

LAPORAN ANALISIS DERET WAKTU
PEMILIHAN MODEL TERBAIK
UNTUK PERAMALAN DATA SAHAM AAPL



Anggota :

1. Adhystira Raihannoeza (2211110007)
2. Meuthia Evawani Sabilla (2211110021)
3. Nahwan Al Hirzi (2211110046)

Dosen Pengampu :

Aina Latifa Riyana Putri, S.Si., M.Mat.

PROGRAM STUDI S1 SAINS DATA
FAKULTAS INFORMATIKA
TELKOM UNIVERSITY PURWOKERTO

2024

DAFTAR ISI

DAFTAR ISI	iii
DAFTAR TABEL	iv
BAB I	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
BAB II	3
2.1. Model <i>Time Series</i> (Deret Waktu)	3
2.2. Identifikasi Model berdasarkan Plot ACF & PACF	4
2.2.1 Model Autoregressive Integrated Moving Average (ARIMA)	4
2.2.2 Akaike Information Creterion (AIC)	5
BAB III.....	6
3.1 Identifikasi Masalah	7
3.2 Pengumpulan Data	7
3.3 <i>Exploratory Data Analysis</i> (EDA)	8
3.4 Preprocessing.....	9
3.5 Uji Stationeritas (ADF Test)	9
3.7 Pemodelan ARIMA.....	10
3.8 Analisis Hasil	11
3.9 Kesimpulan	11
BAB IV.....	12
4.1 Data Harga Saham Apple Inc. (AAPL)	12
4.2 Exploratory Data Analysis (EDA)	13
4.3 Kestationeran Data	15
4.4 Prediksi Harga Saham AAPL.....	26
BAB V.....	30
5.1 Kesimpulan.....	30
5.2 Saran.....	30

DAFTAR TABEL

Tabel 3.1. Harga saham Apple Inc (AAPL).....	7
Tabel 4.1. Data Saham AAPL.....	12
Tabel 4. 2. Tabel nilai differencing satu kali.....	18
Tabel 4. 3. Tabel Hasil perolehan model ARIMA	21
Tabel 4. 4. Tabel Perolehan model terbaik.....	21
Tabel 4. 5. Transformasi Logaritma.....	22
Tabel 4.6. Perbandingan akurasi ketiga model.....	28

BAB I

PENDAHULUAN

1.1 Latar Belakang

Saham merupakan salah satu investasi yang paling populer di pasar keuangan global. Pergerakan harga saham yang dinamis mencerminkan bagaimana kondisi pasar, sentimen investor, dan berbagai faktor ekonomi lainnya. Salah satu saham yang memiliki perhatian besar dari para investor adalah saham Apple Inc. (AAPL), yang merupakan perusahaan teknologi terbesar di dunia berdasarkan kapitalisasi pasar. AAPL dikenal sebagai indikator penting dari sektor teknologi, sehingga memprediksi pergerakan saham ini menjadi tantangan yang menarik sekaligus signifikan.

Peramalan data saham ini adalah bagian penting dalam pengambilan keputusan investasi. Dalam hal ini, analisis deret waktu menjadi metode utama untuk memahami pola historis dan memprediksi pergerakan harga saham di masa depan. Namun dalam pemilihan model terbaik untuk analisis deret waktu memerlukan pendekatan khusus dengan teliti dan sistematis, karena setiap model memiliki kekuatan dan kelemahan yang berbeda-beda. Beberapa model yang sering digunakan untuk memprediksi masa depan yaitu *Autoregressive Integrated Moving Average (ARIMA)*.

Pemilihan model terbaik tidak hanya penting untuk memberikan hasil prediksi yang lebih akurat, tetapi juga pemilihan model terbaik digunakan untuk memahami karakteristik unik dari data saham AAPL. Oleh karena itu, laporan ini bertujuan untuk menganalisis data historis saham AAPL, mengevaluasi berbagai model deret waktu, dan memilih model terbaik berdasarkan kriteria akurasi prediksi. dengan pendekatan ini, laporan ini diharapkan dapat memberikan wawasan yang bermanfaat bagi investor maupun akademisi dalam memahami pergerakan saham AAPL dan dapat memperkuat strategi pengambilan keputusan pada pasar saham.

1.2 Rumusan Masalah

1. Model apa yang paling tepat untuk menganalisis deret waktu harga saham Apple (AAPL) dengan variabel "Terakhir" sebagai fokus utama?
2. Apakah metode ARIMA dapat menghasilkan prediksi yang akurat untuk prediksi harga saham Apple (AAPL)?
3. Seberapa akurat hasil peramalan yang dihasilkan oleh model terbaik tersebut?

1.3 Tujuan

1. Menentukan model ARIMA yang paling sesuai untuk memprediksi harga saham Apple (AAPL)
2. Memprediksi harga saham Apple (AAPL)
3. Mengevaluasi performa model ARIMA dalam menghasilkan prediksi harga saham Apple (AAPL)

BAB II

DASAR TEORI

2.1. Model *Time Series* (Deret Waktu)

Model *time series* merupakan representasi statistik dari sebuah pengamatan sesuatu yang diambil dengan cara berurutan dalam interval waktu yang sama. Dengan model *time series* tersebut dapat dilakukan sebuah analisis untuk mempelajari tren data dari waktu ke waktu dengan memeriksa ketergantungan sekuensial dan temporal dalam suatu kumpulan data.

Adapun rumus *Autocorrelation Function* (ACF) yang dapat dilihat pada persamaan berikut:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \text{Corr}(X_t, X_{t+k}) \quad (1)$$

$$\gamma_k = \text{Cov}(X_t, X_{t+k})$$

Dengan keterangan:

ρ_k = nilai ACF pada lag ke - k

γ_k = kovarian pada lag ke - k

γ_0 = varian

Selain itu adapun *Partial Autocorrelation Function* (PACF) yang digunakan untuk menghitung tingkat keeratan antara X_t dan X_{t+k} , yang mana pengaruh masing-masing lag dianggap berbeda. Berikut ini adalah persamaan *Partial Autocorrelation Function* (PACF):

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{kj-1} \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{kj-1} \rho_{k-j}}$$

$$\phi_{kk} = \phi_{k-1,j} - \phi_{kk} \phi_{k-1,j-1}$$

Untuk :

J : 1,2,...,k-1

Dengan :

ϕ_{kk} = nilai PACF pada lag ke-k k = time lag

2.2. Identifikasi Model berdasarkan Plot ACF & PACF

Pada data *time series* fungsi *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) digunakan untuk menentukan model ARIMA yang sesuai. ACF digunakan untuk melihat korelasi antara observasi peristiwa pada suatu waktu dengan observasi pada waktu sebelumnya, sedangkan PACF digunakan untuk mengukur korelasi antara observasi pada suatu waktu dengan observasi pada waktu sebelumnya dengan menghilangkan korelasi melalui waktu yang lebih pendek.

Dalam proses memilih model yang tepat dapat dilakukan dengan mengidentifikasi orde AR dan MA pada grafik ACF serta PACF sebagai berikut:

1. Jika autokorelasi secara eksponens melemah mendekati nol dan autokorelasi parsial terputus setelah lag ke p artinya terjadi proses AR (p).
2. Jika autokorelasi terputus setelah lag ke p dan autokorelasi parsial melemah secara eksponensial mendekati nol artinya terjadi proses MA (q)
3. Jika keduanya melemah artinya terjadi proses ARIMA ($p, 0, q$)

2.2.1 Model Autoregressive Integrated Moving Average (ARIMA)

Model *autoregressive* merupakan sebuah model yang menyatakan bahwa nilai data saat ini dipengaruhi oleh nilai data dari periode sebelumnya. Model ini biasanya dinotasikan dengan AR (p), di mana p merupakan ordo model tersebut. Kemudian, model *Moving Average* merupakan sebuah model yang menggambarkan hubungan antara nilai pengamatan kesalahan peramalan pada saat ini dengan kesalahan peramalan dari periode sebelumnya. Model ini biasanya dinotasikan sebagai MA(q), di mana q merupakan ordo model.

Model *Autoregressive Moving Average* (ARMA) merupakan sebuah model yang mengkombinasikan model AR(p) dan MA(q). Model ini mengasumsikan bahwa nilai data saat ini dipengaruhi oleh nilai data dari periode sebelumnya serta nilai kesalahan dari periode sebelumnya. Model ARMA dengan ordo p dan q biasanya dinotasikan sebagai ARMA(p, q). Menurut Mulyono, bentuk umum model ARIMA adalah sebagai berikut:

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

Dengan:

Z_t = Variabel *Time Series*

μ = konstanta

$\theta_1, \theta_2, \dots, \theta_q$ = Koefisien parameter *moving average ke - q*

$\phi_1, \phi_2, \dots, \phi_p$ = koefisien parameter *agressive ke - p*

e_{t-q} = sisaan pada saat ke $t - q$

2.2.2 Akaike Information Creterion (AIC)

AIC adalah metode statistik yang digunakan untuk memilih model terbaik dari sekumpulan model kandidat dengan estimasi kesalahan prediksi dan kualitas relatif model. AIC dihitung dengan rumus :

$$AIC = -2l + 2p$$

Dengan keterangan :

l = *log likelihood*

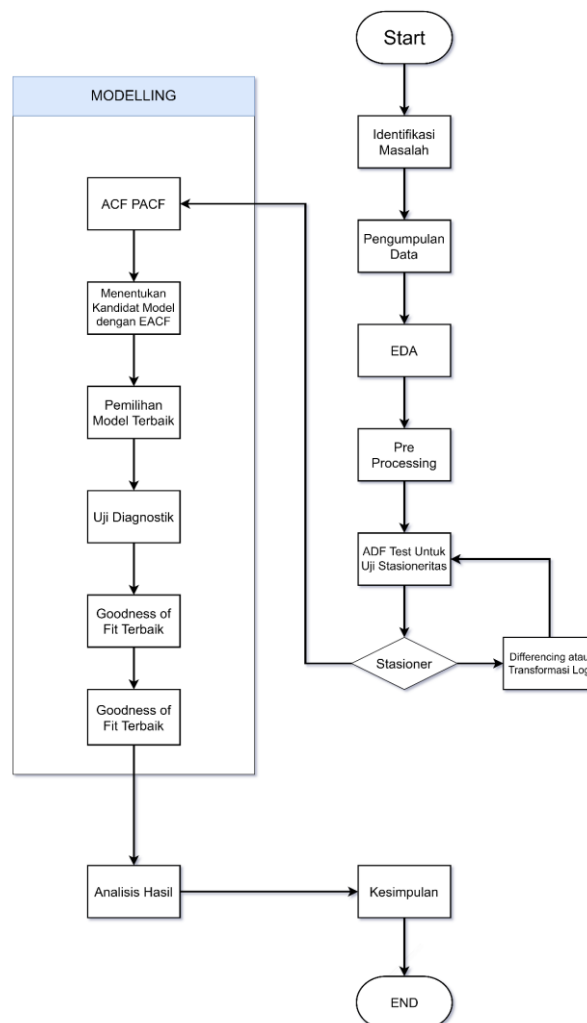
p = *jumlah parameter*

Nilai AIC bergantung pada *log-likelihood* dari suatu fungsi kepadatan peluang. Nilai AIC yang dikenal dapat digunakan sebagai acuan untuk memilih metode terbaik dalam mengestimasi parameter. AIC mengutamakan model dengan kecocokan yang baik (nilai likelihood tinggi) tetapi juga memberikan hambatan untuk jumlah parameter yang besar guna mencegah *overfitting*. Model dengan nilai AIC terendah dipilih karena itu berarti model tersebut memiliki keseimbangan yang baik antara kecocokan data dan tingkat kerumitan model, sehingga kehilangan informasi yang terjadi lebih sedikit dibandingkan dengan model-model lain dalam kumpulan data. Meskipun AIC berguna untuk membandingkan model secara relatif, AIC tidak memberikan penilaian mutlak terhadap kualitas model.

BAB III

METODE PENELITIAN

Dalam menganalisis data runtun waktu perlu dilakukan pemilihan model terlebih dahulu dimana model yang digunakan bergantung terhadap karakteristik atau sifat data yang digunakan. Penelitian ini menggunakan jenis data sekunder dengan pendekatan kualitatif. Sumber data yang digunakan pada penelitian ini merupakan data saham AAPL dari 3 Januari 2022 sampai 1 November 2024 yang didapatkan dari situs id.investing.com yang berisikan harga jual beli saham. Metode analisis yang digunakan pada penelitian ini adalah analisis deret waktu yang mengolah data dengan metode peramalan (*forecasting*) terhadap harga saham AAPL. Dalam tahap analisis deret waktu, metode ARIMA akan digunakan untuk memodelkan data historis harga saham AAPL dengan tujuan memprediksi tren harga saham hingga beberapa hari mendatang.



Gambar 3.1. Diagram Alir Penelitian

Berikut adalah serangkaian langkah yang disajikan dalam diagram alir penelitian dengan tujuan memberikan panduan yang jelas :

3.1 Identifikasi Masalah

Tahap awal dilakukan identifikasi masalah, yaitu prediksi harga saham AAPL dari 3 Januari 2022 sampai 1 November 2024. Proses identifikasi masalah melibatkan pemahaman pola pertumbuhan harga saham dan pencarian model ARIMA yang tepat.

3.2 Pengumpulan Data

Pengumpulan data untuk penelitian ini diambil dari situs id.investing.com yaitu data harga saham Apple Inc (AAPL) yang tercatat mulai dari 3 Januari 2022 sampai 1 November 2024. Berikut merupakan tampilan data yang akan digunakan untuk prediksi model Arima

Tabel 3.1. Harga saham Apple Inc (AAPL)

Tanggal	Terakhir	Pembukaan	Tertinggi	Terendah	Volume	Perubahan
2022-01-03	182,01	177,83	182,88	177,71	104,70M	2,50%
2022-01-04	179,70	182,63	182,94	179,12	99,31M	-1,27%
2022-01-05	174,92	179,61	180,17	174,64	94,54M	-2,66%
2022-01-06	172,00	172,70	175,30	171,64	96,90M	-1,67%
2022-01-07	172,17	172,89	174,14	171,03	86,71M	0,10%
2022-01-10	172,19	169,08	172,50	168,17	106,77M	0,01%
2022-01-11	175,08	172,32	175,18	170,82	76,14M	1,68%
2022-01-12	175,53	176,12	177,18	174,82	74,81M	0,26%
2022-01-13	172,19	175,78	176,62	171,79	84,51M	-1,90%
.
.
.

Tanggal	Terakhir	Pembukaan	Tertinggi	Terendah	Volume	Perubahan
2022-01-03	182,01	177,83	182,88	177,71	104,70M	2,50%
2022-01-04	179,70	182,63	182,94	179,12	99,31M	-1,27%
2022-01-05	174,92	179,61	180,17	174,64	94,54M	-2,66%
2022-01-06	172,00	172,70	175,30	171,64	96,90M	-1,67%
2022-01-07	172,17	172,89	174,14	171,03	86,71M	0,10%
2022-01-10	172,19	169,08	172,50	168,17	106,77M	0,01%
2024-10-21	236,48	234,45	236,85	234,45	36,25M	0,63%
2024-10-22	235,86	233,88	236,22	232,60	38,85M	-0,26%
2024-10-23	230,76	234,08	235,14	227,76	52,29M	-2,16%
2024-10-24	230,57	229,98	230,82	228,41	31,11M	-0,08%
2024-10-25	231,41	229,74	233,22	229,57	38,80M	0,36%
2024-10-28	233,40	233,32	234,73	232,55	36,09M	0,86%
2024-10-29	233,67	233,10	234,32	232,32	35,42M	0,12%
2024-10-30	230,10	232,61	233,47	229,55	47,07M	-1,53%
2024-10-31	225,91	229,34	229,83	225,37	64,37M	-1,82%
2024-11-01	222,91	220,96	225,35	220,27	65,28M	-1,33%

3.3 Exploratory Data Analysis (EDA)

Proses *Exploratory Data Analysis* diperlukan untuk memahami karakteristik dasar dari data, seperti melihat missing value data dan melihat statistik deskriptif data, sebelum melakukan analisis lebih lanjut atau membuat model statistik. Pada prediksi kali ini hal yang dilakukan dalam proses EDA yaitu melihat dimensi data, struktur dataset, mengecek dan visualisasi *missing value*, serta analisis statistik deskriptif.

3.4 Preprocessing

Pada tahap ini yang dilakukan yaitu mengubah tipe kolom dataset. Dalam dataset type setiap variabel memiliki tipe data character dan menghapus karakter yang mengganggu perhitungan saat dilakukan modelling dan mengganti karakter “,” menjadi titik.

3.5 Uji Stationeritas (ADF Test)

Pada tahap pemeriksaan stasioneritas, menggunakan uji seperti *Augmented Dickey-Fuller* (ADF) untuk menilai apakah data memiliki sifat stasioner. Jika data tidak memenuhi kriteria stasioner, dilakukan differencing (pengambilan selisih antar titik data) dan transformasi log untuk menjadikan data menjadi stasioner. Uji ADF memiliki persamaan :

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-i} + \gamma_i \sum_{i=1} \Delta Y_{t-1} + \varepsilon_t$$

Dimana :

Y_t = nilai seri waktu pada waktu t

ΔY_t = perubahan (*differencing*) dari Y_t

β_1 = konstanta (*intercept*)

β_2 = koefisien regresi untuk trend

δ = koefisien regresi untuk lag Y

γ = koefisien regresi untuk differencing lag Y

ε = error

m = lag

t = waktu

Dengan hipotesis :

- $H_0 (\delta \neq 0)$ = tidak stasioner
- $H_1 (\delta = 0)$ = stasioner Dengan keputusan :
- Jika p-value $< \alpha$, maka H_0 ditolak, dan dapat disimpulkan bahwa data stasioner.
- Jika p-value $> \alpha$, maka H_0 diterima, dan dapat disimpulkan bahwa data tidak stasioner.

3.6 Differencing atau Transformasi Log

Proses differencing atau transformasi log dilakukan untuk mengubah data yang tidak stasioner menjadi stasioner. Dalam differencing tingkat pertama, nilai data pada suatu waktu dikurangi dengan nilai data pada waktu sebelumnya. Jika data tetap tidak stasioner, differencing tingkat kedua dilakukan dengan cara yang sama, yaitu mengurangi nilai data yang telah di-differencing pada suatu waktu dengan nilai data yang telah di-differencing pada waktu sebelumnya.

Selain differencing, transformasi logaritmik dapat digunakan untuk mengurangi variabilitas data dan membuatnya lebih stasioner. Transformasi ini mengubah setiap nilai data menjadi logaritmanya. Jika data yang telah ditransformasi logaritmik masih tidak stasioner, differencing kemudian diterapkan pada data yang telah ditransformasi log dengan menghitung perbedaan antara logaritma nilai data pada dua waktu yang berurutan.

3.7 Pemodelan ARIMA

Pemodelan ARIMA (*AutoRegressive Integrated Moving Average*) melibatkan beberapa langkah penting untuk memastikan model yang dipilih dapat memprediksi dengan akurat. Langkah pertama adalah menganalisis *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) dari data yang telah dibuat stasioner. Pada ACF Plot, sumbu vertikal (ACF) menunjukkan kekuatan korelasi antara nilai saat ini dan nilai pada lag (jeda waktu) tertentu. Sumbu horizontal (LAG) menunjukkan jumlah jeda waktu atau interval antar pengamatan. Garis biru putus-putus pada plot ini adalah batas signifikansi. Jika batang ACF melampaui garis ini, itu menunjukkan korelasi yang signifikan. Pada PACF Plot, sumbu vertikal (PACF) menunjukkan kekuatan korelasi antara nilai saat ini dan nilai pada lag tertentu setelah mengendalikan pengaruh interval yang lebih pendek. Sumbu horizontal (LAG) menunjukkan jumlah jeda waktu atau interval antar pengamatan. Garis biru putus-putus pada plot ini juga merupakan batas signifikansi yang sama seperti pada plot ACF.

Setelah menganalisis ACF dan PACF, *Extended Autocorrelation Function* (EACF) dapat digunakan untuk menentukan kandidat model ARIMA yang paling sesuai dengan data. Langkah berikutnya adalah memilih model terbaik dari kandidat

yang dihasilkan dengan menggunakan kriteria informasi, yaitu *Akaike Information Criterion* (AIC) yang mengukur kualitas model berdasarkan keseimbangan antara kecocokan dan kompleksitas model.

Setelah model terbaik dipilih, uji diagnostik dilakukan untuk memeriksa apakah residual (sisa) dari model tersebut memenuhi asumsi *white noise*, yang berarti tidak ada pola yang tersisa dan residuals berdistribusi secara acak. Selanjutnya, goodness of fit dari model dievaluasi untuk memastikan bahwa model tersebut cocok dengan data historis. Ini melibatkan pengujian apakah model tersebut dapat mereproduksi pola yang terlihat dalam data historis. Jika model menunjukkan goodness of fit yang baik, langkah terakhir adalah menggunakan model ARIMA untuk melakukan prediksi harga saham AAPL. Prediksi ini memberikan perkiraan harga saham di masa depan, yang dapat digunakan untuk perencanaan dan pengelolaan transportasi yang lebih efektif.

3.8 Analisis Hasil

Setelah model ARIMA diterapkan, hasil prediksi dibandingkan dengan data aktual untuk menilai akurasi. Proses ini melibatkan perbandingan visual antara nilai prediksi dan nilai aktual dalam bentuk grafik, serta penggunaan metrik statistik seperti *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE) untuk mengukur tingkat kesalahan prediksi.

Evaluasi performa model ARIMA melibatkan analisis kesalahan prediksi dan uji diagnostik lanjutan untuk memastikan residuals tidak menunjukkan pola sistematis. Jika residual berdistribusi secara acak dan model menunjukkan performa yang baik, model tersebut dianggap valid untuk prediksi jangka panjang. Jika model menunjukkan kelemahan, perlu dilakukan perbaikan seperti penyesuaian parameter.

3.9 Kesimpulan

Penelitian ini akan menyimpulkan bahwa model ARIMA efektif dalam memprediksi harga saham AAPL. Melalui analisis data, model ARIMA dapat mengidentifikasi pola pertumbuhan dan fluktuasi musiman dengan baik.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Data Harga Saham Apple Inc. (AAPL)

Data yang digunakan untuk prediksi ini memiliki dimensi data yaitu 7 kolom variabel yang terdiri dari Date, Terakhir, Pembukaan, Tertinggi, Terendah, Volume, Perubahan. Berikut tampilan data yang sudah dilakukan *Preprocessing* dan akan digunakan untuk pemodelan ARIMA.

Tabel 4.1. Data Saham AAPL

Tanggal	Terakhir	Pembukaan	Tertinggi	Terendah	Volume	Perubahan
2022-01-03	182,01	177,83	182,88	177,71	104,70M	2,50%
2022-01-04	179,70	182,63	182,94	179,12	99,31M	-1,27%
2022-01-05	174,92	179,61	180,17	174,64	94,54M	-2,66%
2022-01-06	172,00	172,70	175,30	171,64	96,90M	-1,67%
2022-01-07	172,17	172,89	174,14	171,03	86,71M	0,10%
2022-01-10	172,19	169,08	172,50	168,17	106,77M	0,01%
2022-01-11	175,08	172,32	175,18	170,82	76,14M	1,68%
2022-01-12	175,53	176,12	177,18	174,82	74,81M	0,26%
2022-01-13	172,19	175,78	176,62	171,79	84,51M	-1,90%
.
.
.
2024-10-21	236,48	234,45	236,85	234,45	36,25M	0,63%

Tanggal	Terakhir	Pembukaan	Tertinggi	Terendah	Volume	Perubahan
2022-01-03	182,01	177,83	182,88	177,71	104,70M	2,50%
2022-01-04	179,70	182,63	182,94	179,12	99,31M	-1,27%
2022-01-05	174,92	179,61	180,17	174,64	94,54M	-2,66%
2022-01-06	172,00	172,70	175,30	171,64	96,90M	-1,67%
2022-01-07	172,17	172,89	174,14	171,03	86,71M	0,10%
2022-01-10	172,19	169,08	172,50	168,17	106,77M	0,01%
2024-10-22	235,86	233,88	236,22	232,60	38,85M	-0,26%
2024-10-23	230,76	234,08	235,14	227,76	52,29M	-2,16%
2024-10-24	230,57	229,98	230,82	228,41	31,11M	-0,08%
2024-10-25	231,41	229,74	233,22	229,57	38,80M	0,36%
2024-10-28	233,40	233,32	234,73	232,55	36,09M	0,86%
2024-10-29	233,67	233,10	234,32	232,32	35,42M	0,12%
2024-10-30	230,10	232,61	233,47	229,55	47,07M	-1,53%
2024-10-31	225,91	229,34	229,83	225,37	64,37M	-1,82%
2024-11-01	222,91	220,96	225,35	220,27	65,28M	-1,33%

4.2 Exploratory Data Analysis (EDA)

Sebelum dilakukan prediksi dengan pemodelan ARIMA pada data, langkah yang harus dilakukan terlebih dahulu yaitu memahami karakteristik dasar dari data. Berikut beberapa proses yang dilakukan di dalam *Exploratory Data Analysis* :

a. Melihat Dimensi Data

[1] 713 7

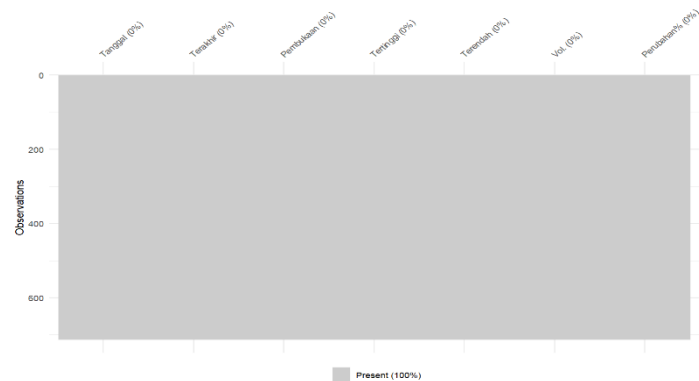
Data yang digunakan untuk prediksi harga saham yaitu terdiri dari 7 kolom variabel dan masing- masing kolom memiliki 713 baris data.

b. Melihat Struktur Data

Data pada variabel “Tanggal” mencakup 713 tanggal, dimulai dari "03-01-2022" hingga "01-11-2024".

c. Melihat *Missing Value*

Gambar 4.1. *Plot test Missing Value*



Data yang digunakan untuk prediksi tidak memiliki *missing value* pada masing-masing variable, khususnya variabel Terakhir sehingga tidak diperlukan tahapan *preprocessing* di bagian ini.

d. Analisis Deskriptif

Gambar 4.2. Gambar hasil analisis deskriptif

```
> summary(Data_AAPL)
```

Tanggal		Terakhir	Pembukaan
Min.	:2022-01-03 00:00:00.0	Min. :125.0	Min. :126.0
1st Qu.	:2022-09-19 00:00:00.0	1st Qu.:154.7	1st Qu.:154.6
Median	:2023-06-05 00:00:00.0	Median :172.7	Median :172.3
Mean	:2023-06-04 01:51:04.7	Mean :174.8	Mean :174.7
3rd Qu.	:2024-02-20 00:00:00.0	3rd Qu.:189.4	3rd Qu.:189.3
Max.	:2024-11-01 00:00:00.0	Max. :236.5	Max. :236.5

Tertinggi	Terendah	Vol.	Perubahan%
Min. :127.8	Min. :124.2	Min. : 24050000	Min. :-5.87000
1st Qu.:156.7	1st Qu.:152.9	1st Qu.: 50130000	1st Qu.: -0.88000
Median :173.9	Median :170.9	Median : 64240000	Median : 0.08000
Mean :176.5	Mean :172.9	Mean : 69384222	Mean : 0.04693
3rd Qu.:190.3	3rd Qu.:188.0	3rd Qu.: 81510000	3rd Qu.: 0.98000
Max. :237.5	Max. :234.4	Max. :318680000	Max. : 8.90000

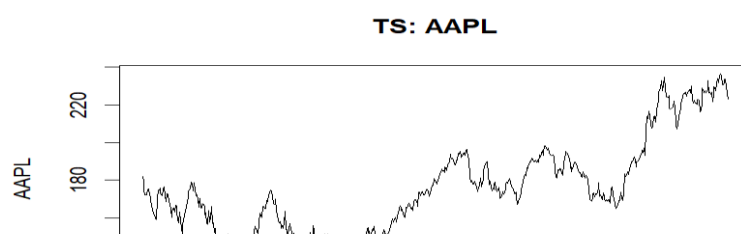
Berdasarkan statistika deskriptif untuk variabel Terakhir (closing price) saham AAPL, data menunjukkan rentang harga yang cukup luas dari minimal \$125.0 hingga maksimal \$236.5. Nilai rata-rata (mean) closing price berada di \$174.8 dengan nilai tengah (median) \$172.7, dimana kedekatan nilai ini mengindikasikan distribusi yang cenderung simetris. Kuartil pertama (Q1) berada di level \$154.7 dan kuartil ketiga (Q3) di \$189.4, memberikan rentang interkuartil (IQR) sebesar \$34.7. Posisi mean yang sedikit lebih tinggi dari median menunjukkan adanya sedikit kemiringan positif (positive skewness) dalam distribusi harga penutupan. Data ini mencakup periode dari 3 Januari 2022 hingga 1 November 2024, yang memberikan gambaran komprehensif tentang pergerakan harga saham AAPL selama hampir tiga tahun terakhir.

4.3 Kestasioneran Data

a. Plot Time Series

Plot *time series* digunakan untuk mevisualisasikan data yang telah dikumpulkan pada interval waktu yang berurutan. Plot *time series* akan digunakan untuk mengidentifikasi tren yang ada pada dataset.

Gambar 4. 3 Plot Time series stasioner data



Plot di atas menunjukkan data time series (deret waktu) dari harga saham AAPL. Terlihat jelas adanya tren kenaikan secara umum dari waktu ke waktu. Ini mengindikasikan bahwa data tersebut tidak stasioner dalam rata-rata. Artinya, rata-rata dari data ini tidak konstan sepanjang waktu, melainkan cenderung meningkat.

Selain itu, amplitudo fluktuasi harga saham juga cenderung meningkat seiring waktu. Hal ini menunjukkan bahwa data juga tidak stasioner dalam varians. Artinya, varians atau dispersi data tidak konstan, melainkan cenderung meningkat.

b. Uji Stasioneritas

Uji stasioneritas digunakan untuk memastikan bahwa data yang digunakan memiliki statistik dasar (mean, varians, autokorelasi) yang tidak berubah seiring berjalannya waktu. Dalam hal ini, uji stasioneritas yang dilakukan adalah uji Augmented Dickey-Fuller Test :

Tabel 4. 2. Tabel Uji Stasioner

<u>Parameter</u>	<u>Nilai</u>
Dickey-Fuller	-2.6289
p-value	0.3121
λ	1.303541

Uji terhadap dengan

stasioneritas ragam dilakukan transformasi box-

cox yaitu dengan mengestimasi parameter lambda dimana lambda yang mendekati angka 1 artinya data sudah stasioner terhadap ragam. Dikarenakan hasil nilai λ menunjukkan angka 1.303541, dimana angka tersebut tampaknya masih jauh dari angka satu maka data dikatakan tidak stationer secara ragam/variens.

Uji stasioneritas terhadap rata-rata dilakukan dengan ADF test untuk melihat apakah data memiliki unit root yang menunjukkan bahwa data tidak stasioner. Pada uji ini, diperlukan hipotesis yaitu :

H0 : Data tidak stasioner

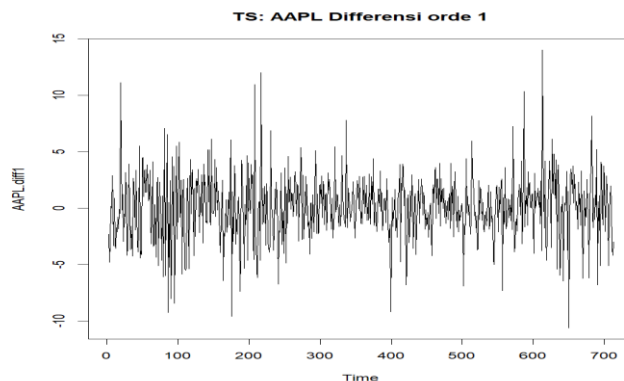
H1 : Data stasioner

H0 ditolak jika $p\text{-value} < \alpha$ (5%). Sehingga dari hasil diatas diperoleh p-values lebih besar dari nilai α ($0.3121 > 0,05$) sehingga H0 diterima yang berarti data tidak stasioner secara rata-rata.

Kesimpulan: Kedua uji di atas menunjukkan bahwa data tidak stasioner terhadap rata-rata dan ragam sehingga data perlu dilakukan *differencing* atau *transformasi*.

c. **Differencing sekali pada data asli**

Gambar 4.4. Plot Differencing satu kali



Berdasarkan grafik differensi orde 1 untuk data saham AAPL, dapat dilihat bahwa data telah menunjukkan karakteristik yang cukup stasioner baik dalam varians maupun rata-rata. Dari segi stasioneritas rata-rata, fluktuasi data bergerak konsisten di sekitar nilai nol ($\text{mean} = 0$) tanpa menunjukkan tren naik atau turun yang signifikan, ini mengindikasikan bahwa data telah stasioner dalam rata-rata setelah dilakukan differencing. Untuk stasioneritas varians, dapat diamati bahwa sebaran atau amplitudo fluktuasi relatif konstan sepanjang waktu dengan mayoritas nilai bergerak antara -5 hingga +5, meskipun terdapat beberapa spike atau lonjakan ekstrem yang mencapai nilai -10 hingga +15. Keberadaan spike ini masih dapat diterima dalam konteks data finansial karena

merepresentasikan momen-momen volatilitas tinggi yang wajar terjadi di pasar saham. Secara keseluruhan, hasil differencing orde pertama ini menunjukkan bahwa transformasi yang dilakukan telah berhasil menghasilkan data yang memenuhi asumsi stasioneritas, baik dalam rata-rata maupun varians.

Kemudian, akan di uji lagi stasioneritas setelah differencing pertama dan hasilnya menunjukkan :

Tabel 4. 2. Tabel nilai differencing satu kali

<u>Parameter</u>	<u>Nilai</u>
Dickey-Fuller	-8.6118
p-value	< 0.01
λ	1.303541

- Uji Stasioneritas terhadap Ragam

Dikarenakan hasil nilai λ menunjukkan angka Nilai lambda terbaik: 0.7647302, dimana angka tersebut cukup mendekati angka 1, maka data dapat dikatakan cukup stationer secara ragam/variens.

- Uji Stasioneritas terhadap Rataan

H0 : Data tidak stationer

H1 : Data stationer

H0 ditolak jika p-value < α (5%). Sehingga dari hasil diatas diperoleh p-values kurang dari nilai α (0,01 < 0,05) sehingga H0 ditolak yang berarti data stationer secara rata-rata.

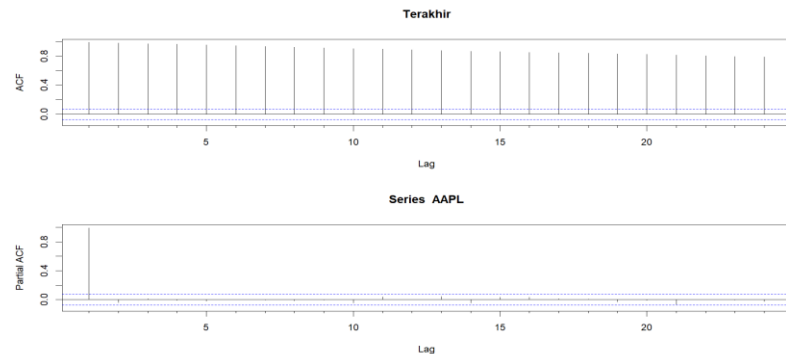
d. Plot ACF dan PACF

Plot ACF (*Autocorrelation Function*) dan PACF (*Partial Autocorrelation Function*) digunakan untuk memahami pola data yang berulang atau keterkaitan

antar waktu.

- **Plot ACF dan PACF sebelum *differencing***

Gambar 4. 5. Ploat ACF dan PACF sebelum di differencing

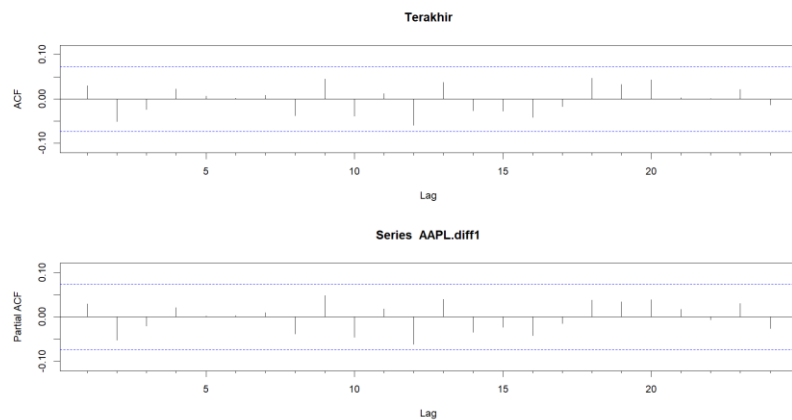


Pada plot ACF terlihat bahwa bar pertama di lag 0 selalu 1, karena setiap titik data 100% berkorelasi dengan dirinya sendiri dan bar-bar selanjutnya menunjukkan korelasi antara titik data dengan lag tertentu. Pada plot ACF ini, dapat dilihat bahwa korelasi tetap signifikan (lebih tinggi dari garis biru putus-putus) hingga sekitar lag 20. Ini menunjukkan adanya pola autokorelasi yang kuat dalam data, yang berarti nilai-nilai masa lalu masih mempengaruhi nilai-nilai masa kini hingga beberapa periode ke belakang. Plot ACF menunjukkan adanya autokorelasi yang kuat dan berkepanjangan hingga sekitar lag 20, menunjukkan bahwa data ini mungkin memiliki pola musiman atau tren yang panjang.

Pada plot PACF terlihat bar pertama di lag 0 juga selalu 1 dan bar kedua di lag 1 menunjukkan korelasi parsial yang signifikan. Setelah lag 1, korelasi parsial tampaknya berkurang dan sebagian besar berada dalam batas signifikan (di antara garis biru putus-putus). Plot PACF menunjukkan bahwa setelah mengendalikan pengaruh lag pertama, korelasi signifikan berkurang drastis, yang dapat mengindikasikan bahwa model ARIMA dengan p (order of AR) kecil mungkin cukup untuk menangkap pola ini.

- **Plot ACF dan PACF sesudah *differencing***

Gambar 4. 6. Gambar plot ACF dan PACF setekah di diferencing



Plot ACF setelah *differencing* sebagian besar batangnya berada dalam batas signifikan (di antara garis biru putus-putus) dan hanya ada beberapa lag yang menunjukkan korelasi signifikan, yang berarti pola autokorelasi dalam data asli telah berkurang.

Plot PACF setelah *differencing* menunjukkan korelasi parsial juga berkurang signifikan dan sebagian besar batang PACF berada dalam batas signifikan dan hanya ada sedikit lag yang menunjukkan korelasi parsial signifikan, menunjukkan bahwa sebagian besar pola korelasi jangka pendek telah dihilangkan.

Pada plot ACF dan PACF setelah differencing terlihat pola autokorelasi yang kuat dalam data asli telah berkurang dan sebagian besar korelasi dalam plot ACF dan PACF berada dalam batas signifikansi, menunjukkan bahwa data yang di-differencing ini lebih mendekati sifat stasioner (tidak bergantung pada waktu).

e. EACF

Extended Autocorrelation Function digunakan untuk menentukan nilai terbaik dari parameter p (order AR) dan q (order MA) dalam model ARIMA. EACF menggabungkan informasi dari autocorrelation function (ACF) dan partial autocorrelation function (PACF). Hasil dari EACF akan digunakan untuk menentukan kandidat model ARIMA.

Gambar 4. 7. Gambar hasil EACF 1

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

Gambar 4. 8. Gambar hasil EACF 2

symbol		0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	"x"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"x"	"o"	"o"	"x"	"x"	"o"
1	"x"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"
2	"x"	"o"	"x"	"o"	"o"	"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"	"o"
4	"x"	"x"	"o"	"o"	"x"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"
5	"x"	"o"	"x"	"o"	"x"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"
6	"x"	"x"	"x"	"x"	"x"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"
7	"x"	"x"	"x"	"x"	"x"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"	"o"

- Kandidat Model**

Diperoleh 33 kandidat model yang akan digunakan untuk mencari model yang terbaik, dengan 26 kandidat model tanpa tranformasi dan 7 kandidat data dengan tranformasi logaritma. Kandidat model yang diperoleh diantaranya yaitu:

Tabel 4. 3. Tabel Hasil perolehan model ARIMA

ARIMA(1,1,1)	ARIMA(0,1,2)	ARIMA(0,1,3)	ARIMA(0,1,4)	ARIMA(0,1,5)
ARIMA(1,1,2)	ARIMA(1,1,3)	ARIMA(1,1,4)	ARIMA(1,1,5)	ARIMA(2,1,2)
ARIMA(2,1,3)	ARIMA(2,1,4)	ARIMA(2,1,5)	ARIMA(3,1,3)	ARIMA(3,1,4)
ARIMA(3,1,5)	ARIMA(1,2,1)	ARIMA(2,2,1)	ARIMA(3,2,1)	ARIMA(4,2,1)
ARIMA(5,2,1)	ARIMA(6,2,1)	ARIMA(7,2,1)	ARIMA(8,2,1)	ARIMA(9,2,1)
ARIMA(10,2,1)				

Tabel 4. 4. Tabel Perolehan model terbaik

MODEL	AIC
ARIMA(1,1,1)	3541.25

ARIMA(0,1,2)	3538.95
ARIMA(0,1,3)	3540.54
ARIMA(0,1,4)	3542.13
ARIMA(0,1,5)	3544.12
ARIMA(1,1,2)	3540.76
ARIMA(1,1,3)	3542.92
ARIMA(1,1,4)	3542.92
ARIMA(1,1,5)	3546.1
ARIMA(2,1,2)	3542.24
ARIMA(2,1,3)	3544.21
ARIMA(2,1,4)	3546.13
ARIMA(2,1,5)	3548.15
ARIMA(3,1,3)	3544.2
ARIMA(3,1,4)	3548.13
ARIMA(3,1,5)	3550.09
ARIMA(1,2,1)	3542.96
ARIMA(2,2,1)	3543.11
ARIMA(3,2,1)	3544.82
ARIMA(4,2,1)	3546.49
ARIMA(5,2,1)	3548.48
ARIMA(6,2,1)	3550.46
ARIMA(7,2,1)	3552.38
ARIMA(8,2,1)	3553.35
ARIMA(9,2,1)	3553.67
ARIMA(10,2,1)	3554.28

Diperoleh nilai AIC terkecil ada pada model ARIMA(2,1,4) sehingga model tersebut akan dicoba untuk uji diagnostik dan menentukan *goodness of fit* nya.

Tabel 4. 5. Transformasi Logaritma

MODEL	AIC
-------	-----

ARIMA(1,1,1)	- 3752.99
ARIMA(1,1,2)	-3747.63
ARIMA(2,1,2)	-3745.99
ARIMA(2,1,1)	-3747.7
ARIMA(1,2,2)	-3738.37
ARIMA(1,2,1)	-3734.52
ARIMA(2,2,2)	-3732.08

f. Uji Diagnostik dan Goodness of Fit

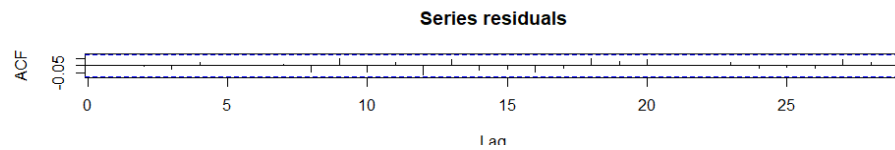
Kami telah melakukan uji diagnostik pada 33 kandidat model yang diperoleh dengan menggunakan nilai AIC terkecil pertama, nilai AIC terkecil kedua, dan nilai AIC terkecil dari data transformasi didapatkan dari 30 kandidat model diantaranya menunjukkan nilai L-Jung box yang p-values nya masih dibawah dan juga masih ada yang bersinggungan dengan garis putus-putus, sehingga model tersebut belum bisa dikatakan baik dan dapat digunakan. Namun ada 3 model yang nilai p-values L-jung Box nya sudah diatas 0,05. Berikut Uji Diagnostik dari beberapa kandidat model

- Data Asli tanpa transformasi

F.1. ARIMA(0,1,2)

❖ ACF Residual

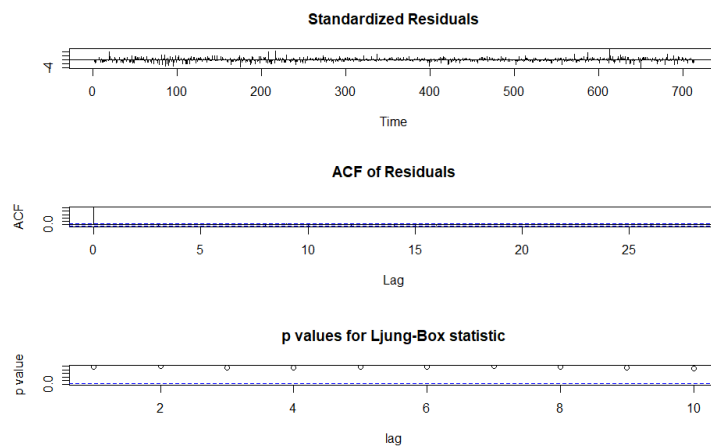
Gambar 4.9. Plot ACF Residual



Pada plot di atas, model ARIMA(0,1,2) telah berhasil menghilangkan semua autokorelasi pada residual, bisa dilihat pada plot bahwa semua garis berada di antara garis biru (batas signifikan).

❖ L-Jung Box Statistic

Gambar 4. 10. Plot L-Jung Box Statistic

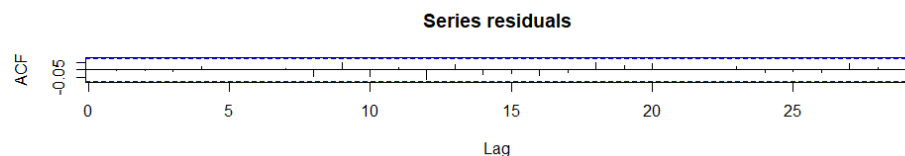


Terlihat bahwa nilai p-values dari L-jung box model ARIMA(0,1,2) sudah berada di atas 0,05 sehingga dapat dikatakan model yang digunakan sudah baik.

F.2. ARIMA(1,1,2)

❖ ACF Residual

Gambar 4. 11. ACF Residual ARIMA (1,1,2)

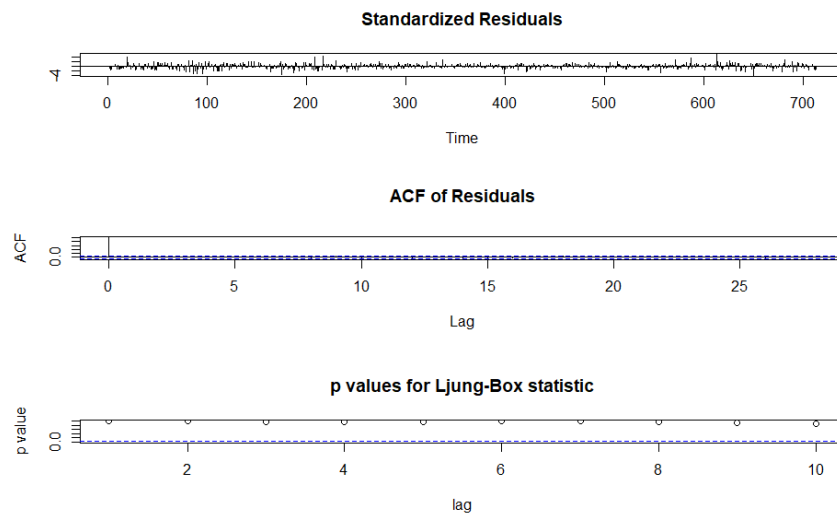


Pada plot di atas, model ARIMA(1,1,2) telah berhasil

menghilangkan semua autokorelasi pada residual, bisa dilihat semua garis berada di antara garis biru (batas signifikan).

❖ L-Jung Box Statistic

Gambar 4. 12. Gambar L-Jung Box Statistic ARIMA (1,1,2)

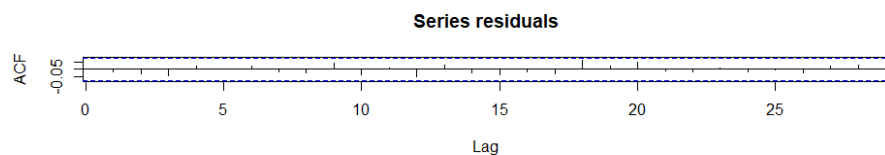


Terlihat bahwa nilai p-values dari L-jung box model ARIMA(1,1,2) sudah berada di atas 0,05 sehingga dapat dikatakan model yang digunakan sudah baik.

F.3. ARIMA (1,1,1)

❖ ACF Residual

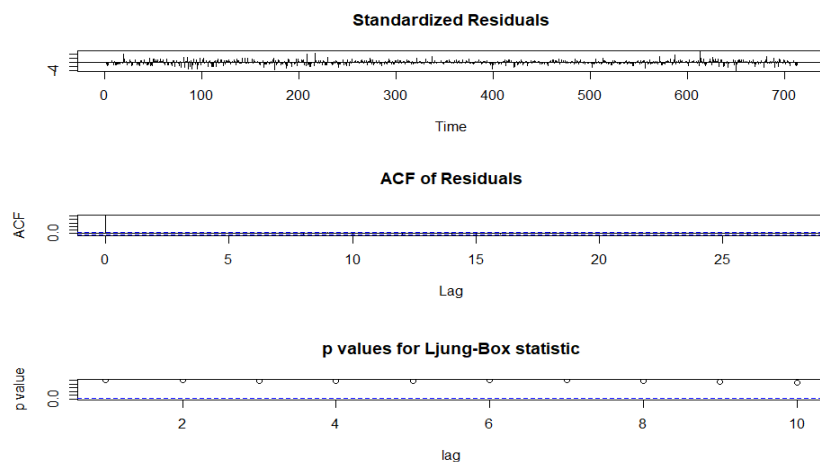
Gambar 4.13. Gambar ACF Residual ARIMA (1,1,1)



Pada plot di atas, model ARIMA(1,1,1) telah berhasil menghilangkan semua autokorelasi pada residual, bisa dilihat semua garis berada di antara garis biru (batas signifikan).

❖ L-Jung Box Statistic

Gambar 4.14. Plot L-Jung Box ARIMA (1,1,1)



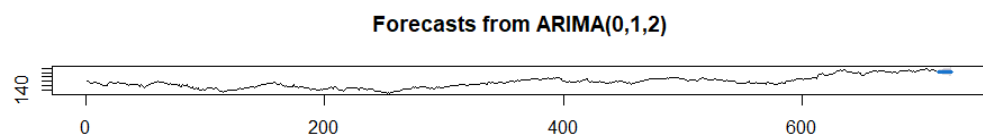
Terlihat bahwa nilai p-values dari L-jung box model ARIMA(1,1,1) sudah berada di atas 0,05 sehingga dapat dikatakan model yang digunakan sudah baik.

4.4 Prediksi Harga Saham AAPL

Data Asli tanpa transformasi

4.4.1. ARIMA (0,1,2)

Gambar 4.15. Plot Hasil Forecast ARIMA (0,1,2)



Gambar di atas merupakan hasil dari analisis *time series* yang menunjukkan prediksi nilai-nilai masa depan. Kami melakukan prediksi untuk 12 hari kedepan berdasarkan data historis dari 3 Januari 2022 sampai 1 November 2024 yang menghasilkan forecasting yang divisualisasikan dengan garis biru. Berikut adalah nilai prediksi 12 hari kedepan :

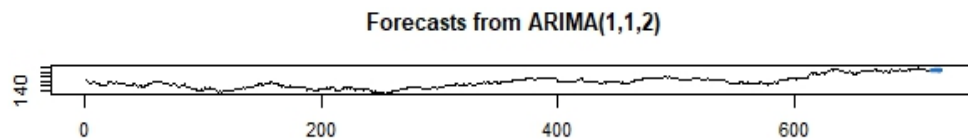
Gambar 4.16. Gambar hasil forecastin ARIMA (0,1,2)

Point	Forecast
714	223.0138
715	223.1584
716	223.1584
717	223.1584
718	223.1584
719	223.1584
720	223.1584
721	223.1584
722	223.1584
723	223.1584
724	223.1584
725	223.1584

Dari forecast di atas kita hanya memiliki perbedaan nilai di forecast data ke 714 dan 715, selain data tersebut semuanya sama dengan data ke 715, yang berarti model ini belum sepenuhnya bagus dalam hal meramalkan

4.4.2. ARIMA (1,1,2)

Gambar 4.17. Plot hasil forecasting ARIMA (1,1,2)



Gambar di atas merupakan hasil dari analisis *time series* yang menunjukkan prediksi nilai-nilai masa depan. Kami melakukan prediksi untuk 12 hari kedepan berdasarkan data historis dari 3 Januari 2022 sampai 1 November 2024 yang menghasilkan forecasting yang divisualisasikan dengan garis biru. Berikut adalah nilai prediksi 12 hari kedepan :

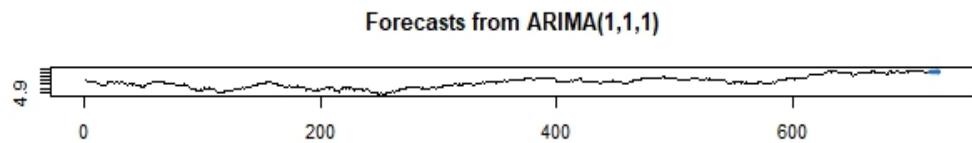
Gambar 4.18. Gambar hasil nilai forecasting ARIMA (1,1,2)

Point	Forecast
714	223.0558
715	223.2645
716	223.3176
717	223.3312
718	223.3346
719	223.3355
720	223.3357
721	223.3358
722	223.3358
723	223.3358
724	223.3358
725	223.3358

Dari forecast di atas kita hanya memiliki perbedaan nilai di forecast data ke 714 sampai 721, selain data tersebut semuanya sama dengan data ke 721, yang berarti model ini belum sepenuhnya bagus dalam hal meramalkan atau hanya bisa meramalkan sampai data ke 721.

4.4.3. ARIMA (1,1,1)

Gambar 4.19. Plot hasil forecasting ARIMA (1,1,1)



Gambar di atas merupakan hasil dari analisis *time series* yang menunjukkan prediksi nilai-nilai masa depan. Kami melakukan prediksi untuk 12 hari kedepan berdasarkan data historis dari 3 Januari 2022 sampai 1 November 2024 yang menghasilkan forecasting yang divisualisasikan dengan garis biru. Berikut adalah nilai prediksi 12 hari kedepan:

Gambar 4.20. Gambar hasil nilai forecasting (1,1,1)

Point	Forecast Transformasi	Forecast Actual
714	5.40776	223.1312136
715	5.406864	222.9313775
716	5.407674	223.1120251
717	5.406943	222.9489898
718	5.407603	223.0961847
719	5.407007	222.963259
720	5.407545	223.0832455
721	5.407059	222.9748534
722	5.407498	223.0727608
723	5.407101	222.9842185
724	5.407459	223.0640612
725	5.407136	222.9920231

Dari forecast di atas kita memiliki perbedaan nilai di semua forecast data, walaupun hanya antar data hanya selisih sedikit.

Tabel 4.6. Perbandingan akurasi ketiga model

Model	ME	RMSE	MAE	MAPE
Model 2	0.07245682	2.895894	2.158104	1.278846

Model 6	0.07363287	2.89566	2.159143	1.279527
Model 27	168.703	170.515	168.703	96.99077

BAB V

PENUTUP

5.1 Kesimpulan

Dari ketiga model di atas, kami melkaukan perbadign ME, RMSE, MAE, MPE, dan MAPA yang terkecil adalah model 2, ARIMA(0,1,2), tapi model tersebut hanya bisa memprediksi 2 hari setelahnya, yang mana hari berikutnya sama nilai forecastnya, dan begitu juga model 6, ARIMA(1,1,2). Berbeda dengan model 27 kita mendapatkan nilai forecast dari 12 hari berikutnya yang berbeda walaupun selisih kecil antara data forecastnya. Jadi kesimpulanya model 27. Jadi kesimpulanya kami memilih model 27

5.2 Saran

Berdasarkan hasil analisis, model 27 dipilih karena mampu memberikan prediksi hingga 12 hari ke depan dengan variasi nilai forecast meskipun selisihnya kecil. Disarankan untuk terus memvalidasi kinerja model ini dengan data terbaru, mempertimbangkan pola musiman jika relevan, dan mengevaluasi residual untuk memastikan akurasi. Selain itu, eksplorasi model alternatif seperti SARIMA atau machine learning dapat dilakukan untuk meningkatkan performa prediksi di masa mendatang.