



Laporan Analisis Data *Time Series*

Peramalan Indeks Harga Konsumen Gabungan Sektor Transportasi Menggunakan Model SARIMAX (Regresi *Time Series*) untuk Memodelkan Intervensi Kebijakan dan Struktural

Nurrulia Rayhan Naifah¹, Adinda Putri Novianti², Selvi Maulidyah³,
Veronica Febriani Putri⁴, Dandy Harsa Fathoni⁵, Naufal Mahdy Nashrullah⁶
Analisis Deret Waktu Kelas D

¹Departemen Statistika, Fakultas Sains dan Analitika Data (FSAD),
Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Jawa Timur, Indonesia

Abstrak

Penelitian ini bertujuan untuk memproyeksikan pergerakan Indeks Harga Konsumen Gabungan sektor Transportasi periode Januari 2020 hingga September 2025 menggunakan model SARIMAX dengan variabel tambahan berupa *outlier* dan intervensi. Analisis deret waktu IHKG Transportasi menunjukkan adanya volatilitas tinggi dan anomali struktural, terutama lonjakan harga pasca penyesuaian harga BBM bersubsidi (September 2022) dan penurunan diskontinu yang masif di awal 2024 yang diakibatkan oleh perubahan tahun dasar (*rebasing*) IHKG oleh Badan Pusat Statistik. Model dasar ARIMA (0, 1, 0) terbukti tidak memadai karena gagal mereplikasi diskontinuitas ini. Oleh karena itu, dikembangkan model SARIMAX yang memasukkan variabel eksogen *level shift* dan *additive outlier*. Meskipun model SARIMAX dengan penambahan *outlier* memenangkan kriteria *parsimony* AIC sebesar 91,99, model terintegrasi SARIMAX dengan *outlier* dan intervensi menunjukkan kinerja peramalan terbaik pada data pengujian dengan RMSE terendah sebesar 0,3296231. Hasil ini menegaskan bahwa penanganan simultan terhadap peristiwa struktural permanen dan pencilan sementara secara signifikan meningkatkan daya prediksi model deret waktu ekonomi.

Kata Kunci: ARIMA, SARIMAX, IHKG Transportasi, Regresi *Time Series*

Tabel Spesifikasi

Subject area	Statistika Terapan
More specific subject area	Pemodelan Deret Waktu pada Data Ekonomi
Method	1. Metode ARIMA klasik 2. Metode SARIMAX 3. Metode regresi <i>time series</i>
Name and reference of original method	1. Hyndman, R.J., & Athanasopoulos, G. 2021. <i>Forecasting: principles and practice</i> , 3rd edition. Melbourne: OTexts 2. Cryer, J. D., & Chan, K.-S. 2008. <i>Time Series Analysis: With Applications in R</i> (2nd ed.). New York: Springer 3. Minarnik. 2009. <i>Analisis Time Series terhadap Indeks Harga Konsumen (IHK) di Indonesia dengan Menggunakan Metode Intervensi untuk Mengatasi Outlier</i> . Institut Teknologi Sepuluh Nopember
Resource availability	Data berupa Indeks Harga Konsumen Gabungan (IHKG) di sektor transportasi dipublikasikan oleh Bank Indonesia dalam bentuk dokumen <i>excel</i> pada periode Januari 2020 hingga September 2025.



Bab I Pendahuluan

1.1 Latar Belakang

Indeks Harga Konsumen (IHK) merupakan indikator utama yang digunakan untuk mengukur perubahan harga dari sekelompok barang dan jasa yang dikonsumsi masyarakat. Salah satu komponennya adalah IHK kategori transportasi yang memiliki peranan penting dalam menggambarkan biaya mobilitas penduduk serta dinamika kegiatan ekonomi. Pergerakan harga pada kelompok transportasi cenderung dipengaruhi oleh faktor musiman, pola permintaan, serta berbagai kebijakan dari pemerintah yang menyebabkan perubahan harga secara tiba-tiba.

Dalam beberapa tahun terakhir pola IHK pada kategori transportasi menunjukkan adanya fluktuasi yang tidak sepenuhnya bersifat musiman. Terdapat periode tertentu yang memperlihatkan perubahan signifikan dan berbeda dari pola historis sebelumnya yang mengindikasikan adanya *structural break* dalam pergerakan data. Perubahan seperti ini tidak dapat dijelaskan secara memadai oleh model deret waktu konvensional seperti ARIMA yang hanya mengandalkan data masa lalu tanpa mempertimbangkan kejadian khusus di luar pola historisnya.

Oleh karena itu, pendekatan Model Intervensi menjadi relevan untuk digunakan dalam analisis ini. Model Intervensi memungkinkan analisis untuk menangkap dampak suatu kejadian tertentu yang menyebabkan perubahan mendadak maupun perubahan jangka panjang pada data *time series*. Pendekatan ini diharapkan mampu memberikan gambaran yang lebih akurat mengenai pola perubahan indeks sekaligus menghasilkan prediksi yang lebih dapat diandalkan untuk mendukung perencanaan dan evaluasi kebijakan yang terkait dengan dinamika harga pada sektor transportasi.

1.2 Tujuan Penelitian

Adapun tujuan studi ini, antara lain:

1. Menganalisis pola pergerakan Indeks Harga Konsumen (IHK) Transportasi.
2. Mendeteksi dan mengevaluasi adanya kejadian khusus seperti outlier dan intervensi.
3. Membangun model peramalan terbaik.
4. Menghasilkan proyeksi nilai IHK Transportasi untuk periode mendatang.

1.3 Manfaat Penelitian

Adapun manfaat studi ini, antara lain:

1. Memahami pola pergerakan IHK Transportasi.
2. Meningkatkan kemampuan analisis *time series* yang komprehensif.
3. Menyediakan metode yang dapat direplikasi untuk studi akademik berikutnya.
4. Memberikan dasar informasi berbasis data untuk merumuskan kebijakan bagi pemangku kebijakan.



Bab II

Tinjauan Pustaka

2.1 Indeks Harga Konsumen

Indeks Harga Konsumen (IHK) merupakan salah satu indikator statistik yang digunakan untuk mengukur perubahan harga barang dan jasa yang dikonsumsi rumah tangga dalam suatu periode waktu tertentu. IHK menjadi alat utama dalam pengukuran inflasi serta indikator penting bagi pemerintah dan pelaku ekonomi dalam merumuskan kebijakan fiskal dan moneter (Badan Pusat Statistik, 2023).

IHK disusun berdasarkan hasil dari pengolahan Survey Harga Konsumen (SHK). Survei tersebut mencakup berbagai kelompok pengeluaran, seperti bahan makanan, perumahan, sandang, kesehatan, pendidikan, hingga transportasi. Dalam kelompok transportasi, survei ini mencakup komoditas seperti tarif angkutan, bahan bakar kendaraan, dan jasa perbaikan kendaraan. Perubahan harga pada kelompok ini sangat dipengaruhi oleh dinamika permintaan masyarakat, fluktuasi harga energi, kondisi ekonomi, serta kebijakan pemerintah seperti penyesuaian harga BBM atau tarif angkutan umum.

Survei harga dilakukan untuk mendapatkan data harga konsumen yang biasa dibayar dalam kondisi normal yang kemudian digunakan untuk menghitung indeks ini. Pengumpulan data ini mencerminkan perubahan harga dari paket komoditas yang dikonsumsi oleh rumah tangga, sehingga perubahan IHK menggambarkan tingkat inflasi atau deflasi yang terjadi pada berbagai kelompok pengeluaran, termasuk transportasi. Hal ini penting sebagai dasar kebijakan pemerintah dan indikator ekonomi secara umum.

Singkatnya, IHK menggunakan data survei harga dari berbagai kelompok pengeluaran dan dalam kelompok transportasi meliputi tarif angkutan, bahan bakar kendaraan, dan jasa perbaikan, yang dipengaruhi oleh faktor permintaan, harga energi, ekonomi, dan kebijakan pemerintah terkait BBM dan tarif angkutan publik.

2.2 Statistika Deskriptif

Statistika Merupakan metode pengumpulan data, analisis, interpretasi, dan pengambilan keputusan berdasarkan hasil analisis. Statistika dibagi menjadi statistika deskriptif dan statistika inferensia. Statistika deskriptif didefinisikan sebagai metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna (Walpole, 1995). Statistika deskriptif digunakan untuk memberikan informasi seputar data tanpa mengambil suatu keputusan (inferensi). Salah satu ukuran pemusatan data yang digunakan pada penelitian ini adalah rata-rata. Sedangkan ukuran penyebaran data yang digunakan pada penelitian ini adalah standar deviasi.

2.2.1 Rata-rata (*Mean*)

Rata-rata atau *mean* merupakan jumlah keseluruhan data yang diamati, kemudian dibagi dengan banyaknya data. Secara matematis, rata-rata dapat ditulis sebagai berikut.

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$$



Laporan Analisis Data *Time Series*

Dengan Y_i merupakan data pengamatan ke- i dan n merupakan banyaknya data pengamatan.

2.2.2 Standar Deviasi

Standar deviasi merupakan nilai yang menunjukkan penyebaran data yang dianalisis. Semakin kecil nilai standar deviasi, maka semakin kecil pula penyebaran data pengamatan, dalam arti lain, variasi data yang diamati semakin kecil dan semakin seragam, demikian juga sebaliknya. Secara matematis, standar deviasi dapat ditulis sebagai berikut.

$$s = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}}$$

dengan,

\bar{Y} = rata-rata pengamatan

Y_i = data pengamatan ke- i

n = banyak data pengamatan

2.2.3 Plot *Time Series*

Plot *time series* merupakan plot yang menggambarkan nilai variabel random berdasarkan nilai variabel random berdasarkan runtun waktu. Nilai variabel hasil pengamatan berada pada sumbu vertikal, sedangkan interval waktu berada pada sumbu horizontal. Plot *time series* dapat digunakan untuk mengidentifikasi pola data yang meliputi ada tidaknya unsur trend serta ada tidaknya pola *seasonal*.

Selain itu, plot *time series* dapat digunakan untuk mengidentifikasi ke-stasioner-an suatu data secara visual. Menurut Tufte dalam Cryer & Chan (2008: 8-9), plot *time series* merupakan desain grafik yang paling sering digunakan karena keteraturan pencatatan dalam detik, menit, jam, hari, pekan, bulan, tahun, serta satuan waktu lain yang menjadikan interpretasi lebih efektif. Hal ini tidak dapat ditemukan pada komposisi grafik lain.

2.3 *Time Series*

Deret Waktu (*Time Series*) didefinisikan sebagai serangkaian observasi yang diurutkan berdasarkan waktu t . Data yang diamati pada waktu t , Y_t , diasumsikan dipengaruhi oleh waktu dan pola historisnya. Analisis deret waktu bertujuan untuk memahami mekanisme yang mendasari data, serta memproyeksikan nilai di masa depan (peramalan) (Box and Jenkins, 1994).

Secara klasik, setiap deret waktu Y_t dapat diuraikan menjadi komponen-komponen utama yang menggambarkan sifat dan pola pergerakan data. Dekomposisi ini membantu dalam identifikasi karakteristik data sebelum pemodelan. Model dekomposisi dapat bersifat aditif atau multiplikatif (Makridakis, Wheelwright, & Hyndman, 1998).



2.4 Autocorrelation Function (ACF)

Autokorelasi atau yang dikenal dengan *autocorrelation function* (ACF) merupakan koefisien korelasi antara Y_t dan Y_{t-k} pada lag ke- k . ACF biasanya dinotasikan dengan k dan didefinisikan sebagai berikut (Wei, 2006).

$$\rho_k = \frac{\text{Cov}(Y_t, Y_{t+k})}{\sqrt{\text{Var}(Y_t) \cdot \text{Var}(Y_{t+k})}} = \frac{\text{Cov}(Y_t, Y_{t+k})}{\text{Var}(Y_t)} = \frac{\gamma_k}{\gamma_0}$$

dengan,

- $\rho_k = \text{ACF lag-}k$
- $\text{Var}(Y_t) = \text{Var}(Y_{t+k}) = 0$
- $k = \text{Cov}(Y_t, Y_{t-k}) = E(Y_t - \mu)E(Y_{t+k} - \mu)$

Secara matematis, fungsi autokorelasi yang dihitung berdasarkan sampel ($\hat{\rho}_k$) dari suatu data *time series* $\{Y_t\}$ dapat dituliskan sebagai berikut.

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}; k = 0, 1, 2, \dots$$

dengan $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$

2.5 Partial Autocorrelation Function (PACF)

Partial Autocorrelation Function (PACF) merupakan sebuah fungsi dari ACF yang sangat berguna untuk menentukan orde p dari model AR (Tsay, 2010). Pada analisis *time series*, PACF dinotasikan dengan $\hat{\phi}_{kk}$ persamaan sebagai berikut (Wei, 2006).

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_j}$$

dan

$$\hat{\phi}_{k+1,j} = \hat{\phi}_{k+1,k+1} \hat{\phi}_{k,k+1-j}; j = 1, 2, 3, \dots, k$$

dengan,

- $\hat{\rho}_k$ = sampel ACF lag- k
- $\hat{\phi}_{k+1,j}$ = sampel PACF lag- k

2.6 Autoregressive Integrated Moving Average (ARIMA)

Model ARIMA merupakan model gabungan atau kombinasi dari model *Autoregressive* (AR) orde p dan *Moving Average* (MA) orde q serta diikuti oleh proses *differencing* orde d . Model ini biasa digunakan untuk data yang bersifat linier dan univariat. Secara umum, bentuk model ARIMA (p, d, q) adalah sebagai berikut (Wei, 2006).

$$\phi(B)(1-B)^d Y_t = \theta_0 + \theta(B)a_t$$

Keterangan:

- p = orde dari *autoregressive* (AR)
- q = orde dari *moving average* (MA)
- d = orde dari *differencing*



Laporan Analisis Data *Time Series*

- $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$
- $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$
- $\theta_0 = (1 - \phi_1 - \dots - \phi_p)$
- $(1 - B)Y_t = Y_t - Y_{t-1}$
- a_t merupakan barisan *white noise* dengan rata-rata nol dan varians konstan sehingga $a_t \sim WN(0, \sigma^2)$
- B merupakan operator *backshift*.

2.7 Model SARIMAX

Model SARIMAX (*Seasonal Autoregressive Integrated Moving Average with Exogeneous Regressors*) adalah perluasan metodologis yang signifikan dari model ARIMA yang dikembangkan oleh Box dan Jenkins, yang dirancang khusus untuk menganalisis data deret waktu yang dipengaruhi oleh faktor internal maupun eksternal. Model ini secara efektif memadukan dua konsep pemodelan utama, yaitu proses stokastik deret waktu dan model regresi.

Inti dari SARIMAX adalah model ARIMA(p, d, q) dan SARIMA (P, D, Q)_s. Model ARIMA memodelkan ketergantungan deret waktu pada nilai masa lalunya dan *error* masa lalu setelah deret distasionerkan melalui diferensiasi. Ketika data menunjukkan pola musiman, komponen musiman SARIMA ditambahkan untuk menangkap keteraturan yang terjadi pada interval tertentu.

Model SARIMAX mengintegrasikan variabel eksogen ke dalam struktur SARIMA. Tujuan utamanya adalah memisahkan dampak guncangan eksternal dari pola internal deret waktu yang dimodelkan oleh ARIMA atau SARIMA. Secara umum, sebuah deret waktu yang diobservasi (Y_t) dapat dirumuskan sebagai berikut.

$$Y_t = f(I_t) + N_t$$

dimana,

- Y_t = Deret waktu yang diobservasi
- $f(I_t)$ = Fungsi Transfer yang menjelaskan dampak intervensi I_t
- N_t = Komponen deret waktu yang tidak dipengaruhi intervensi, dimodelkan oleh ARIMA(p, d, q)(P, D, Q)

Dalam model SARIMAX, $f(I_t)$ diwakili oleh model regresi dengan variabel prediktor, menjadikan model ini secara teknis disebut regresi *time series*. Pendekatan ini memungkinkan analisis untuk mengukur secara eksplisit dampak variabel prediktor.

2.8 Model Evaluasi

Evaluasi model dilakukan untuk mengukur tingkat akurasi model yang umumnya menggunakan dua pendekatan utama yaitu *Root Means Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE). Nilai RMSE yang lebih kecil mengindikasikan bahwa model memberikan prediksi yang lebih akurat terhadap data aktual. RMSE didefinisikan sebagai berikut.



$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n - p}}$$

dengan p adalah jumlah parameter.

Sementara itu, MAPE mengukur kesalahan prediksi relatif terhadap nilai aktual dalam bentuk persentase menggunakan persamaan sebagai berikut.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100$$

2.9 Peramalan

Peramalan merupakan metode untuk memprediksi suatu nilai pada masa yang akan datang dengan memperhatikan data masa lalu maupun data saat ini. Dalam ilmu sosial, berbagai fenomena cenderung bersifat tidak pasti, berbeda dengan ilmu eksakta. Data seperti jumlah penduduk, pendapatan per kapita, volume penjualan, dan tingkat konsumsi selalu mengalami perubahan yang dipengaruhi oleh banyak faktor-faktor yang sangat kompleks. Karena dinamika tersebut sulit diprediksi secara tepat, dalam hal ini diperlukan adanya peramalan (*forecast*).

Seperti telah dijelaskan, fenomena dalam ilmu sosial penuh dengan ketidakpastian dan sulit diprediksi secara akurat. Oleh karena itu, tujuan utama peramalan adalah mengurangi dampak ketidakpastian tersebut terhadap hasil prediksi. Dengan kata lain, peramalan diarahkan untuk menghasilkan estimasi yang memiliki tingkat kesalahan sekecil mungkin (*forecast error*), yang umumnya diukur menggunakan mean squared error, mean absolute error, dan ukuran lainnya (Subagyo, 2009).

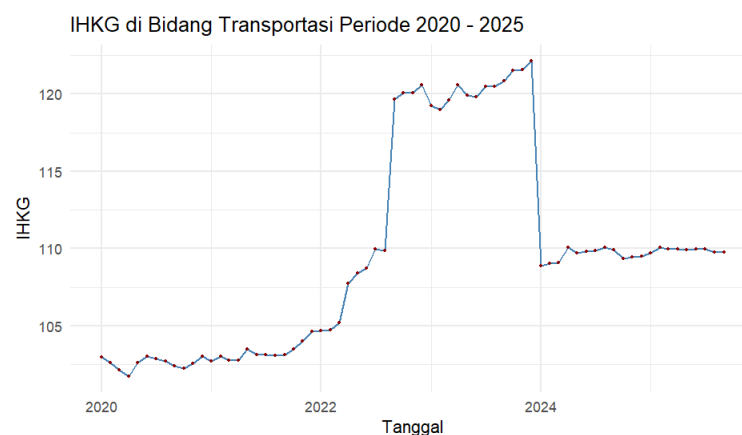


Bab III

Analisis dan Pembahasan

3.1 Dataset

Data yang dianalisis adalah data Indeks Harga Konsumen Gabungan (IHKG) bulanan yang difokuskan secara spesifik pada sektor transportasi. Dataset ini mencakup periode waktu yang relatif panjang, yaitu Januari 2020 hingga September 2025, menghasilkan 69 observasi bulanan. Setiap titik data merepresentasikan nilai indeks pada bulan yang bersangkutan, yang secara esensial mengukur tingkat inflasi atau perubahan harga rata-rata yang dibayar konsumen untuk barang dan jasa di sektor transportasi. Transformasi data menjadi objek *time series* dengan frekuensi bulanan dilakukan untuk memfasilitasi penerapan model deret waktu.



Gambar 1. Plot IHKG Transportasi Periode 2020 – 2025

Visualisasi deret waktu, yang ditunjukkan pada Plot IHKG Transportasi 2020 – 2025, menunjukkan beberapa hal penting berupa tren kenaikan signifikan dari tahun 2020 hingga puncaknya di akhir 2023 yang mencerminkan periode inflasi yang tinggi dan anomali ekstrem berupa penurunan tajam pada awal tahun 2024. Pergerakan yang mendadak ini mengindikasikan bahwa data IHKG transportasi sangat sensitif terhadap peristiwa eksternal (intervensi), menjadikannya kandidat ideal untuk pemodelan menggunakan analisis intervensi dan deteksi *outlier*.

Kesesuaian data IHKG transportasi untuk dimodelkan menggunakan analisis intervensi didasarkan pada karakteristik data yang diamati. Tren kenaikan tajam pada tahun 2022 hingga 2023 kemungkinan besar disebabkan oleh kenaikan harga energi global atau kebijakan harga domestik, yang berperan sebagai intervensi lingkungan dan memicu pergeseran tren. Lebih lanjut, anomali penurunan drastis di awal 2024 hampir pasti merupakan hasil dari intervensi struktural. Analisis intervensi memungkinkan variabel *dummy* eksogen dimasukkan secara eksplisit ke dalam model ARIMA untuk mengukur dan mengisolasi dampak terpisah dari intervensi permanen terhadap proses stokastik IHKG.

Sementara itu, deteksi *outlier* dibutuhkan untuk mengidentifikasi dan memodelkan pencilan yang sifatnya lebih sementara, seperti *additive outlier* (AO) atau *pulse*, yang mungkin disebabkan oleh lonjakan musiman yang tidak tertangkap oleh komponen musiman model. Dengan menggabungkan kedua teknik, model dapat menjelaskan variabilitas data yang



disebabkan oleh peristiwa fundamental dan peristiwa anomali, yang esensial untuk memitigasi bias estimasi parameter dan meningkatkan akurasi peramalan.

Untuk memastikan validitas eksternal model yang dikembangkan, dataset dibagi secara temporal menjadi dua segmen:

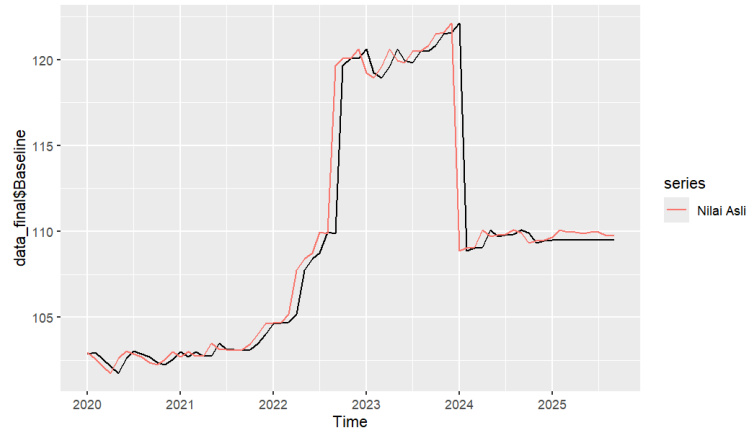
- a. Data pelatihan: periode yang digunakan untuk estimasi parameter model. Segmen ini mencakup 60 observasi, yaitu dari Januari 2020 hingga Desember 2024. Pemilihan batas ini didasarkan pada strategi untuk memastikan model yang telah dilatih mencakup intervensi struktural besar yang terjadi pada Januari 2024, sehingga model dapat mempelajari dan menginternalisasi pergeseran level baru sebelum melakukan prediksi.
- b. Data pengujian: periode yang digunakan untuk validasi peramalan. Segmen ini terdiri dari 9 observasi data, yaitu dari Januari 2025 hingga September 2025. Kinerja peramalan pada data pengujian dievaluasi menggunakan metrik akurasi standar untuk menentukan model yang paling unggul dalam memproyeksikan pergerakan IHKG di masa depan.

3.2 Pemodelan dasar ARIMA untuk Peramalan Indeks Harga Konsumen Gabungan Sektor Transportasi

Model dasar (*baseline*) untuk peramalan IHKG Transportasi diperoleh melalui prosedur pemilihan model otomatis di *Rstudio* yang menghasilkan spesifikasi ARIMA(0, 1, 0) tanpa konstanta. Berikut adalah model ARIMA yang terbentuk.

$$Y_t - Y_{t-1} = \alpha_t$$

Model ini adalah bentuk yang paling sederhana, mengimplikasikan bahwa data IHKG menjadi stasioner setelah dilakukan diferensiasi tingkat pertama dan tidak memerlukan komponen *Autoregressive* maupun *Moving Average* untuk memodelkan korelasi sisa. Secara praktis, model ARIMA(0, 1, 0) menunjukkan bahwa nilai IHKG bulan ini adalah perkiraan terbaik untuk nilai bulan berikutnya setelah memperhitungkan perbedaan sebelumnya. Berikut adalah plot perbandingan antara data IHKG dengan hasil ramalannya.

**Gambar 2.**

Data Asli IHKG vs. Data Peramalan dengan ARIMA(0, 1, 0)

Plot perbandingan antara nilai asli IHKG Transportasi dengan hasil *fitting* dan peramalan model ARIMA(0, 1, 0) menunjukkan bahwa model memiliki kemampuan yang memadai untuk menangkap tren umum data dan mereplikasi nilai data pelatihan selama periode 2020 hingga 2023. Namun, kelemahan mendasar model tersebut ditampilkan secara mencolok pada awal 2024, di mana terjadi kegagalan fatal dalam memodelkan penurunan vertikal diskontinu IHKG yang disebabkan oleh intervensi struktural, seperti *rebasing*. Model ARIMA(0, 1, 0), yang hanya mengandalkan nilai periode sebelumnya, tidak memiliki mekanisme untuk mengantisipasi dan menyesuaikan diri dengan pergeseran level eksternal yang masif tersebut.

Meskipun model kemudian berhasil menangkap tingkat stabilitas harga rata-rata yang baru, ketidakmampuan untuk menjelaskan diskontinuitas ini menegaskan bahwa model ARIMA(0, 1, 0) tidak valid secara struktural, sehingga membenarkan urgensi penggunaan analisis intervensi dan deteksi *outlier* untuk meningkatkan akurasi peramalan.

3.3 Pemodelan Lanjutan ARIMA dengan Penambahan *Outlier* untuk Peramalan Indeks Harga Konsumen Gabungan Sektor Transportasi

Model lanjutan diperoleh dengan menambahkan variabel eksogen *Level Shift* (LS) yang diidentifikasi dari deteksi *outlier* ke dalam model SARIMA(1, 0, 0)(1, 0, 0)[12]. Spesifikasi model ini, yaitu regresi dengan error ARIMA(1, 0, 0)(1, 0, 0)[12], ditujukan untuk memodelkan proses stokastik dasar data IHKG Transportasi sembari mengisolasi efek dari tiga pergeseran level harga yang signifikan, yaitu pada data ke-28, 33, dan 49. Berikut adalah bentuk model SARIMAX yang terbentuk.

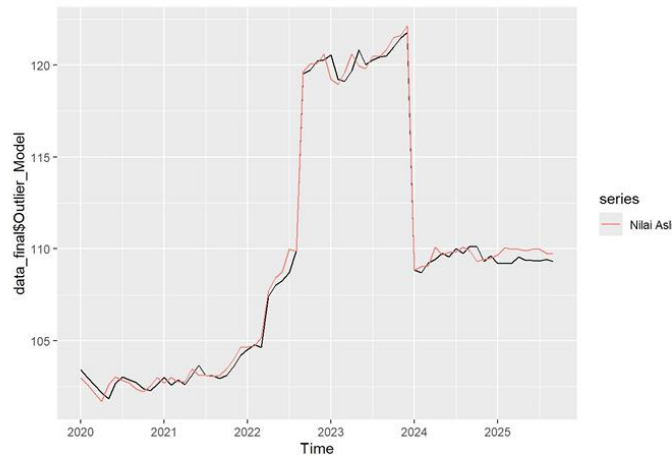
$$Y_t = \mu_t + 0,3988(Y_{t-12} - \mu_{t-12}) + 0,9812(Y_{t-1} - \mu_{t-1}) - 0,3909(Y_{t-13} - \mu_{t-13}) + \alpha_t$$

$$\text{dengan } \mu_t = 106,5808 + 2,2451.LS(28)_t + 9,6532.LS(33)_t - 12,6811.LS(49)_t$$

Model ini memiliki koefisien AR(1) sebesar 0,9812 dan koefisien SAR(1) sebesar 0,3988 yang dapat menangkap ketergantungan musiman dan non-musiman yang kuat dalam data. Lebih penting lagi, koefisien LS49 sebesar -12,6811 secara eksplisit mengukur dan menyesuaikan dampak penurunan level struktural yang diamati pada Januari 2024, yang tidak dapat dijelaskan oleh model dasar sebelumnya. Model ini menunjukkan peningkatan dramatis



dalam kemampuan *fitting* data, yang memungkinkannya mereplikasi secara akurat titik diskontinuitas tersebut, serta menghasilkan peramalan yang jauh lebih stabil dan kredibel pada periode 2024 hingga 2025. Berikut adalah plot perbandingan antara data IHKG dengan hasil ramalannya.



Gambar 3.

Data Asli IHKG vs. Data Peramalan dengan
 $ARIMA(1, 0, 0)(1, 0, 0)_{12} + \text{Regresi}(LS28, LS33, LS49)$

Plot perbandingan antara nilai asli IHKG Transportasi dengan hasil *fitting* dan peramalan model *outlier* ini menunjukkan adanya perbaikan besar dalam kemampuan model untuk menyesuaikan diri dengan anomali struktural data. Model regresi dengan *error* SARIMA kini mampu mereplikasi pergeseran level yang sangat tajam pada awal 2024 secara akurat, ditandai dengan garis hitam yang hampir berimpit dengan garis merah di titik diskontinuitas tersebut. Selain itu, model menunjukkan *fitting* yang sangat baik untuk periode kenaikan harga yang tinggi pada tahun 2022 hingga 2023. Kemampuan model *outlier* secara eksplisit memodelkan anomali ini menghasilkan peramalan yang jauh lebih stabil dan akurat pada periode 2024 hingga 2025 dibandingkan dengan model dasar sebelumnya.

3.4 Pemodelan Lanjutan ARIMA dengan Penambahan Intervensi untuk Peramalan Indeks Harga Konsumen Gabungan Sektor Transportasi

Pengembangan model ARIMA dengan pendekatan intervensi dilakukan untuk mengatasi kelemahan model dasar dalam menangani perubahan struktural pada data IHKG Transportasi. Berdasarkan analisis pola data, teridentifikasi adanya intervensi eksternal pada pertengahan tahun 2022 dan awal tahun 2024 yang menyebabkan perubahan tingkat permanen dalam deret waktu. Pada bulan September 2022, terjadi penyesuaian harga Bahan Bakar Minyak (BBM) bersubsidi yang memicu inflasi tajam dan permanen di sektor transportasi, sedangkan pada Januari 2024, kemungkinan besar penurunan drastis IHKG Transportasi disebabkan oleh perubahan tahun dasar (*rebasing*) IHKG oleh Badan Pusat Statistik (BPS) yang mulai diberlakukan pada awal tahun tersebut.

Untuk menangani masalah intervensi tersebut, diperoleh model lanjutan dengan menambahkan variabel eksogen *level shift* yang merepresentasikan dua pergeseran level kunci yang terobservasi, ke dalam model $SARIMA(0, 1, 0)(1, 0, 0)[12]$. Spesifikasi model ini, yaitu



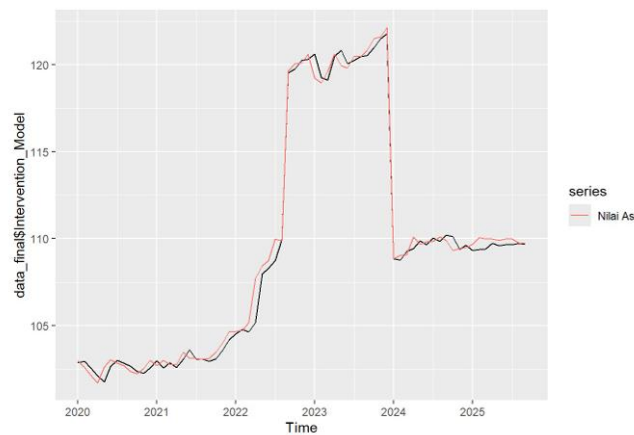
Laporan Analisis Data *Time Series*



regresi dengan *error* SARIMA(0, 1, 0)(1, 0, 0)₁₂, bertujuan untuk memodelkan proses stokastik dasar IHKG Transportasi sembari mengisolasi efek dari dua intervensi signifikan. Berikut adalah model SARIMA yang terbentuk.

$$Y_t = Y_{t-1} + 0,3448(Y_{t-12} - Y_{t-13}) + 9,6306.LS(33)_t - 12,8107.LS(49)_t + \alpha_t$$

Model ini memiliki koefisien SAR(1) sebesar 0,3448 yang menangkap ketergantungan musiman, koefisien LS33 sebesar 9,6306 yang mengukur dampak kenaikan harga pada September 2022, dan koefisien LS49 yang secara eksplisit mengukur dan menyesuaikan dampak penurunan level struktural (seperti *rebasing*) yang diamati pada Januari 2024. Model ini menunjukkan peningkatan dramatis dalam kemampuan *fitting* data dibandingkan model *baseline*, yang memungkinkannya mereplikasi secara akurat titik diskontinuitas tersebut, serta menghasilkan peramalan yang lebih stabil dan kredibel pada periode 2024 hingga 2025. Berikut adalah plot perbandingan data IKHG dengan hasil ramalannya.



Gambar 4.

Data Asli IHKG vs. Data Peramalan dengan
SARIMA(0, 1, 0)(1, 0, 0)₁₂ + Regresi(LS33, LS49)

Plot perbandingan antara nilai asli IHKG Transportasi dengan hasil *fitting* dan peramalan model intervensi ini menunjukkan adanya perbaikan besar dalam kemampuan model untuk menyesuaikan diri dengan anomali struktural data. Model regresi dengan *error* SARIMA kini mampu mereplikasi pergeseran level yang sangat tajam pada awal 2024 secara akurat, ditandai dengan garis prediksi yang hampir berimpit dengan nilai aktual di titik diskontinuitas tersebut. Selain itu, model menunjukkan *fitting* yang sangat baik untuk periode kenaikan harga yang tinggi pada tahun 2022 hingga 2023. Kemampuan model intervensi secara eksplisit memodelkan anomali ini menghasilkan peramalan yang jauh lebih stabil dan akurat pada periode 2024 hingga 2025 dibandingkan dengan model dasar sebelumnya.

3.5 Pemodelan Lanjutan ARIMA dengan Integrasi Outlier dan Intervensi untuk Peramalan Indeks Harga Konsumen Gabungan Sektor Transportasi

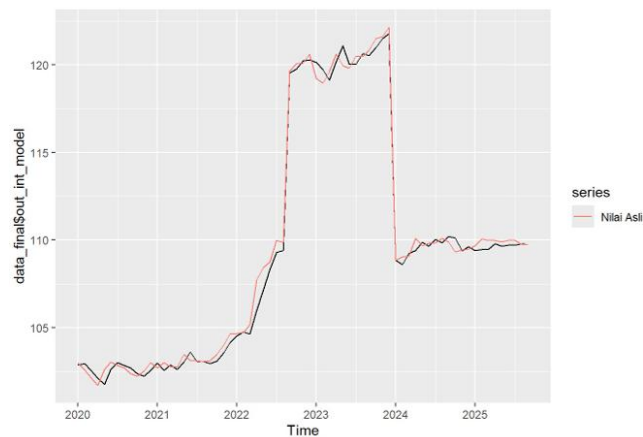
Pengembangan model ARIMA dengan penambahan intervensi dilanjutkan untuk meningkatkan akurasi *fitting* dan peramalan dengan mengakomodasi pencila sementara (*additive outlier*) selain pergeseran level struktural (*level shift*). Model yang diperoleh adalah

regresi dengan error SARIMA(0, 1, 0)(1, 0, 0)₁₂, yang memasukkan LS33 dan LS49, serta tiga variabel AO28, AO31, dan AO37 yang merepresentasikan *error* sementara (*pulse*) dari residual model intervensi sebelumnya. Berikut adalah model SARIMA yang terbentuk.

$$Y_t = Y_{t-1} + 0,3219(Y_{t-12} - Y_{t-13}) + 9,6375.LS(33)_t - 12,9995.LS(49)_t + \mu_t + \alpha_t$$

$$\text{dengan } \mu_t = 0,7879.AO(28)_t + 0,5506.AO(31)_t + 0,4876.AO(37)_t$$

Model ini memiliki koefisien SAR(1) sebesar 0,3219 yang menangkap ketergantungan musiman dan koefisien LS49 sebesar -12,9995 yang mengukur dampak penurunan level permanen IHKG pada Januari 2024. Penanganan simultan terhadap *additive outlier* dan *level shift* ini menghasilkan spesifikasi model yang paling akurat. Berikut adalah plot perbandingan data IHKG dengan hasil ramalannya.



Gambar 4. Data Asli IHKG vs. Data Peramalan dengan SARIMA(0, 1, 0)(1, 0, 0)₁₂ + Regresi(LS33, LS49, AO28, AO31, AO37)

Plot perbandingan antara nilai asli IHKG Transportasi dengan hasil *fitting* dan peramalan model *outlier* dan intervensi ini menunjukkan adanya perbaikan substansial dalam kemampuan model untuk menyesuaikan diri dengan semua anomali data. Model regresi dengan error SARIMA kini mampu mereplikasi pergerakan level yang sangat tajam di awal 2024 secara akurat, ditandai dengan garis prediksi yang hampir berimpit dengan nilai aktual di titik diskontinuitas tersebut. Selain itu, model menunjukkan *fitting* yang sangat halus pada fluktuasi minor selama periode kenaikan harga di tahun 2022 hingga 2023. Kemampuan model gabungan *outlier* dan intervensi untuk memodelkan anomali baik yang bersifat permanen maupun sementara ini menghasilkan peramalan yang paling stabil dan akurat pada periode 2024 hingga 2025 dibandingkan dengan model-model sebelumnya.

3.6 Pemilihan Model Terbaik

Pemilihan model peramalan IHKG Transportasi terbaik dilakukan berdasarkan evaluasi komparatif metrik akurasi pada data pengujian karena ini merupakan indikator yang paling valid untuk kinerja prediksi di masa depan. Ada empat model yang diuji, yaitu model dasar ARIMA, model SARIMA dengan pendekatan *outlier*, model SARIMA dengan pendekatan intervensi, dan model SARIMA dengan integrasi *outlier* dan intervensi. Berikut adalah tabel perbandingan antar model dengan berbagai kriteria.



Tabel 1. Pemilihan Model Terbaik				
Nama Model	ARIMA (0, 1, 0)	SARIMA (1, 1, 0) (1, 0, 0)[12] + <i>outlier</i>	SARIMA (0, 1, 0) (1, 0, 0)[12] + <i>intervensi</i>	SARIMA (0, 1, 0) (1, 0, 0)[12] + <i>outlier</i> + <i>intervensi</i>
Signifikansi Parameter	Tidak ada parameter	√	√	X
<i>White Noise</i>	√	√	√	√
Distribusi Normal	X	√	X	X
Homokedastisitas	√	√	√	√
AIC	263.75	91.99	104.51	101.21
BIC	265.82	106.65	112.82	115.75
RMSE <i>in-sample</i>	2,2052948	0,4403786	0,5368553	0,4970509
RMSE <i>out-sample</i>	0,4310839	0,5773956	0,3791941	0,3296231

Tabel 1 menyajikan hasil komparasi empat model deret waktu, di mana model SARIMA dengan penambahan *outlier* dan *intervensi* terbukti unggul dalam hal akurasi peramalan, meskipun gagal pada beberapa uji diagnostik *in-sample*. Secara diagnostik residual, semua model lanjutan berhasil memenuhi asumsi fundamental *white noise* dan homoskedastisitas. Namun, hanya model SARIMA dengan penambahan *outlier* yang memiliki residual terdistribusi normal. Selain itu, model SARIMA dengan penambahan *outlier* dan *intervensi* menunjukkan adanya parameter yang tidak signifikan, mengindikasikan adanya *over-specification* atau kelebihan parameter yang tidak perlu, sementara model SARIMA dengan penambahan *outlier* memenangkan kriteria informasi AIC dan BIC terendah.

Meskipun model SARIMA dengan penambahan *outlier* unggul berdasarkan kriteria *parsimony* dan diagnostik *in-sample*, nilai akurasi pada RMSE data pengujian berfungsi sebagai penentuan akhir. SARIMA dengan penambahan *outlier* dan *intervensi* secara tegas menunjukkan kinerja prediksi terbaik dengan mencatatkan RMSE data pengujian terendah sebesar 0,3296231, diikuti oleh SARIMA dengan penambahan *intervensi* dan model dasar. Keunggulan ini menunjukkan bahwa penanganan simultan terhadap *level shift* permanen dan *additive outlier* menghasilkan model yang paling *robust* dan memiliki daya generalisasi tertinggi untuk memprediksi IHKG Transportasi di masa depan. Kegagalan pada uji normalitas pada model terbaik sering diabaikan selama asumsi *white noise* terpenuhi dan akurasi prediksi tertinggi dicapai.

3.7 Keterbatasan Penelitian

Adapun keterbatasan dalam studi ini, antara lain :

1. Penentuan titik *intervensi* dan *outlier* hanya didasarkan pada analisis statistik tanpa verifikasi kualitatif lebih lanjut mengenai kejadian spesifik yang menyebabkan perubahan struktural pada data IHKG Transportasi.
2. Model yang dikembangkan sangat bergantung pada pola historis dan mungkin kurang responsif terhadap perubahan struktural mendadak atau kebijakan baru yang belum tercermin dalam data periode pelatihan.



Laporan Analisis Data *Time Series*



3. Integrasi antara pendekatan outlier dan intervensi meningkatkan kompleksitas model secara signifikan, yang dapat menurunkan kemampuan generalisasi model pada data baru di masa depan.



Bab IV

Penutup

4.1 Kesimpulan

Penelitian ini bertujuan memproyeksikan pergerakan Indeks Harga Konsumen Gabungan Sektor Transportasi (IHKG Transportasi) periode Januari 2020 hingga September 2025. Hasil analisis menunjukkan bahwa model *baseline* ARIMA(0, 1, 0) terbukti tidak memadai karena gagal mereplikasi anomali struktural diskontinu, seperti penurunan level tajam pada Januari 2024 akibat *rebasing* IHKG, yang tercermin dari nilai RMSE *out-sample* (data pengujian) yang sangat tinggi (4,310,839). Berdasarkan evaluasi komparatif akurasi peramalan, model terbaik yang dipilih adalah SARIMA(0, 1, 0)(1, 0, 0)[12] dengan integrasi *outlier* dan intervensi (SARIMAX).

Keunggulan model ini didukung oleh nilai RMSE *out-sample* **terendah (0,3296)**, yang merupakan indikator validitas kinerja prediksi masa depan. Integrasi variabel regresi eksogen (*Level Shift* dan *Additive Outlier*) secara eksplisit memungkinkan model mengestimasi dampak peristiwa non-stokastik yang signifikan (seperti *rebasing* dan penyesuaian harga BBM), menghasilkan *fitting* data yang akurat pada titik diskontinuitas dan, pada akhirnya, peramalan yang paling stabil dan kredibel untuk IHKG Transportasi dalam jangka pendek.

4.2 Saran

Penulis mengucapkan terima kasih kepada dosen pengampu mata kuliah Pengantar Analisis Deret Waktu atas bimbingan yang diberikan selama ini. Terima kasih juga disampaikan kepada seluruh rekan tim yang telah bekerja sama dalam melakukan analisis data dan penyusunan laporan.



Daftar Pustaka

- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control* (5th ed.). John Wiley & Sons.
- Box, G. E. P., & Tiao, G. C. (1975). Intervention Analysis with Applications to Economic and Environmental Problems. *Journal of the American Statistical Association*, 70(352), 70-79.
- Budiarti, L., Tarno, & Warsito, B., (2013). Analisis Intervensi dan Deteksi Outlier pada Data Wisatawan Domestik (Studi Kasus di Daerah Istimewa Yogyakarta). *Jurnal Gaussian*, 2(1), 39 – 48.
- Chen, C., & Liu, L. M. (1993). Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association*, 88(421), 284-290.
- Cryer, J. D., & Chan, K.-S. (2008). *Time Series Analysis With Applications in R, Second Edition*. New York: Springer
- Direktorat Statistik Harga. (2025). *Indeks Harga Konsumen 38 Provinsi di Indonesia 2024 (2022 = 100)*. Badan Pusat Statistik
- Minarnik. 2009. *Analisis Time Series terhadap Indeks Harga Konsumen (IHK) di Indonesia dengan Menggunakan Metode Intervensi untuk Mengatasi Outlier*. Institut Teknologi Sepuluh Nopember
- Subdirektorat Statistik Harga Konsumen. (2020). *Indeks Harga Konsumen di 82 Kota di Indonesia (2012 = 100) 2019*. BPS RI.
- Wei, W. W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods, Second Edition*. United States: Pearson Education, Inc.



Lampiran

Pemodelan ARIMA dasar (*baseline*)

Series: IHKG_ts_train\$Indeks

ARIMA(0,1,0)

$\sigma^2 = 4.946$: log likelihood = -130.87

AIC=263.75 AICc=263.82 BIC=265.82

Ljung-Box test

data: Residuals from ARIMA(0,1,0)

$Q^* = 1.4908$, df = 12, p-value = 0.9999

Model df: 0. Total lags used: 12

Shapiro-Wilk normality test

data: base_model\$residuals

$W = 0.42481$, p-value = $5.845e-14$

ARCH LM-test; Null hypothesis: no ARCH effects

data: base_model\$residuals

Chi-squared = 1.4201, df = 12, p-value = 0.9999

	ME	RMSE	MAE
Training set	0.1102162	2.2052948	0.7778828
Test set	0.4122222	0.4310839	0.4122222
	MPE	MAPE	MASE
Training set	0.08510272	0.6973543	0.10808619
Test set	0.37498373	0.3749837	0.05727795
	ACF1	Theil's U	
Training set	-0.014130847	NA	
Test set	0.001750664	2.735638	

Pemodelan SARIMA(1, 0, 0)(1, 0, 0)[12] + [LS28, LS33, LS49]



Laporan Analisis Data *Time Series*



Series: IHKG_ts_train\$Indeks
Regression with ARIMA(1,0,0)(1,0,0)[12] errors

Coefficients:

	ar1	sar1	intercept	LS28	LS33
	0.9812	0.3988	106.5805	2.2451	9.6532
s.e.	0.0181	0.1356	2.9370	0.4235	0.4162
					LS49
					-12.6811
s.e.					0.4865

sigma^2 = 0.2155: log likelihood = -39
AIC=91.99 AICc=94.15 BIC=106.65

Ljung-Box test

data: Residuals from Regression with
ARIMA(1,0,0)(1,0,0)[12] errors
Q* = 4.0208, df = 10, p-value = 0.9464

Model df: 2. Total lags used: 12

Shapiro-wilk normality test

data: model_with_outlier\$residuals
W = 0.97153, p-value = 0.1733

ARCH LM-test; Null hypothesis: no ARCH
effects

data: model_with_outlier\$residuals
Chi-squared = 12.487, df = 12, p-value =
0.4074

	ME	RMSE	MAE
Training set	0.0749322	0.4403786	0.3465628
Test set	0.5555534	0.5773956	0.5555534
	MPE	MAPE	MASE
Training set	0.06702853	0.3153338	0.04815463
Test set	0.50542818	0.5054282	0.07719370
	ACF1	Theil's U	
Training set	0.06068401	NA	
Test set	0.12537337	3.560998	

N = 60	Koefisien (df = 54)					
k = 6	Intercept	AR(1)	SAR(1)	LS28	LS33	LS49
Nilai	106,5805	0,9812	0,3988	2,2451	9,6532	-12,6811
s.e.	2,937	0,0181	0,1356	0,4235	0,4162	0,4865
t	36,2889	54,20994	2,941003	5,301299	23,19366	-26,066
signifikan atau tidak?	Signifikan	Signifikan	Signifikan	Signifikan	Signifikan	Signifikan



Pemodelan SARIMA(0, 1, 0)(1, 0, 0)[12] + [LS33, LS49]

Series: IHKG_ts_train\$Indeks

Regression with ARIMA(0,1,0)(1,0,0)[12] errors

Coefficients:

	sar1	xreg1	xreg2
	0.3448	9.6306	-12.8107
s.e.	0.1242	0.5130	0.5681

sigma² = 0.3088: log likelihood = -48.26

AIC=104.51 AICc=105.25 BIC=112.82

Ljung-Box test

data: Residuals from Regression with
ARIMA(0,1,0)(1,0,0)[12] errors

Q* = 13.489, df = 11, p-value = 0.2626

Model df: 1. Total lags used: 12

Shapiro-Wilk normality test

data: model_with_intervention\$residuals

W = 0.87458, p-value = 1.756e-05

ARCH LM-test; Null hypothesis: no ARCH
effects

data: model_with_intervention\$residuals

Chi-squared = 5.5272, df = 12, p-value =
0.938

	ME	RMSE	MAE
Training set	0.1086696	0.5368553	0.3668113
Test set	0.3196470	0.3791941	0.3196470
	MPE	MAPE	MASE
Training set	0.1012754	0.3353200	0.05096813
Test set	0.2907348	0.2907348	0.04441469
	ACF1	Theil's U	
Training set	0.1548722	NA	
Test set	0.4597214	2.30733	



Laporan Analisis Data *Time Series*



N = 59	Koefisien (df = 56)		
k = 3	SAR(1)	LS33	LS49
Nilai	0,3448	9,6306	-12,8107
s.e.	0,1242	0,513	0,5681
t	2,77617	18,7731	-22,5501
signifikan atau tidak?	Signifikan	Signifikan	Signifikan

Pemodelan SARIMA(0, 1, 0)(1, 0, 0)[12] + [LS33, LS49, AO28, AO31, AO37]

Series: IHKG_ts_train\$Indeks

Regression with ARIMA(0,1,0)(1,0,0)[12] errors

Coefficients:

```
      sar1      xreg1      xreg2      xreg3      xreg4
0.3219  9.6375 -12.9995  0.7879  0.5506
s.e.   0.1276  0.4786   0.5293  0.3401  0.3391
      xreg5
-0.4876
s.e.   0.3482
```

sigma^2 = 0.2797: log likelihood = -43.6

AIC=101.21 AICc=103.4 BIC=115.75

Ljung-Box test

data: Residuals from Regression with

ARIMA(0,1,0)(1,0,0)[12] errors

Q* = 17.307, df = 11, p-value = 0.09912

Model df: 1. Total lags used: 12

Shapiro-Wilk normality test

data: outlier_intervention\$residuals

W = 0.95967, p-value = 0.04535

ARCH LM-test; Null hypothesis: no ARCH
effects

data: outlier_intervention\$residuals

Chi-squared = 7.3497, df = 12, p-value =
0.8337



Laporan Analisis Data *Time Series*



	ME	RMSE	MAE
Training set	0.1153417	0.4970509	0.3826915
Test set	0.2650114	0.3296213	0.2700009
	MPE	MAPE	MASE
Training set	0.1070135	0.3485037	0.05317468
Test set	0.2410178	0.2455636	0.03751641
	ACF1	Theil's U	
Training set	0.249621	NA	
Test set	0.447244	2.022488	

N = 59	Koefisien (df = 53)					
k = 6	SAR(1)	LS33	LS49	AO28	AO31	AO37
Nilai	0,3219	9,6375	-12,9995	0,7879	0,5506	-0,4876
s.e.	0,1276	0,4786	0,5293	0,3401	0,3391	0,3482
t	2,52273	20,13686	-24,5598	2,316672	1,62371	-1,40034
signifikan atau tidak?	Signifikan	Signifikan	Signifikan	Signifikan	Tidak Signifikan	Tidak Signifikan