

Aplikasi Model ARIMAX dengan Efek Variasi Kalender untuk Peramalan Trend Pencarian Kata Kunci “Zalora” pada Data Google Trends

Andrea Tri Rian Dani^{1*}, Sri Wahyuningsih¹, Fachrian Bimantoro Putra¹,
Meirinda Fauziyah¹, Sri Wigantono², Hardina Sandariria², Qonita Qurrota
A'yun², Muhammad Aldani Zen¹

Received: 25 January 2023
Revised: 16 September 2023
Accepted: 19 September 2023

¹Program Studi Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman, Samarinda

²Program Studi Matematika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman, Samarinda

*Corresponding author: andreatriandani@fmipa.unmul.ac.id

ABSTRAK – ARIMAX menjadi salah satu metode dalam analisis runtun waktu yang digunakan untuk memodelkan suatu peristiwa dengan menambahkan variabel eksogen sebagai informasi tambahan. Saat ini, model ARIMAX dapat diterapkan pada data runtun waktu yang memiliki efek variasi kalender. Secara singkat, variasi kalender terjadi akibat perubahan komposisi pada kalender. Tujuan dari penelitian ini adalah mengaplikasikan model ARIMAX dengan efek variasi kalender untuk peramalan trend pencarian kata kunci “Zalora”. Data dikumpulkan dimulai dari periode Januari 2018 sampai dengan November 2022 dalam bentuk series mingguan. Berdasarkan hasil analisis, diperoleh model ARIMAX dengan efek variasi kalender dengan sisaan ARIMA (1,1,1). Ketepatan akurasi peramalan menggunakan Mean Absolute Percentage Error (MAPE) sebesar 10.47%. Hasil peramalan untuk 24 periode kedepan cenderung berfluktuasi dan diperkirakan pada bulan April 2023 akan terjadi peningkatan trend pencarian kata kunci “Zalora”.

Kata kunci– ARIMAX, variasi kalender, ZALORA

ABSTRACT – ARIMAX is a method in time series analysis that is used to model an event by adding exogenous variables as additional information. Currently, the ARIMAX model can be applied to time series data that has calendar variation effects. In short, calendar variations occur due to changes in the composition of the calendar. The purpose of this study is to apply the ARIMAX model with the effects of calendar variations to forecast search trends for the keyword "Zalora". Data were collected starting from January 2018 to November 2022 in the form of a weekly series. Based on the results of the analysis, the ARIMAX model is obtained with calendar variation effects with ARIMA residuals (1,1,1). Forecasting accuracy using the Mean Absolute Percentage Error (MAPE) of 10.47%. Forecasting results for the next 24 periods tend to fluctuate and it is estimated that in April 2023 there will be an increase in search trends for the keyword "Zalora".

Keywords– ARIMAX, calendar variations, ZALORA

I. PENDAHULUAN

Runtun waktu adalah serangkaian observasi yang dikolektifkan berdasarkan runtutan waktu dalam suatu periode tertentu [1], [2]. Metode analisis yang biasa digunakan pada data runtun waktu adalah analisis runtun waktu [3], [4], [5]. Peramalan (*forecasting*) adalah metode yang biasa diterapkan untuk menduga peristiwa yang akan terjadi pada masa yang akan datang dengan memperhatikan data historis maupun data sekarang [6]. Perencanaan peramalan tentunya memerlukan ketepatan dalam pemilihan metode. Hal ini dilakukan guna meminimumkan kesalahan dalam proses [7].

Metode yang sangat familiar dalam melakukan peramalan adalah *Autoregressive Integrated Moving Average* (ARIMA) [8], [9]. ARIMA dalam Box dan Jenkins [10] dijelaskan sebagai model yang terdiri dari tiga aspek, yaitu *Autoregressive* (AR), *Integrated*, dan *Moving Average* (MA). Prinsip dari pemodelan ARIMA dilakukan dengan lima tahapan pemodelan yaitu identifikasi, estimasi parameter model ARIMA, pengujian signifikansi, pemeriksaan diagnostik dan tahap akhir adalah melakukan peramalan [11], [12]. Beberapa penelitian terkini yang mengkaji menggunakan model ARIMA diantaranya: [13]–[16]. Salah satu kelemahan ARIMA adalah tidak dapat mengakomodir informasi tambahan yang diduga mempengaruhi data runtun waktu. Pada nyatanya, sering kali dijumpai beberapa data runtun waktu yang bukan hanya saja dipengaruhi oleh unsur waktu, sehingga model ARIMA tidak mampu mempotret keseluruhan informasi pada data.

Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX) dikembangkan dari model ARIMA dengan mempertimbangkan factor selain waktu dengan menggunakan variabel eksogen [7]. Variabel eksogen disini memiliki arti sebagai suatu variabel yang diduga mempengaruhi dan atau yang menjadi sebab perubahan nilai di variabel respon, baik secara positif maupun negatif [17]. Model ARIMAX dapat diterapkan pada data runtun waktu yang memiliki efek variasi kalender. Variasi kalender diindikasikan sebagai terjadinya pola berulang dengan panjang periode yang cukup beragam yang diakibatkan sebagai pengaruh penanggalan kalender yang bervariasi / berbeda-beda setiap tahunnya. Beberapa penelitian terkini yang mengkaji menggunakan model ARIMAX dengan efek variasi kalender diantaranya: [7], [18]–[20].

Zalora adalah bagian dari *Global Fashion Group* termuka didunia. Zalora merupakan salah satu *E-Commerce* yang ramai diperbincangkan di Indonesia dalam hal busana pakaian. Menjual berbagai macam brand baik lokal dan internasional yang berkualitas, sehingga tentunya akan menjadi daya tarik konsumen. Google trends merupakan laman resmi dari google yang mempotret kejadian di masyarakat berdasarkan kata kunci pencarian. Kata kunci pencarian yang akan dikaji dalam artikel ini adalah "Zalora" (<https://trends.google.com/trends/?geo=ID>). Peramalan trend pencarian kata kunci "Zalora" menjadi menarik dikarenakan pola data runtun waktu yang dihasilkan menunjukkan pola berulang akibat adanya efek variasi kalender, diduga berhubungan dengan bulan Ramadan, Natal, dan Tahun Baru.

Beberapa penelitian sebelumnya yang mengaplikasikan model ARIMAX pada data google trends diantaranya Lingga, dkk (2021) dalam penelitiannya menerapkan metode variasi kalender ARIMAX untuk pemodelan dan peramalan kedatangan wisatawan ke tempat wisata dengan google trends [21]. Dani, dkk (2023) melakukan peramalan tren pencarian kata kunci "Sarung Wadimor" di Indonesia menggunakan model time series regression with calender variation dan ARIMA Box-Jenkins [22].

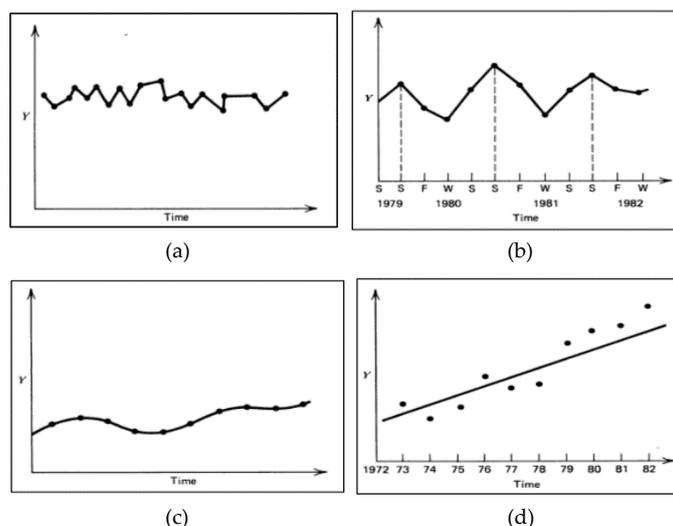
Berdasarkan uraian, maka peneliti tertarik untuk mengaplikasikan model ARIMAX dengan efek variasi kalender untuk peramalan trend pencarian kata kunci "Zalora" pada data Google Trends. Akurasi kebaikan menggunakan nilai *Mean Absolute Percentage Error* (MAPE).

II. REVIEW LITERATUR

Pada bagian ini diuraikan secara singkat definisi dari data runtun waktu, model regresi dummy, model ARIMA, dan model ARIMAX:

A. Runtun Waktu

Data runtun waktu dapat didefinisikan sebagai serangkaian / runtutan dari obervasi berdasarkan indeks waktu secara berurutan dengan selang waktu yang cenderung tetap. Pola dalam data runtun waktu secara umum dapat dibedakan menjadi empat jenis [4] seperti pada Gambar 1.



Gambar 1 Pola Data Runtun Waktu (a) Pola Horizontal; (b) Pola Musiman; (c) Pola Siklis; (d) Pola Trend

B. Regresi Dummy

Regresi Dummy merupakan generalisasi dari regresi linear yang dimana variabel prediktornya adalah kategorik [14]. Dituliskan model regresi dummy seperti pada Persamaan (1).

$$Y_t = \beta_0 + \beta_1 D_{1:t} + \beta_2 D_{2:t} + \dots + \beta_k D_{k:t} + \varepsilon_t \tag{1}$$

Dimana $\varepsilon_t \sim IIDN(0, \sigma_\varepsilon^2)$. Jika terdapat n observasi dan k variabel prediktor, maka untuk setiap $t = 1, 2, \dots, n$ dapat dijabarkan sebagai berikut:

$$Y_1 = \beta_0 + \beta_1 D_{1:1} + \beta_2 D_{2:1} + \dots + \beta_k D_{k:1} + \varepsilon_1$$

$$Y_2 = \beta_0 + \beta_1 D_{1:2} + \beta_2 D_{2:2} + \dots + \beta_k D_{k:2} + \varepsilon_2$$

⋮

$$Y_n = \beta_0 + \beta_1 D_{1:n} + \beta_2 D_{2:n} + \dots + \beta_k D_{k:n} + \varepsilon_n$$

Berdasarkan hasil penjabaran, maka dapat dinyatakan dalam notasi matriks seperti pada Persamaan (2)

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & D_{1:1} & \dots & D_{k:1} \\ 1 & D_{1:2} & \dots & D_{k:2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & D_{1:n} & \dots & D_{k:n} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2}$$

Sehingga Persamaan (2) dapat ditulis seperti pada Persamaan (3).

$$y = X\beta + \varepsilon. \tag{3}$$

Estimasi parameter dari $\hat{\beta}$ menggunakan *Ordinary Least Squares* (OLS) dituliskan pada Persamaan (4).

$$\hat{\beta} = (X^T X)^{-1} X^T y. \tag{4}$$

C. ARIMA

Model ARIMA dapat dirincikan dalam penulisan sebagai ARIMA (p, d, q) dimana orde p adalah derajat proses AR, d adalah orde *differencing*, dan orde q adalah derajat proses MA [11]. Model ARIMA (p, d, q) dapat dituliskan dalam Persamaan (5)

$$\phi_p(B)(1 - B)^d Y_t = \theta_q(B) a_t \tag{5}$$

dengan:

- $\phi_p(B) = 1 - \phi_1(B) - \phi_2(B)^2 - \dots - \phi_p(B)^p$: operator backshift (B) proses AR;
- $\theta_q(B) = 1 - \theta_1(B) - \theta_2(B)^2 - \dots - \theta_q(B)^q$: operator backshift (B) proses MA;
- Y_t : data pada waktu ke- t ;
- B : operator backshift;
- $(1 - B)^d$: operator *differencing*; dan
- d : orde *differencing*.

D. ARIMAX

Model ARIMAX pada penelitian ini dibangun dengan menambahkan komponen regresi dummy dan pemodelan residual dari model ARIMA [18]. Model ARIMAX dapat dituliskan pada Persamaan (6).

$$Y_t = \gamma t + \beta_1 D_{1:t} + \beta_2 D_{2:t} + \dots + \beta_k D_{k:t} + \frac{\theta_q(B) a_t}{\phi_p(B)(1-B)^d} Y_t. \tag{6}$$

Tahapan dalam ARIMAX hampir sama dengan tahap pemodelan ARIMA yang meliputi estimasi parameter dan pengujian signifikansi parameter, melakukan pemeriksaan diagnostik residual, dan melakukan peramalan model ARIMAX dengan efek variasi kalender.

E. Mean Absolute Percentage Error (MAPE)

Kriteria pengukuran MAPE dirumuskan seperti pada Persamaan (7).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\%. \tag{7}$$

Berdasarkan Persamaan (7), jika diperoleh nilai MAPE yang semakin kecil, berarti nilai taksiran diindikasikan akan semakin mendekati nilai aktual.

III. METODOLOGI

Pada bagian ini diuraikan mengenai tahapan analisis dari metode yang digunakan serta sumber data.

1. Data dan Sumber Data

Data trend pencarian kata kunci "Zalora" merupakan data sekunder yang diperoleh dari website <https://trends.google.com/trends/explore?date=today%205-y&geo=ID&q=zalora>. Data dikumpulkan dimulai dari periode Januari 2018 sampai dengan November 2022 dalam bentuk *series* mingguan. Adapun variabel yang digunakan adalah data trend pencarian kata kunci "Zalora" yang disimbolkan sebagai Y_t .

2. Tahapan Analisis Data

Langkah-langkah yang dilakukan dalam proses analisis dirincikan sebagai berikut:

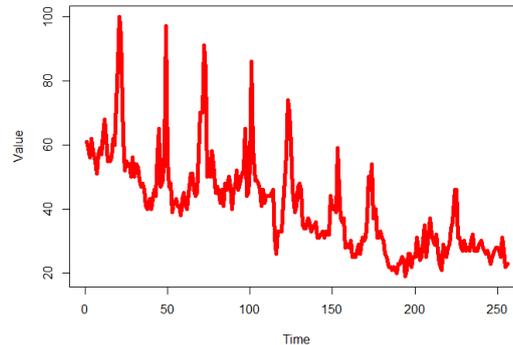
- a. Melakukan eksplorasi dan identifikasi menggunakan grafik runtun waktu.
- b. Memodelkan trend deterministik serta efek variasi kalender menggunakan regresi linier berganda terhadap data trend pencarian kata kunci "Zalora" sebagai variabel respon, variabel t dan dummy sebagai variabel prediktor.
- c. Melakukan pemeriksaan asumsi independensi residual (*white noise*) dengan menggunakan grafik Autocorrelation Function (ACF) dan pengujian hipotesis Ljung-Box. Jika residual tidak memenuhi asumsi maka langkah dilanjutkan dengan pemodelan ARIMA.
- d. Pemodelan residual pada tahap (c) dengan menggunakan model ARIMA.
- e. Pemodelan secara simultan dengan menggunakan informasi variabel t dan dummy yang signifikan, serta orde dari model ARIMA yang terbaik untuk memodelkan data trend pencarian kata kunci "Zalora" menggunakan ARIMAX.
- f. Peramalan data trend pencarian kata kunci "Zalora" menggunakan model ARIMAX terbaik.
- g. Tahap evaluasi dengan mengukur nilai akurasi kebaikan menggunakan MAPE.

IV. HASIL DAN DISKUSI

Bagian ini membahas dan menguraikan proses analisis pemodelan trend pencarian kata kunci “Zalora” menggunakan model ARIMAX.

A. Eksplorasi Data menggunakan Grafik Runtun Waktu

Grafik runtun waktu adalah langkah pertama yang dilakukan sebagai bagian dari eksplorasi data. Grafik runtun waktu disajikan pada Gambar 2.



Gambar 2 Grafik Runtun Waktu data trend pencarian kata kunci “Zalora”

Grafik runtun waktu yang disajikan pada Gambar 2 menunjukkan jika trend pencarian kata kunci “Zalora” memiliki pola yang cenderung turun dari tahun ke tahun. Namun, diidentifikasi jika pada minggu-minggu mendekati hari raya idul fitri, natal dan tahun baru terjadi peningkatan trend pencarian kata kunci “Zalora”. Hal ini ditandai dengan adanya pola kenaikan di periode-periode tertentu.

B. Pemodelan ARIMAX

a. Regresi Dummy

Berdasarkan Gambar 2, terlihat pada grafik runtun waktu jika trend pencarian kata kunci “Zalora” mengalami peningkatan di periode-periode tertentu. Peningkatan banyaknya pencarian kata kunci “Zalora” diindikasikan terjadi karena efek adanya libur Hari Raya Idul Fitri, Natal, dan Tahun Baru.

$$D_{i,t} = \begin{cases} 1, & \text{minggu dengan kejadian tertentu} \\ 0, & \text{minggu lainnya} \end{cases}$$

dengan $i = 1,2,3,4,5,6$.

Sehingga yang menjadi variabel *Dummy* dirincikan sebagai berikut:

- $D_{1,t}$: satu minggu sebelum Hari Raya Idul Fitri
- $D_{2,t}$: minggu saat Hari Raya Idul Fitri
- $D_{3,t}$: satu minggu setelah Hari Raya Idul Fitri
- $D_{4,t}$: satu minggu sebelum Natal dan Tahun Baru
- $D_{5,t}$: minggu Natal dan Tahun Baru
- $D_{6,t}$: satu minggu setelah Natal dan Tahun Baru

Selain itu, pada penelitian ini juga menambahkan efek variasi kalender dari bulan untuk menangkap efek musiman (seasonal) sebagai berikut:

$$M_{1,t} = \begin{cases} 1, & \text{bulan Januari} \\ 0, & \text{untuk periode bulan lainnya} \end{cases}$$

$$M_{2,t} = \begin{cases} 1, & \text{bulan Februari} \\ 0, & \text{untuk periode bulan lainnya} \end{cases}$$

⋮

$$M_{12,t} = \begin{cases} 1, & \text{bulan Desember} \\ 0, & \text{untuk periode bulan lainnya.} \end{cases}$$

Berdasarkan Gambar 2 terlihat bahwa grafik runtun waktu trend pencarian kata kunci “Zalora” memiliki kecenderungan pola turun, sehingga perlu dilakukan pemodelan trend deterministik. Proses pemodelan gabungan antara trend deterministik dan efek variasi kalender menggunakan regresi linear berganda dengan formula dituliskan pada Persamaan (8).

$$Y_t = \delta_0 + \delta_1 t + \beta_1 D_{1,t} + \beta_2 D_{2,t} + \dots + \beta_6 D_{6,t} + \beta_7 M_{1,t} + \dots + \beta_{18} M_{12,t} + \varepsilon_t \tag{8}$$

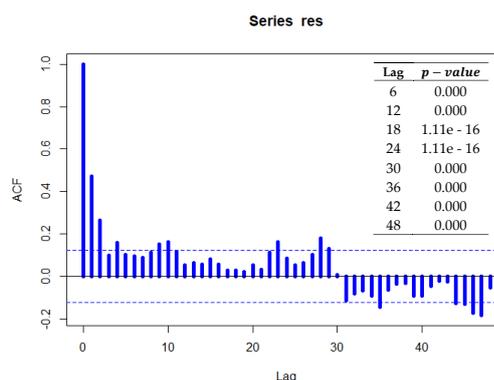
dengan $\varepsilon_t \sim IIDN(0, \sigma^2)$.

Berdasarkan hasil estimasi parameter, pengujian signifikansi baik secara simultan dan parsial, dan melakukan proses eliminasi variabel prediktor yang tidak signifikan, sehingga diperoleh parameter-parameter yang signifikan dan disajikan pada Tabel 1.

Tabel 1 Hasil Estimasi Parameter Model Regresi Dummy

Variabel	Parameter	Koefisien	p - value
Konstanta	$\hat{\delta}_0$	56,64	$< 2e - 16$
t	$\hat{\delta}_1$	-0,15	$< 2e - 16$
$D_{1;t}$	$\hat{\beta}_1$	18,58	$4,32e - 07$
$D_{4;t}$	$\hat{\beta}_4$	-11,18	0,0109
$D_{5;t}$	$\hat{\beta}_5$	-11,53	0,0087
$M_{4;t}$	$\hat{\beta}_{10}$	4,76	0,0040
$M_{5;t}$	$\hat{\beta}_{11}$	15,32	$9,94e - 11$
$M_{6;t}$	$\hat{\beta}_{12}$	5,68	0,0018
$M_{11;t}$	$\hat{\beta}_{17}$	17,67	$< 2e - 16$
$M_{12;t}$	$\hat{\beta}_{18}$	5,37	0,00267

Dengan menggunakan statistik uji F, secara simultan diperoleh nilai $p - value$ sebesar $< 2e - 16$, sehingga diperoleh keputusan Tolak H_0 yang berarti secara simultan variabel prediktor berpengaruh terhadap variabel respon. Model regresi linear berganda yang terbentuk memiliki nilai koefisien determinasi sebesar 75.26%. Proses selanjutnya adalah melakukan pemeriksaan *white noise* dari residual model regresi. Pemeriksaan asumsi *white noise* menggunakan grafik ACF dan Ljung-Box test.

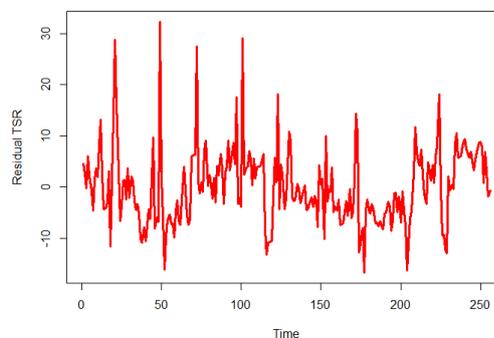


Gambar 3 Grafik ACF dari Residual Model Regresi

Berdasarkan Gambar 3, grafik ACF dan dengan melakukan pengujian Ljung-Box test, diperoleh jika *lag-lag* menunjukkan hasil belum *white noise* dengan nilai $p - value$ kurang dari taraf signifikansi 5%. Maka dapat disimpulkan residual dari model regresi linear perlu dimodelkan kembali menggunakan model ARIMA.

b. Pemodelan dengan Autoregressive Integrated Moving Average (ARIMA)

Berdasarkan uji asumsi pada residual model regresi (lihat Gambar 4 sebagai ilustrasi), diperoleh kesimpulan jika residual masih belum *white noise*. Hal ini berarti residual pada model regresi tidak memenuhi asumsi independensi, dan perlu dimodelkan menggunakan model ARIMA, untuk pemodelan ARIMAX.



Gambar 4 Grafik Runtun Waktu data Residual

1. Identifikasi Stasioneritas

Stasioneritas di kategorikan menjadi dua, yaitu stasioneritas dalam variansi dan rata-rata. Pemeriksaan stasioneritas dalam variansi menggunakan Transformasi Box-Cox, sedangkan stasioneritas dalam rata-rata menggunakan pengujian Augmented Dickey-Fuller. Pada data residual, terdapat beberapa nilai observasi yang bernilai negatif, sehingga menjadi hal yang perlu untuk ditambahkan suatu nilai konstanta agar semua data pengamatan bernilai positif. Konstanta yang digunakan adalah sebesar +20. Pemilihan konstanta sebesar 20, dimaksudkan agar semua data bernilai positif. Setelah dilakukan pemeriksaan dengan Box-Cox, didapatkan nilai λ sebesar 0,48, sehingga diindikasikan jika data masih belum stasioner dalam variansi. Untuk

mengatasinya, maka dilakukan suatu transformasi Box-Cox pada data residual, dengan $Y_t^* = (Y_t)^\lambda$ (transformasi pangkat) dimana λ adalah 0,48, Y_t adalah data residual, dan Y_t^* nilai hasil transformasi Box-Cox, $Y_t^* = (Y_t)^{0.48}$.

Setelah ditransformasi, kemudian dilakukan pemeriksaan kembali dengan Box-Cox dari data transformasi dan didapatkan nilai λ sebesar 1, sehingga diindikasikan data telah memenuhi stasioneritas dalam variansi. Langkah selanjutnya, akan dilakukan pengujian hipotesis dengan menggunakan Augmented Dickey-Fuller untuk memeriksa stasioneritas dalam rata-rata

Tabel 2 Pengujian ADF Data Residual setelah Transformasi Box-Cox

Variabel	Koefisien	t_{hit}	τ_{tabel}
ResidualTf	-0,01167	-1,097	-1,95

Berdasarkan Tabel 2, diperoleh keputusan Gagal Tolak H_0 yang artinya data masih belum stasioner dalam rata-rata. Penyelesaian permasalahan tersebut dengan melakukan *differencing*. Setelah dilakukan *differencing*, dilakukan pemeriksaan kembali menggunakan Augmented Dickey-Fuller.

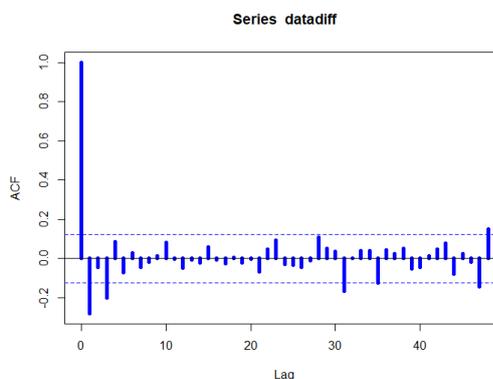
Tabel 3 Pengujian ADF Data setelah *Differencing*

Variabel	Koefisien	t_{hit}	τ_{tabel}
ResidualTf*	-1,45189	-14,504	-1,95

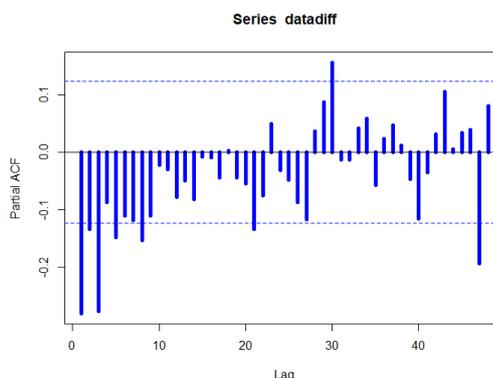
Pada Tabel 3, terlihat nilai $|t_{hit}| > |\tau_{tabel}|$, sehingga diperoleh keputusan Tolak H_0 , yang berarti data setelah *differencing* telah stasioner dalam rata-rata.

2. Pembentukan Model ARIMA

Pembentukan model ARIMA dikerjakan terhadap data residual dari model regresi linear yang sudah stasioner, baik dalam variansi dan rata-rata. Identifikasi dilakukan dengan membuat grafik ACF dan PACF yang ditampilkan pada Gambar 5 dan 6.



Gambar 5 Grafik ACF Data Setelah *Differencing*



Gambar 6 Grafik PACF Data Setelah *Differencing*

Proses selanjutnya yaitu menentukan orde pada model ARIMA. Untuk menentukan orde pada model AR dan MA dapat melihat pada grafik ACF dan PACF. Berdasarkan Gambar 5, grafik ACF, dapat diketahui orde untuk MA adalah 1 (nilai ACF *cut off* setelah lag 1). Untuk grafik PACF pada Gambar 6, dapat diketahui orde untuk AR adalah 3 (nilai PACF *cut off* setelah lag 3), dan orde untuk *differencing* (d) adalah 1.

3. Estimasi, Pengujian Signifikansi, dan Pemeriksaan Diagnostik Parameter Model ARIMA

Estimasi parameter dari model ARIMA menggunakan metode *Maximum Likelihood Estimation* (MLE). Hasil estimasi dan uji signifikansi parameter ditampilkan pada Tabel 4.

Tabel 4 Estimasi Parameter Model ARIMA

Model ARIMA Sementara	Estimasi Parameter	P-Value	Independensi Residual (P-Value)	Normalitas Residual (P-Value)		
ARIMA (0,1,1)	$\hat{\theta}_1 : -0,563$	0,000	Lag 6	0,000	0,069	
			Lag 12	0,000		
			Lag 18	0,000		
			Lag 24	0,000		
			Lag 30	0,000		
			Lag 36	0,000		
			Lag 42	0,000		
ARIMA (1,1,0)	$\hat{\phi}_1 : -0,279$	0,000	Lag 6	0,000	0,0124	
			Lag 12	0,004		
			Lag 18	0,038		
			Lag 24	0,049		
			Lag 30	0,023		
			Lag 36	0,003		
			Lag 42	0,005		
ARIMA (2,1,0)	$\hat{\phi}_1 : -0,317$	0,000	Lag 6	0,000	0,0172	
			Lag 12	0,004		
	$\hat{\phi}_2 : -0,132$	0,032	Lag 18	0,036		
			Lag 24	0,052		
				Lag 30		0,023
				Lag 36		0,006
				Lag 42		0,001
ARIMA (3,1,0)	$\hat{\phi}_1 : -0,353$	0,000	Lag 6	0,026	0,0359	
			Lag 12	0,143		
			Lag 18	0,436		
			Lag 24	0,469		
			Lag 30	0,443		
			Lag 36	0,268		
			Lag 42	0,427		
ARIMA (1,1,1)*	$\hat{\theta}_1 : -0,563$	0,000	Lag 6	0,375	0,1473	
			Lag 12	0,671		
			Lag 18	0,924		
	$\hat{\phi}_1 : -0,353$	0,000	Lag 24	0,894		

Model ARIMA Sementara	Estimasi Parameter	P-Value	Independensi Residual (P-Value)	Normalitas Residual (P-Value)
			Lag 30	0,788
			Lag 36	0,471
			Lag 42	0,626

Keterangan:

* Model ARIMA yang signifikan, *white noise*, dan residual berdistribusi Normal

Setelah dilakukan pengujian signifikansi parameter dan pemeriksaan diagnostik, didapatkan untuk model ARIMA terbaik adalah ARIMA (1,1,1).

c. *Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX)*

Pemodelan ARIMAX dilakukan dengan menggunakan informasi hasil pemodelan dengan regresi dummy dan ARIMA. Hasil regresi *dummy* menunjukkan jika

- $D_{1;t}$: satu minggu sebelum Hari Raya Idul Fitri
- $D_{4;t}$: satu minggu sebelum Natal dan Tahun Baru
- $D_{5;t}$: minggu Natal dan Tahun Baru
- $M_{4;t}$: *Dummy* untuk Bulan April
- $M_{5;t}$: *Dummy* untuk Bulan Mei
- $M_{6;t}$: *Dummy* untuk Bulan Juni
- $M_{11;t}$: *Dummy* untuk Bulan November
- $M_{12;t}$: *Dummy* untuk Bulan Desember

adalah variabel *dummy* yang signifikan, dan akan digunakan sebagai informasi dalam pemodelan ARIMAX sebagai variabel eksogen. Pada pemodelan ARIMA, diperoleh model ARIMA yang terbaik yaitu ARIMA (1,1,1). Berdasarkan hasil estimasi parameter, pengujian signifikansi dan melakukan proses eliminasi variabel eksogen yang tidak signifikan, sehingga diperoleh parameter-parameter yang signifikan dan disajikan pada Tabel 5.

Tabel 5 Estimasi Parameter Model ARIMAX

Variabel	Estimasi Parameter	SE	P-Value
t	$\hat{\gamma}_1$: -0,149	0,013	0,000
$D_{1;t}$	$\hat{\beta}_1$: 11,121	2,515	0,000
$D_{4;t}$	$\hat{\beta}_2$: -13,225	3,246	0,000
$D_{5;t}$	$\hat{\beta}_3$: -10,600	3,477	0,002
$M_{5;t}$	$\hat{\beta}_4$: 9,419	2,339	0,000
$M_{12;t}$	$\hat{\beta}_5$: 8,874	2,720	0,001
AR (1)	$\hat{\phi}_1$: 0,615	0,055	0,000
MA (1)	$\hat{\theta}_1$: -1,000	0,013	0,000

Berdasarkan hasil pengujian model ARIMAX dengan efek variasi kalender secara simultan yang ditampilkan pada Tabel 5, terlihat jika semua parameter didalam model telah signifikan dengan taraf signifikansi 5%. Langkah selanjutnya adalah melakukan pemeriksaan diagnostik dari residual model ARIMAX.

1. Independensi Residual

Pengujian independensi residual menggunakan uji Ljung-Box dengan rumusan hipotesis:

Hipotesis

H_0 : $\rho_1 = \rho_2 = \dots = \rho_k = 0$ (*Residual White Noise*)

H_1 : minimal terdapat satu $\rho_k \neq 0$ (*Residual belum White Noise*)

Tabel 6 merupakan hasil pengujian residual dengan menggunakan statistik uji Ljung-Box.

Tabel 6 Hasil Pengujian Residual Model ARIMAX dengan Menggunakan Statistik Uji Ljung-Box

Model ARIMAX dengan Efek Variasi Kalender	Lag	P-Value
	6	0,190
	12	0,519
	18	0,818
	24	0,395
	30	0,068
	36	0,056
	42	0,058

2. Normalitas Residual

Pengujian normalitas residual menggunakan uji Kolmogorov-Smirnov dengan rumusan hipotesis:

Hipotesis

H_0 : residual berdistribusi normal

H_1 : residual tidak berdistribusi normal

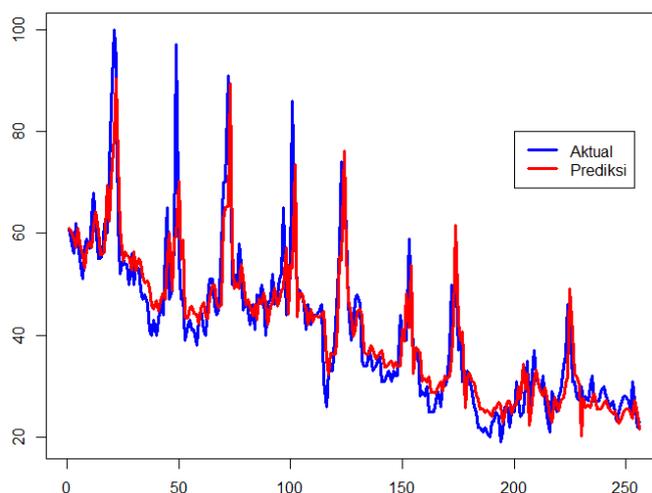
Tabel 7 merupakan hasil pengujian residual dengan menggunakan statistik uji Kolmogorov-Smirnov.

Tabel 7 Pengujian Residual dengan Menggunakan Statistik Uji Kolmogorov-Smirnov

Model ARIMAX dengan Efek Variasi Kalender	P-Value	0,453
	Kesimpulan	Residual berdistribusi normal

d. Ketepatan Akurasi dan Peramalan

Pengukuran akurasi kebaikan pada penelitian ini menggunakan MAPE. Visualisasi grafik perbandingan data aktual dan prediksi menggunakan model ARIMAX dengan efek Variasi Kalender ditampilkan pada Gambar 7.



Gambar 7 Grafik Perbandingan Data Aktual dengan Prediksi dari Model ARIMAX Terbaik

Berdasarkan Gambar 7, diketahui grafik runtun waktu antara data aktual, nilai prediksinya hampir mengikuti pola data sesungguhnya dengan tingkat akurasi peramalan menggunakan MAPE sebesar 10,47%. Model ARIMAX dapat dijabarkan pada Persamaan (9).

$$\hat{Y}_t = -0,149t + 11,121D_{1:t} - 13,225D_{2:t} - 10,600D_{4:t} + 9,419M_{5:t} + 8,874M_{12:t} + 0,615N_{t-1} + a_t + a_{t-1} \quad (9)$$

Model ARIMAX untuk peramalan l periode ke depan dapat dijabarkan pada Persamaan (10).

$$\hat{Y}_{t+l} = -0,149(t+l) + 11,121D_{1:t+l} - 13,225D_{2:t+l} - 10,600D_{4:t+l} + 9,419M_{5:t+l} + 8,874M_{12:t+l} + 0,615N_{t+l-1} + a_{t+l} + a_{t+l-1} \quad (10)$$

Hasil peramalan trend pencarian kata kunci "Zalora" untuk 24 periode kedepan ditampilkan pada Tabel 8.

Tabel 8 Hasil Peramalan Tren Pencarian Kata Kunci "Zalora" untuk 24 periode Kedepan

Minggu	Data trend pencarian kata kunci "Zalora"	Minggu	Data trend pencarian kata kunci "Zalora"	Minggu	Data trend pencarian kata kunci "Zalora"	Minggu	Data trend pencarian kata kunci "Zalora"
04/12/2022	30.970	15/01/2023	20.058	26/02/2023	19.100	09/04/2023	18.201
11/12/2022	30.356	22/01/2023	19.883	05/03/2023	18.949	16/04/2023	29.173
18/12/2022	16.695	29/01/2023	19.718	12/03/2023	18.799	23/04/2023	17.902
25/12/2022	18.996	05/02/2023	19.559	19/03/2023	18.649	30/04/2023	27.172
01/01/2023	20.464	12/02/2023	19.404	26/03/2023	18.499	07/05/2023	27.022
08/01/2023	20.248	19/02/2023	19.251	02/04/2023	18.350	14/15/2023	26.873

Berdasarkan Tabel 8, terlihat jika hasil peramalan Data trend pencarian kata kunci "Zalora" untuk 24 periode kedepan berfluktuasi dan menunjukkan pola yang berulang di beberapa periode tertentu. Diperkirakan di bulan April 2023 akan terjadi peningkatan trend pencarian kata kunci "Zalora" dikarenakan bertepatan dengan Hari Raya Idul Fitri di Tahun 2023.

V. KESIMPULAN DAN SARAN

Aplikasi model ARIMAX dengan efek variasi kalender untuk meramalkan trend pencarian kata kunci "Zalora" pada data Google Trends telah berhasil dilakukan. Berdasarkan hasil analisis data, diperoleh model ARIMAX sebagai berikut:

$$\hat{Y}_t = -0.149t + 11.121D_{1:t} - 13.225D_{2:t} - 10.600D_{4:t} + 9.419M_{5:t} + 8.874M_{12:t} + 0.615N_{t-1} + a_t + a_{t-1}$$

Dengan melihat grafik perbandingan data aktual dan prediksi dari model ARIMAX, dapat diketahui prediksi dari model ARIMAX cenderung mengikuti pola data aktual dengan akurasi kebaikan menggunakan MAPE sebesar 10,47%. Selanjutnya berdasarkan hasil peramalan untuk 24 periode kedepan, diperkirakan di bulan April 2023 akan terjadi peningkatan trend pencarian kata kunci "Zalora" dikarenakan bertepatan dengan Hari Raya Idul Fitri di Tahun 2023. Dampak yang terjadi adalah perlu diperhatikan nya kestabilan antara supply dan demand untuk brand fashion Zalora mendekati bulan April 2023. Saran untuk penelitian selanjutnya yaitu dapat mempertimbangkan menggunakan model ARIMA Subset dan juga mengaplikasikan model Hybrid.

DAFTAR PUSTAKA

- [1] A. T. R. Dani, S. Wahyuningsih, and N. A. Rizki, "Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu," *Jambura Journal of Mathematics*, vol. 1, no. 2, pp. 64–78, 2019.
- [2] A. Harfianto, R. F. Mukhlas, and R. A. Wahyukomala, "Dampak Krisis Global terhadap Industri Hasil Tembakau di Indonesia: Sebuah Analisis Runtun Waktu," *Pajak dan Keuangan Negara*, vol. 4, pp. 326–332, 2022.
- [3] L. A. Nurkhasanah, S. Suparti, and S. Sudarno, "Perbandingan Metode Runtun Waktu Fuzzy-Chen dan Fuzzy-Markov Chain untuk Meramalkan Data Inflasi di Indonesia," *Gaussian*, vol. 4, no. 4, pp. 917–926, 2015.
- [4] M. Arumsari and A. T. R. Dani, "Peramalan Data Runtun Waktu menggunakan Model Hybrid Time Series Regression – Autoregressive Integrated Moving Average," *Jurnal Siger Matematika*, vol. 2, no. 1, pp. 1–12, 2021.
- [5] R. Novidianto and A. T. R. Dani, "Analisis Klaster Kasus Aktif COVID-19 Menurut Provinsi di Indonesia Berdasarkan Data Deret Waktu," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 5, pp. 15–24, 2020.
- [6] A. R. Saputra, S. Wahyuningsih, and M. Siringoringo, "Peramalan Jumlah Titik Panas Provinsi Kalimantan Timur Menggunakan Analisis Intervensi Fungsi Pulse Forecasting Number of Hotspots in East Borneo using Pulse Function of Intervention Analysis," *Jurnal EKSPONENSIAL*, vol. 12, no. 1, 2021.
- [7] Q. S. Akmalia and L. Sucipto, "Metode ARIMA , ARIMAX , dan SARIMA : Sebuah Meta- Analisis Perbedaan Tingkat Akurasi Peramalan Data Time Series," *Jurnal Informatika Kaputama (JIK)*, vol. 6, no. 3, 2022.
- [8] K. Ramadani, S. Wahyuningsih, and M. N. Hayati, "Forecasting Stock Price PT. Telkom Using Hybrid Time Series Regression Linear– Autoregressive Integrated Moving Average Model," *Jurnal Matematika, Statistika dan Komputasi*, vol. 18, no. 2, pp. 293–307, 2022.
- [9] S. W. Rizki, S. Statistika, and Yundari, "Combined Model Time Series Regression – ARIMA on Stocks Prices," *TENSOR: Pure and Applied Mathematics Journal*, vol. 3, no. 2, pp. 65–72, 2022.
- [10] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. California: Holden-Day, 1976.
- [11] Suhartono, *Analisis Data Statistik dengan R*. Surabaya: Lab Statistik Komputasi, ITS, 2008.
- [12] A. Emiro, I. K. Hasan, and N. Achmad, "Perbandingan Model ARIMA-RBF dan ARIMA-GARCH dalam Peramalan Time Series Inflasi Provinsi Gorontalo," *Research in the Mathematical and Natural Sciences*, vol. 2, no. 1, pp. 9–17, 2023.
- [13] C. B. Satrio, W. Darmawan, B. U. Nadia, and N. Hanafiah, "Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET," *Procedia Comput Sci*, vol. 179, no. 2020, pp. 524–532, 2021.
- [14] B. S. Wirdyacharya and M. Prastuti, "Peramalan Permintaan Semen di PT. XYZ Menggunakan Time Series Regression dan ARIMA," *Jurnal Sains dan Seni ITS*, vol. 11, no. 1, 2022.
- [15] R. Yuliyanti and E. Arliani, "Peramalan Jumlah Penduduk Menggunakan Model ARIMA," *Kajian dan Terapan Matematika*, vol. 8, no. 2, pp. 114–128, 2022.
- [16] G. Perone, "Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy," *European Journal of Health Economics*, vol. 23, no. 6, pp. 917–940, 2022.
- [17] D. Kusnandar and N. Imro, "Peramalan Harga Saham Syariah Jakarta Islamic Index Dengan Model Arimax-Garch," *Buletin Ilmiah Math. Stat dan Terapannya (Bimaster)*, vol. 11, no. 2, pp. 263–272, 2022.
- [18] L. Widyarsi and H. Usman, "Penggunaan Data Google Trends untuk Peramalan Tingkat Pengangguran Terbuka di Tingkat Nasional dan Regional di Provinsi Jawa Barat," *Seminar Nasional Official Statistics*, vol. 2021, no. 1, pp. 980–990, 2021.
- [19] S. Hidayat and N. Hakim, "Peramalan Ekspor Luar Negeri Banten Menggunakan Model Arimax," *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 2, no. 2, pp. 204–213, 2021.
- [20] R. Permatasari, "Pemodelan dan Peramalan Runtun Waktu Nonlinear dengan Metode Exponential Smooth Transition Autoregressive (Estar)," *Indonesian Journal of Mathematics and Natural Sciences*, vol. 45, no. 1, pp. 20–29, 2022.
- [21] N. N. Lingga, I. Indwiarti, and A. A. Rohmawati, "Pemodelan dan Peramalan Kedatangan Wisatawan ke Tempat Wisata dengan Google Trends Menggunakan Metode Variasi Kalender ARIMAX," *e-Proceeding of Engineering*, vol. 8, no. 2, pp. 3361–3372, 2021.
- [22] A. T. R. Dani, M. Fauziah, H. Sandariria, and Q. Q. A'yun, "Forecasting The Search Trends of The Keyword 'Sarung Wadimor' In Indonesia on Google Trends Data Using Time Series Regression with Calender Variation and Arima Box-Jenkins," *Jurnal Matematika, Statistika dan Komputasi*, vol. 19, no. 3, pp. 447–459, May 2023.



© 2023 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (<http://creativecommons.org/licenses/by-sa/4.0/>).