &(\ln(Y_t) - \ln(Y_{t-1})) - (\ln(Y_{t-12}) - \ln(Y_{t-13})) = -0.5530((\ln(Y_{t-1}) - \ln(Y_{t-2})) - \\ &(\ln(Y_{t-11}) - \ln(Y_{t-12}))) + -0.3128((\ln(Y_{t-2}) - \ln(Y_{t-3})) - (\ln(Y_{t-10}) - \ln(Y_{t-11}))) \\ &- 0.6093\epsilon_{t-12} + \epsilon_t \end{aligned}$$

4. **Evaluasi Residual:** Semua hasil analisis residual telah memenuhi asumsi diagnostik model. Hal ini menunjukkan bahwa model telah mengakomodasi struktur data dengan baik.
5. **Peramalan:** Model terpilih memiliki keakuratan yang cukup baik dalam memprediksi. Untuk hasil peramalan 5 tahun ke depan, diperoleh hasil bahwa tren data yang naik akan terus berlanjut. Selain itu batas prediksi juga terus melebar seiring berjalannya waktu yang menunjukkan bahwa prediksi semakin tidak pasti seiring berjalannya waktu. Namun secara keseluruhan, model memberikan wawasan yang berguna untuk perencanaan masa depan dengan mempertimbangkan tren dan pola musiman, meskipun dengan tingkat

ketidakpastian yang meningkat seiring waktu

6. **Rekomendasi:** Berdasarkan hasil analisis, model SARIMA $(2,1,0)(0,1,1)[12]$ dapat digunakan untuk membuat prediksi jangka pendek mengenai jumlah penumpang pesawat di Changi Airport. Namun, penting untuk melakukan evaluasi berkala dan menyesuaikan model dengan data terbaru untuk menjaga akurasi prediksi.

6.2 Saran dan Rekomendasi

Berdasarkan kesimpulan analisis time series kami pada jumlah penumpang pesawat di Changi Airport, berikut adalah rekomendasi dari kami:

1. Implementasi Model SARIMA: Disarankan untuk mengimplementasikan Model SARIMA $(2,1,0)(0,1,1)[12]$ sebagai alat prediksi untuk memperkirakan jumlah penumpang pesawat di masa mendatang. Model ini terbukti efektif dalam menangkap pola musiman dan tren yang ada dalam data historis.
2. Pemantauan dan Evaluasi Berkala: Penting untuk melakukan pemantauan dan evaluasi berkala terhadap kinerja model. Hal ini akan membantu dalam mengidentifikasi perubahan pola atau tren baru yang mungkin terjadi, serta memastikan bahwa model tetap relevan dan akurat.
3. Penyesuaian Model dengan Data Terbaru: Model SARIMA perlu disesuaikan dengan data terbaru secara teratur untuk meningkatkan keakuratan prediksi. Penyesuaian ini mencakup pemilihan ulang parameter dan evaluasi ulang struktur model sesuai dengan dinamika baru dalam jumlah penumpang pesawat.
4. Kolaborasi dengan Pihak Terkait: Kerja sama yang erat dengan pihak terkait seperti maskapai penerbangan, otoritas bandara, dan lembaga terkait lainnya penting untuk memperbaiki perencanaan kapasitas, manajemen antrian, dan strategi pelayanan pelanggan di Changi Airport.
5. Inovasi Teknologi dan Analisis Data: Investasi dalam teknologi canggih untuk analisis data, seperti kecerdasan buatan dan big data, dapat meningkatkan kemampuan prediksi dan pengambilan keputusan yang lebih baik di masa depan.

Dengan menerapkan rekomendasi ini, diharapkan Changi Airport dapat terus memperkuat posisinya sebagai salah satu hub penerbangan terkemuka di dunia dengan layanan yang handal dan responsif terhadap perubahan pasar dan kebutuhan penumpang.

DAFTAR PUSTAKA

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: With R examples* (4th ed.). Springer.

Chatfield, C. (2019). *The analysis of time series: An introduction* (7th ed.). Chapman & Hall/CRC.

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd ed.). Springer.

Cowpertwait, P. S. P., & Metcalfe, A. V. (2009). *Introductory time series with R*. Springer.

Cryer, J. D., & Chan, K.-S. (2008). *Time series analysis: With applications in R* (2nd ed.). Springer.

Enders, W. (2014). *Applied econometric time series* (4th ed.). Wiley.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.

Tsay, R. S. (2014). *Analysis of financial time series* (3rd ed.). Wiley.

LAMPIRAN

Link Google Drive untuk Code R dan Video Presentasi

□ Lampiran Code dan Video Project UAS Metode Peramalan_Changi Airport

Kontribusi Anggota Kelompok

No	Nama	NPM	Kontribusi	Tingkat Kontribusi
1	Kirono Dwi Saputro	2106656365	Aktif berdiskusi, mencari data, membantu pengerjaan code, pembuatan ppt dan video	100%
2	Luthfi Athallah Herdita Wiryaman	2206826980	Aktif berdiskusi, mencari data, membantu pengerjaan code, menyusun laporan, pembuatan ppt dan video	100%
3	Zavier Raiyan Dana	2206829875	Aktif berdiskusi, mencari data, membantu pengerjaan code, menyusun laporan, pembuatan ppt dan video	100%