

Peramalan Tren Pencarian Kata Kunci "Wardah" Menggunakan ARIMAX dan *Random Forest*

DIAN NURMAWATI, ATIK WINTARTI*, DINDA GALUH GUMINTA
Program Studi Sarjana Sains Data, Universitas Negeri Surabaya, Indonesia

Abstrak

Industri Kecantikan di Indonesia mengalami pertumbuhan yang pesat, menempatkan salah satu produk lokal seperti Wardah sebagai pemimpin pasar yang diminati konsumen. Perubahan minat konsumen secara digital dapat dipantau melalui aktivitas pencarian di Internet seperti *Google Trends*. Penelitian ini bertujuan untuk mengeksplorasi data dan meramalkan tren pencarian kata kunci "Wardah" periode 1 Januari 2021 hingga 28 Februari 2026. Pendekatan yang digunakan adalah metode statistik ARIMAX dan metode *machine learning Random Forest* dengan fitur *time-lag*. Melalui proses *grid search*, diperoleh model terbaik yaitu ARIMAX (0,0,10). Variabel yang signifikan meliputi tren waktu, *dummy* bulan, 7 hari sebelum dan saat Idul Fitri, hari saat Natal, 7 hari setelah Natal, 7 hari setelah Idul Adha, 7 hari sebelum Imlek dan hari saat Imlek, *dummy weekend*, serta *double date*. Model ARIMAX pada data *in-sample* menghasilkan RMSE 6,74; MAE 4,99; dan MAPE 6,59%, sedangkan pada data *out-sample* menghasilkan RMSE 7,71; MAE 6,17; dan MAPE 7,45%. Sementara itu, *Random Forest* dengan *hyperparameter* terbaik pada data *in-sample* menghasilkan RMSE 2,35; MAE 1,74; dan MAPE 2,27%, sedangkan pada data *out-sample* menghasilkan RMSE 6,38; MAE 5,03; dan MAPE 6,16%. Berdasarkan hasil evaluasi, *Random Forest* memberikan performa lebih baik dibandingkan ARIMAX pada data *in-sample* maupun *out-sample*.

Kata Kunci ARIMAX; *Google Trends*; Peramalan; *Random Forest*; Wardah

Abstract

The beauty industry in Indonesia is experiencing rapid growth, positioning a local product like Wardah as a market leader highly demanded by consumers. Changes in digital consumer interest can be monitored through online search activities such as Google Trends. This study aims to explore data and forecast the search trend for the keyword "Wardah" from January 1, 2021, to February 28, 2026. The approaches used are the statistical method ARIMAX and the machine learning method Random Forest utilizing time-lag features. Through the grid search process, the best-selected model obtained is ARIMAX (0,0,10). Significant variables include the time trend, monthly dummy, 7 days before and during Eid al-Fitr, the day of Christmas, 7 days after Christmas, 7 days after Eid al-Adha, 7 days before Chinese New Year, and the day of Chinese New Year, weekend dummy, and double date. The ARIMAX model on in-sample data produces RMSE 6,74; MAE 4,99; and MAPE 6,59%, while on out-sample data it produces RMSE 7,71; MAE 6,17; and MAPE 7,45%. Meanwhile, Random Forest model with the best hyperparameters on in-sample data produces RMSE 2,35; MAE 1,74 and MAPE 2,27%, whereas on out-sample data it produces RMSE 6,38; MAE 5,03 and MAPE 6,16%. Based on the evaluation results,

*Corresponding author. Email: atikwintarti@unesa.ac.id.

Random Forest delivers a better performance compared to ARIMAX on both in-sample and out-sample data.

Keywords *ARIMAX; Forecasting; Google Trends; Random Forest; Wardah*

1 Pendahuluan

Industri kecantikan di Indonesia mengalami pertumbuhan yang pesat dalam beberapa tahun terakhir. Nilai belanja produk kecantikan terus meningkat dari tahun ke tahun. Tingkat pertumbuhan belanja kecantikan di Indonesia lebih tinggi dibandingkan negara-negara di kawasan Asia Pasifik yang tumbuh 6,74% dan Amerika Serikat 5,8% (Santika, 2024). Kondisi ini sejalan dengan meningkatnya permintaan masyarakat terhadap produk perawatan diri, termasuk *skincare* dan kosmetik, sehingga industri kecantikan menjadi salah satu sektor yang berperan penting dalam perekonomian Indonesia (Subekti and Rahayu, 2025).

Salah satu merek kosmetik lokal yang memiliki popularitas tinggi adalah Wardah. Berdasarkan data *e-commerce*, Wardah menempati posisi sebagai salah satu merek *skincare* dan kecantikan terlaris di Shopee Indonesia dengan pangsa pasar yang cukup besar (Compas, 2025) (GoodStats, 2025a) (GoodStats, 2025b). Keberhasilan tersebut menunjukkan tingginya minat dan kepercayaan masyarakat terhadap produk Wardah. Dalam perilaku digital, *Google Trends* dapat digunakan untuk memahami pola pencarian informasi masyarakat di internet sekaligus memantau minat konsumen terhadap suatu merek dari waktu ke waktu. Data *Google Trends* berasal dari kata kunci yang dicari pengguna Google, dan dengan tingkat pengguna Google di Indonesia mencapai 98,23%, data yang dihasilkan dinilai cukup representatif dalam menggambarkan perilaku pencarian masyarakat (Prasojo, 2022).

Pada data deret waktu, perubahan nilai tidak hanya dipengaruhi oleh pola historis, tetapi juga faktor eksternal tertentu. Oleh karena itu, model ARIMAX digunakan karena mampu mengakomodasi variabel eksogen, termasuk variasi kalender yang mencerminkan pola berulang akibat perbedaan penanggalan setiap tahun (Dani et al., 2023b). Beberapa penelitian sebelumnya telah memanfaatkan data *Google Trends*. Penelitian pencarian kata kunci "Zalora" menerapkan ARIMAX dengan variasi kalender (Dani et al., 2023b). Penelitian lain pada tren pencarian "Sarung Wadimor" juga memanfaatkan variasi kalender berupa Ramadan dan bulan (Dani et al., 2023a). Penerapan ARIMAX juga digunakan pada peramalan tingkat hunian hotel di Provinsi Riau (Rizalde et al., 2021).

Di sisi lain, pendekatan *machine learning*, khususnya *Random Forest*, juga banyak digunakan pada peramalan data deret waktu. Pada penelitian prediksi harga saham (Urrochman et al., 2025). Pada penelitian prediksi dan analisis kemiskinan (Handayani and Qutub, 2025). Selain itu, penelitian dengan memanfaatkan data *google trends* pada prediksi okupansi hotel (Rahmawati et al., 2024). Penelitian sebelumnya juga menunjukkan bahwa peristiwa *double date* (tanggal kembar) berpengaruh signifikan terhadap perilaku konsumen, khususnya dalam pembelian produk kosmetik dan aktivitas belanja *online* (Leilani et al., 2023). Selain itu, pada penelitian (Ferdian et al., 2025) dan (Oktaviansyah and Wulansari, 2025) menunjukkan bahwa *weekend* (hari libur) memberikan dampak signifikan terhadap perilaku pembelian konsumen, khususnya dalam peningkatan permintaan.

Berdasarkan uraian, maka penelitian ini menggunakan ARIMAX dengan variabel ekso-gen berupa variasi kalender yang meliputi hari raya Idul Fitri, Idul Adha, Imlek, Natal, bulan, *double date*, dan *weekend*. Selain itu, digunakan model *Random Forest* dengan fitur *time-lag* untuk menangkap pola historis data pencarian. Serta bertujuan untuk membandingkan performa

model ARIMAX dan *Random Forest* dalam meramalkan indeks pencarian kata kunci "Wardah" berdasarkan data *Google Trends*. Evaluasi performa model menggunakan nilai RMSE, MAE, dan MAPE.

2 Metode

2.1 Dataset

Dataset penelitian ini diperoleh dari *Google Trends* (<https://trends.google.co.id/trends/>) menggunakan istilah penelusuran kata kunci "Wardah" pada wilayah Indonesia. Data mencakup periode 1 Januari 2021 hingga 28 Februari 2026 sebanyak 1.885 data harian dengan dua variabel, yaitu Tanggal dan Wardah. Variabel Wardah berupa indeks pencarian dengan skala 0 - 100. Nilai 100 menunjukkan tingkat popularitas tertinggi dari pencarian kata kunci tersebut dalam periode tertentu dan nilai mendekati 0 menunjukkan minat pencarian yang rendah (Zeller, 2023).

Table 1: Data *Google Trends* kata kunci "Wardah"

Tanggal	Wardah
01/01/2021	48
02/01/2021	41
03/01/2021	46
04/01/2021	42
⋮	⋮
28/02/2026	92

2.2 Alur Penelitian

Penelitian ini menggambarkan tahapan yang ditempuh untuk mencapai tujuan penelitian, mulai dari pengambilan data *Google Trends*, Tahap pra-pemrosesan meliputi konversi format tanggal ke *datetime*, pengecekan *missing value*, pengurutan data berdasarkan waktu, serta memastikan variabel numerik. Analisis eksploratif dilakukan melalui grafik deret waktu untuk memahami pola data. Selanjutnya, dibentuk variabel *dummy* variasi kalender sebagai variabel eksogen pada ARIMAX. Data kemudian dibagi menjadi *in-sample* (1 Januari 2021 hingga 31 Desember 2025) dan *out-sample* (1 Januari 2026 hingga 28 Februari 2026) berdasarkan urutan waktu untuk menjaga struktur deret waktu, tahap pemodelan menggunakan ARIMAX dan *Random Forest*, hingga evaluasi model.

• Tahapan pemodelan ARIMAX

1. Persiapan data dilakukan dengan menggunakan dataset hasil pra-pemrosesan yang telah mencakup variabel *dummy* kalender. Variabel *dummy* didefinisikan sebagai berikut:

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

Variabel *dummy* hari besar keagamaan dirinci sebagai berikut:

$D_{1,t}$: 7 hari sebelum Hari Raya Idul Fitri
 $D_{2,t}$: hari saat Hari Raya Idul Fitri
 $D_{3,t}$: 7 hari setelah Hari Raya Idul Fitri
 $D_{4,t}$: 7 hari sebelum Natal
 $D_{5,t}$: hari saat Natal
 $D_{6,t}$: 7 hari setelah Natal
 $D_{7,t}$: 7 hari sebelum Idul Adha
 $D_{8,t}$: hari saat Idul Adha
 $D_{9,t}$: 7 hari setelah Idul Adha
 $D_{10,t}$: 7 hari sebelum Imlek
 $D_{11,t}$: hari saat Imlek
 $D_{12,t}$: 7 hari setelah Imlek
Dummy bulan didefinisikan sebagai:

$$M_{j,t} = \begin{cases} 1, & \text{bulan } j \\ 0, & \text{bulan lainnya} \end{cases} \quad j = 1, 2, \dots, 12$$

Variabel *dummy* bulan dirinci sebagai berikut:

$M_{1,t}$: bernilai 1 untuk bulan Januari, 0 untuk bulan lainnya
 $M_{2,t}$: bernilai 1 untuk bulan Februari, 0 untuk bulan lainnya
 $\dots$
 $M_{12,t}$: bernilai 1 untuk bulan Desember, 0 untuk bulan lainnya

Selain itu, digunakan dua variabel *dummy* tambahan untuk menangkap pola perilaku konsumen di luar hari besar keagamaan dan bulan, yaitu:

Dummy Double Date

$$DoubleDate_t = \begin{cases} 1, & \text{jika tanggal kembar (1.1, 2.2, \dots, 12.12)} \\ 0, & \text{lainnya} \end{cases}$$

Dummy Weekend

$$Weekend_t = \begin{cases} 1, & \text{Sabtu atau Minggu} \\ 0, & \text{hari lainnya} \end{cases}$$

2. Stasioneritas varians dan rata-rata diuji menggunakan *transformasi Box-Cox* dan uji ADF sebelum serta sesudah *differencing* hingga data stasioner (p-value < 0,05).
3. Identifikasi orde ARIMAX dilakukan melalui analisis ACF dan PACF, kemudian kombinasi (p,d,q) diuji menggunakan *grid search* dengan variabel eksogen (*dummy* kalender). Model terbaik dipilih berdasarkan AIC terkecil dan parameternya signifikan.
4. Uji signifikansi parameter dilakukan dengan mengeliminasi variabel eksogen yang tidak signifikan (p-value > 0,05) menggunakan metode *backward elimination* hingga seluruh variabel signifikan.
5. Analisis diagnostik model dilakukan pada residual untuk memastikan terpenuhinya asumsi *white noise* dan normalitas.
6. Peramalan tren pada data *in-sample* dan *out-sample* menggunakan model ARIMAX terbaik dan dilakukan *inverse transformasi Box-Cox* untuk mengembalikan hasil peramalan ke skala data asli.
7. Evaluasi model dilakukan menggunakan RMSE, MAE, dan MAPE pada data *in-sample* dan *out-sample*.

● Tahapan pemodelan *Random Forest*

1. Pemodelan menggunakan dataset hasil pra-pemrosesan pada tahap ARIMAX, termasuk variabel *dummy* kalender yang telah dibuat.
2. Pembuatan fitur *time-lag* untuk menangkap pola historis indeks pencarian kata kunci "Wardah" berdasarkan hasil ACF dan PACF sebelumnya. Selain itu, variabel tren (t) dan variabel *dummy* kalender juga digunakan sebagai prediktor. Setelah pembentukan fitur, data dengan nilai hilang (NaN) akibat lag dihapus sehingga diperoleh dataset akhir untuk pemodelan.
3. Pembentukan model *Random Forest* dan *hyperparameter tuning* dioptimasi menggunakan *GridSearchCV* dengan 5-fold cross-validation untuk memperoleh kombinasi parameter terbaik. Parameter yang diuji meliputi `n_estimators`, `max_depth`, `min_samples_leaf`, `min_samples_split`, `max_features`, dan `bootstrap`. Model dengan performa terbaik dipilih untuk tahap peramalan.
4. Peramalan pada data *in-sample* dan *out-sample* dilakukan menggunakan model *Random Forest* dengan parameter terbaik yang diperoleh dari proses *hyperparameter tuning*.
5. Analisis *feature importance* dilakukan untuk mengukur kontribusi masing-masing variabel terhadap hasil peramalan. Nilai *importance* yang lebih tinggi menunjukkan pengaruh yang lebih besar terhadap model.
6. Evaluasi model dilakukan menggunakan RMSE, MAE, dan MAPE pada data *in-sample* dan *out-sample*.

2.3 ARIMAX

ARIMAX (*Autoregressive Integrated Moving Average with eXogenous variables*) merupakan perluasan dari model ARIMA yang memasukkan variabel eksternal (exogenous variables) ke dalam proses peramalan. Variabel eksogen tersebut bukan bagian dari deret waktu utama, namun dianggap memberikan pengaruh signifikan terhadap nilai variabel dependen sehingga dapat meningkatkan akurasi peramalan (Zaghwan et al., 2026). Sehingga, bentuk umum ARIMAX dapat dituliskan pada persamaan berikut (Rizalde et al., 2021):

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)a_t + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \dots + \beta_p X_{p,t} \quad (1)$$

dengan:

Y_t	=	nilai variabel respon (dependen) pada waktu ke- t (misalnya indeks pencarian Wardah)
$(1-B)^d Y_t$	=	proses <i>differencing</i> orde d
$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$	=	parameter <i>Autoregressive</i> (AR)
$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$	=	parameter <i>Moving Average</i> (MA)
a_t	=	residual (<i>white noise</i>) pada waktu ke- t
$X_{i,t}$	=	variabel eksogen pada waktu ke- t
β_i	=	koefisien variabel eksogen
p	=	orde komponen <i>autoregressive</i> (AR)
q	=	orde komponen <i>moving average</i> (MA)

2.4 Random Forest

Random Forest merupakan salah satu algoritma pembelajaran mesin yang digunakan untuk pemodelan regresi. Algoritma ini termasuk ke dalam metode *ensemble learning* yang bekerja dengan menggabungkan banyak model regresi berbasis *decision tree* untuk meningkatkan akurasi

prediksi dan mengurangi risiko *overfitting* (Yin et al., 2023). Prediksi yang dihasilkan oleh setiap pohon dapat dinyatakan pada persamaan berikut:

$$\hat{y}_b(x) = T_b(x) \quad (2)$$

Prediksi akhir *Random Forest* diperoleh dengan menghitung rata-rata prediksi dari seluruh B pohon, yang dirumuskan sebagai (Du, 2026):

$$\hat{y}(x) = \frac{1}{B} \sum_{b=1}^B \hat{y}_b(x) \quad (3)$$

dengan:

x	=	data input atau fitur yang ingin diprediksi nilainya.
$\hat{y}(x)$	=	nilai prediksi akhir dari model <i>Random Forest</i> untuk input x .
B	=	jumlah pohon keputusan (<i>decision trees</i>) yang dibangun dalam <i>Random Forest</i> .
$b = 1, 2, \dots, B$	=	indeks yang menunjukkan urutan setiap pohon di dalam hutan.
$\hat{y}_b(x)$	=	prediksi yang dihasilkan oleh pohon ke- b untuk input x .
$T_b(x)$	=	fungsi prediksi yang dihasilkan oleh pohon keputusan ke- b untuk input x .

2.5 Akaike's Information Criterion (AIC)

AIC (*Akaike's Information Criterion*) merupakan suatu ukuran yang mengintergrasikan tingkat kesesuaian model dengan nilai likelihood serta jumlah parameter yang ada dalam model tersebut. Model yang dinilai paling baik adalah dengan nilai AIC paling rendah. Nilai AIC dapat dihitung melalui Persamaan berikut (Shumway and Stoffer, 2011):

$$AIC = \log \hat{\sigma}_k^2 + \frac{n + 2k}{n} \quad (4)$$

dengan:

$\hat{\sigma}_k^2$	=	estimasi varians error dari model dengan k parameter. Semakin kecil nilai ini, semakin baik model dalam menjelaskan data.
k	=	jumlah parameter dalam model.
n	=	jumlah total observasi dalam deret waktu.
$\log \hat{\sigma}_k^2$	=	komponen pertama AIC yang mengukur ketepatan model; nilai varians error yang kecil akan menghasilkan nilai log yang lebih kecil.
$\frac{n+2k}{n}$	=	komponen penalti yang membatasi kompleksitas model; semakin banyak parameter (k), semakin besar penalti yang diberikan.

2.6 Uji Stasioneritas

Transformasi diterapkan saat varians data tidak bersifat konstan terhadap waktu, yang dikenal sebagai transformasi Box-Cox. Nilai parameter transformasi λ menentukan jenis transfor-

masi yang digunakan. Apabila $\lambda = 1$, data telah berada dalam kondisi stasioner dalam varians (Shumway and Stoffer, 2011). Transformasi Box-Cox dapat dituliskan sebagai berikut (Wei, 2006):

$$Y_t^{(\lambda)} = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln(Y_t), & \lambda = 0 \end{cases} \quad (5)$$

dengan:

Y_t	=	nilai deret waktu pada waktu ke- t sebelum transformasi Box-Cox.
$Y_t^{(\lambda)}$	=	nilai deret waktu pada waktu ke- t setelah transformasi Box-Cox.
λ	=	parameter transformasi Box-Cox yang menentukan bentuk transformasi.
$\lambda \neq 0$	=	kondisi ketika transformasi menggunakan bentuk pangkat (power transformation).
$\lambda = 0$	=	kondisi ketika transformasi menggunakan logaritma natural.
$\frac{Y_t^\lambda - 1}{\lambda}$	=	bentuk transformasi ketika $\lambda \neq 0$, digunakan untuk menstabilkan varians dan mendekati distribusi normal.
$\ln(Y_t)$	=	bentuk limit transformasi ketika $\lambda = 0$.

Beberapa nilai λ beserta bentuk transformasinya disajikan pada Tabel 2 (Risma and Sahri-man, 2020).

Table 2: Nilai λ dan Model Transformasi

Nilai λ	Model Transformasi
2	Y_t^2
1	Y_t
0,5	$\sqrt{Y_t}$
0	$\ln(Y_t)$
-0,5	$\frac{1}{\sqrt{Y_t}}$
-1	$\frac{1}{Y_t}$
-2	$\frac{1}{Y_t^2}$

Setelah proses pemodelan dan peramalan selesai, hasil peramalan dikembalikan ke skala data asli menggunakan transformasi balik (*inverse Box-Cox*) agar mudah diinterpretasikan. Transformasi balik *Box-Cox* dinyatakan pada Persamaan (6) (Hyndman and Athanasopoulos, 2021).

$$Y_t = \begin{cases} (\lambda Y_t^{(\lambda)} + 1)^{1/\lambda}, & \lambda \neq 0, \\ \exp(Y_t^{(\lambda)}), & \lambda = 0. \end{cases} \quad (6)$$

dengan:

Y_t	=	nilai data pada waktu ke- t pada skala asli.
$Y_t^{(\lambda)}$	=	nilai data pada waktu ke- t setelah transformasi <i>Box-Cox</i> .
λ	=	parameter transformasi <i>Box-Cox</i> .

Selain itu juga dilakukan evaluasi stasioner pada rata-rata menggunakan uji ADF (*Augmented Dicky Fuller*), Hipotesis dari uji ADF adalah sebagai berikut:

$$\begin{aligned} H_0 : \delta &= 0 && \text{data tidak stasioner (ada unit root)} \\ H_1 : \delta &< 0 && \text{data stasioner (tidak ada unit root)} \end{aligned}$$

Penolakan terhadap hipotesis nol (H_0) terjadi jika p -value lebih kecil dari α pada tingkat signifikansi yang ditetapkan, yaitu 0,05 (Gujarati, 2006).

2.7 Analisis Diagnostik

Analisis diagnostik dilakukan bahwa model telah memenuhi asumsi *white noise* (independensi residual) menggunakan statistik uji Ljung-Box, dengan hipotesis (Wei, 2006):

$$\begin{aligned} H_0 : \rho_1 &= \rho_2 = \dots = \rho_K = 0 && \text{(residual } white \text{ noise, tidak terdapat autokorelasi)} \\ H_1 : &\text{terdapat minimal satu } \rho_k \neq 0, k = 1, 2, \dots, K && \text{(residual tidak } white \text{ noise, terdapat autokorelasi)} \end{aligned}$$

Adapun persamaan uji Ljung-Box dapat dituliskan sebagai berikut:

$$Q = n(n+2) \sum_{k=1}^K \frac{\hat{\rho}_k^2}{n-k} \quad (7)$$

dengan:

$$\begin{aligned} Q &= \text{nilai statistik uji Ljung-Box.} \\ n &= \text{jumlah data observasi (pengamatan).} \\ \hat{\rho}_k &= \text{estimasi autokorelasi residual pada lag ke-}k. \\ K &= \text{jumlah lag maksimum yang digunakan dalam pengujian.} \end{aligned}$$

Kriteria pengambilan keputusan adalah H_0 diterima jika nilai p -value $> \alpha$, dengan tingkat signifikansi yang digunakan adalah 0,05. Untuk menguji normalitas residual dalam model, salah satunya dengan menggunakan uji Jarque-Bera, dengan rumusan hipotesis:

$$\begin{aligned} H_0 : &\text{residual berdistribusi normal} \\ H_1 : &\text{residual tidak berdistribusi normal} \end{aligned}$$

Kesimpulan kriteria pengujian H_0 diterima jika nilai p -value $> \alpha$, dengan tingkat signifikansi $\alpha = 0,05$.

2.8 Evaluasi Model

Dalam menentukan hasil peramalan terbaik dari model ARIMAX dan *Random Forest*, dilakukan dengan membandingkan nilai *Root Mean Squared Error* (RMSE), *Mean Absolut Error* (MAE) dan *Mean Absolute Percentage Error* (MAPE). Model terbaik adalah yang memiliki nilai RMSE, MAE dan MAPE paling kecil. RMSE merupakan akar dari rata-rata kuadrat selisih antara nilai aktual dengan dengan nilai prediksi, kemudian dibagi dengan jumlah data. MAE adalah perhitungan rata-rata kesalahan absolut dalam prediksi, semakin kecil nilai MAE maka semakin baik model prediksi (Chai and Draxler, 2014). MAPE menunjukkan persentase kesalahan dari

hasil peramalan (Kim and Kim, 2016). Nilai MAPE <10% menunjukkan bahwa model tersebut cocok digunakan sebagai model peramalan. RMSE dituliskan pada persamaan (8).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (8)$$

MAE dituliskan pada persamaan (9).

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (9)$$

MAPE dituliskan pada persamaan (10).

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100 \quad (10)$$

dengan:

Y_t	=	data aktual pada waktu ke- t .
$\hat{Y}_t$	=	data prediksi pada waktu ke- t .
n	=	jumlah observasi.

3 Hasil dan Pembahasan

3.1 Peramalan Tren Pencarian Kata Kunci “Wardah” Menggunakan ARI-MAX

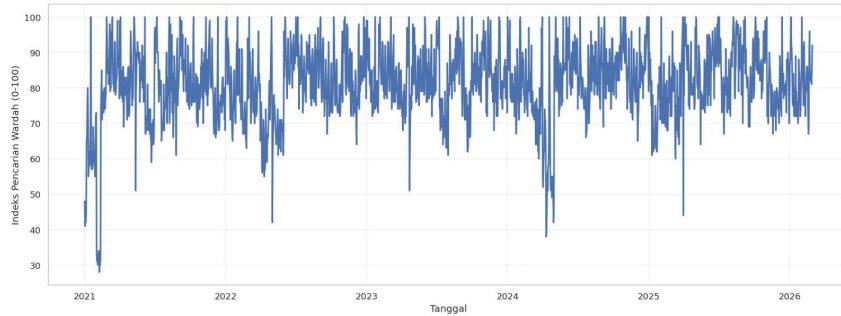


Figure 1: Pola Runtun Waktu Indeks Pencarian Kata Kunci "Wardah"

Berdasarkan gambar 1, indeks pencarian kata kunci “Wardah” menunjukkan fluktuasi yang cukup dinamis selama periode pengamatan. Meskipun terdapat fluktuasi harian, secara umum nilainya cenderung tinggi, yaitu pada kisaran indeks 70 hingga 100, yang mengindikasikan tingginya dan stabilnya minat masyarakat terhadap merek Wardah. Pada beberapa hari di tahun 2021, 2022, 2024, dan 2025 terjadi penurunan pencarian hingga berada pada rentang nilai indeks 30 hingga 50. Namun, tidak berlangsung lama karena nilai indeks kembali ke rentang yang tinggi. Secara keseluruhan, meskipun nilai indeks pencarian mengalami variasi jangka pendek, minat pencarian masyarakat terhadap "Wardah" tetap stabil dan tidak menunjukkan penurunan signifikan sepanjang tahun 2021 hingga 2026.

Uji Stasioneritas

Berdasarkan hasil pengujian pada data asli, diperoleh nilai estimasi λ sebesar 2,49 yang menunjukkan bahwa data belum stasioner dalam varians sehingga diperlukan transformasi Box-Cox. Karena nilai λ mendekati 2, maka transformasi yang digunakan adalah transformasi pangkat dua, yaitu $Y^{(\lambda)} = Y^2$ (Risma and Sahriman, 2020). Setelah transformasi dilakukan, nilai estimasi λ bergeser mendekati 1,00 yang mengindikasikan bahwa varians data telah stabil. Selanjutnya, hasil uji Augmented Dickey-Fuller (ADF) menunjukkan nilai p -value sebesar $0,000 < \alpha$ (0,05), sehingga H_0 ditolak dan data dinyatakan stasioner dalam rata-rata. Dengan demikian, data hasil transformasi Box-Cox telah memenuhi asumsi stasioneritas dalam varians maupun rata-rata, sehingga tidak diperlukan proses differencing ($d = 0$) pada tahap pemodelan ARIMAX.

Pembentukan Model ARIMAX

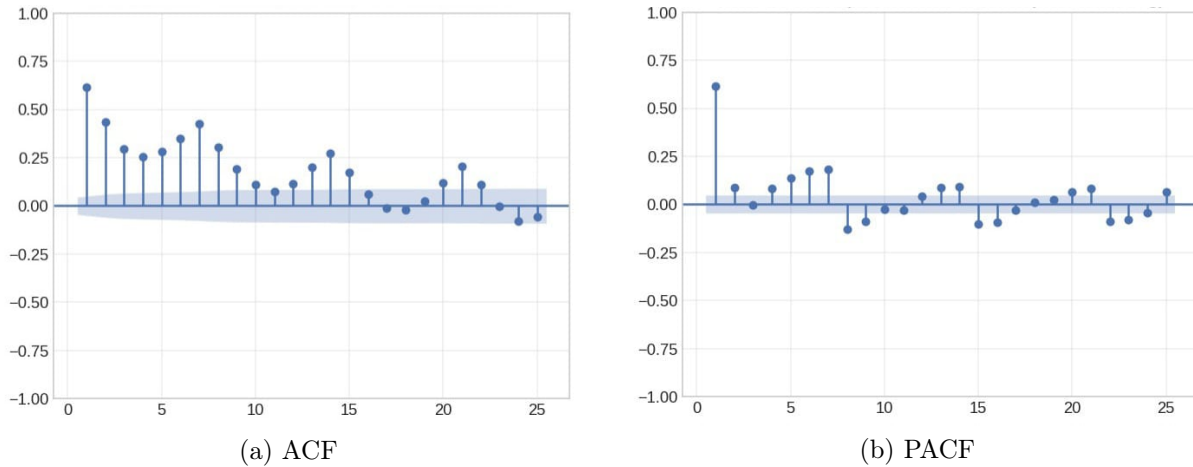


Figure 2: Grafik ACF dan PACF

Pada plot ACF tidak terjadi *cut off* secara tiba-tiba, melainkan menunjukkan pola penurunan bertahap (*decay*) pada beberapa lag awal. Lag 1 memiliki autokorelasi tertinggi, diikuti lag 2 hingga 5 yang masih berada di atas batas signifikansi. Pada lag yang lebih jauh (lag 6 hingga 15 serta lag 21) masih terlihat *spike*, yang mengindikasikan adanya memori jangka panjang pada data. Sementara itu, plot PACF menunjukkan spike signifikan pada lag 1 dan lag 2. Pada lag 3 nilai PACF sempat berada dalam batas kepercayaan, namun tidak menunjukkan *cut off* yang jelas karena kembali muncul spike pada lag 4 hingga 7. Setelah lag 7, sebagian besar nilai berada dalam batas kepercayaan meskipun terdapat beberapa spike kecil pada lag 21. Pola ini mengindikasikan kemungkinan model AR orde rendah atau kombinasi AR dan MA, sehingga model yang sesuai adalah ARIMA ($p, 0, q$). Plot ACF dan PACF digunakan untuk menentukan orde model ARIMAX yang terbentuk.

Dengan mempertimbangkan pola tersebut, pencarian orde AR (p) dan MA (q) dilakukan pada rentang 0–14, dengan $d = 0$ sesuai hasil uji stasioneritas. Rentang ini dipilih agar model dapat menangkap ketergantungan data hingga dua siklus minggu, mengingat pada plot ACF masih terlihat spike signifikan pada lag menengah, sekaligus menjaga efisiensi komputasi pada proses *grid search*. Setiap kombinasi model dievaluasi berdasarkan nilai AIC terkecil serta signifikansi parameter AR dan MA pada taraf 5%. Berdasarkan hasil *grid search*, diperoleh model terbaik yaitu ARIMAX (0,0,10) yang memiliki nilai AIC paling rendah dan seluruh parameter

ARIMAX signifikan pada taraf 5%.

Signifikansi Parameter

Table 3: Estimasi Parameter Model ARIMAX (Awal)

Parameter	Coef	Std. Err	Z-value	P-value
t	1.4515	0.670	2.168	0.030
D_1	5209.733	1085.751	4.798	0.000
D_2	-1.512×10^4	1485.488	-10.182	0.000
D_3	-362.9688	880.564	-0.412	0.680
D_4	1989.9404	1416.615	1.405	0.160
D_5	7122.7177	2037.016	3.497	0.000
D_6	3468.7840	1053.003	3.294	0.001
D_7	1177.6244	2069.836	0.569	0.569
D_8	-3194.7069	2394.690	-1.334	0.182
D_9	3175.5837	1824.009	1.741	0.082
D_{10}	-2471.2490	1435.084	-1.722	0.085
D_{11}	4589.3440	1673.637	2.742	0.006
D_{12}	675.2942	1392.184	0.485	0.628
M_1	1.767×10^4	1014.511	17.418	0.000
M_2	1.819×10^4	1046.097	17.391	0.000
M_3	1.928×10^4	1153.853	17.079	0.000
M_4	1.579×10^4	975.274	16.188	0.000
M_5	1.978×10^4	988.480	20.014	0.000
M_6	2.184×10^4	994.761	21.952	0.000
M_7	1.928×10^4	988.033	19.512	0.000
M_8	2.315×10^4	1164.761	19.884	0.000
M_9	2.074×10^4	1231.674	16.841	0.000
M_{10}	2.194×10^4	1253.046	17.507	0.000
M_{11}	2.083×10^4	1174.756	17.731	0.000
M_{12}	1.828×10^4	1357.828	13.464	0.000
Weekend	5545.0525	252.770	21.937	0.000
Double Date	8252.0970	452.820	18.224	0.000
MA(1)	0.4239	0.021	20.180	0.000
MA(2)	0.3941	0.024	16.274	0.000
MA(3)	0.2781	0.027	10.286	0.000
MA(4)	0.2008	0.028	7.118	0.000
MA(5)	0.1817	0.027	6.636	0.000
MA(6)	0.1344	0.029	4.609	0.000
MA(7)	0.2003	0.029	6.969	0.000
MA(8)	0.1346	0.028	4.886	0.000
MA(9)	0.1167	0.028	4.099	0.000
MA(10)	0.1006	0.025	3.979	0.000
σ^2	2.018×10^7	3.330	6.06×10^6	0.000

Hasil estimasi awal pada Tabel 3 menunjukkan bahwa beberapa variabel *dummy* meliputi D_3 (7 hari setelah idul fitri), D_4 (7 hari sebelum natal), D_7 (7 hari sebelum idul adha), D_8 (hari saat idul adha), D_9 (7 hari setelah idul adha), D_{10} (7 hari sebelum imlek) dan D_{12} (7 hari setelah imlek) memiliki nilai p -value $> 0,05$ sehingga tidak signifikan secara statistik. Hal ini mengindikasikan bahwa keberadaan momen-momen tersebut belum memberikan pengaruh perbedaan volume indeks pencarian kata kunci "Wardah" yang signifikan dibandingkan dengan hari-hari lainnya. Ketidaksignifikanan tersebut menunjukkan bahwa perubahan indeks pencarian pada periode tersebut relatif tidak konsisten atau belum cukup kuat secara statistik untuk dibedakan dari variasi pencarian pada hari-hari lainnya. Selain itu, perubahan indeks pencarian pada momen tersebut dapat dipengaruhi oleh variabel eksogen lain yang terdapat dalam model, seperti variabel *dummy* hari besar keagamaan lainnya, *dummy Double Date*, *dummy Weekend*, dan komponen deret waktu. Oleh karena itu, seluruh variabel yang tidak signifikan ini dievaluasi

dan dikeluarkan menggunakan metode *backward elimination*.

Proses eliminasi variabel menggunakan metode *backward elimination* dengan menghapus variabel-variabel tidak signifikan secara bertahap berdasarkan nilai p -value terbesar. Hal ini dilakukan untuk menjaga efisiensi dan keakuratan model. Setelah variabel yang tidak signifikan dikeluarkan, diperoleh model akhir di mana seluruh parameter eksogen yang tersisa dan komponen *Moving Average* (MA) memiliki nilai p -value $< 0,05$ (signifikan), sebagaimana ditunjukkan pada Tabel 4. Hal ini mengindikasikan bahwa tren waktu, variasi musiman bulanan, serta momen kalender tertentu berpengaruh signifikan terhadap fluktuasi indeks pencarian kata kunci "Wardah".

Table 4: Estimasi Parameter Model ARIMAX (Akhir)

Parameter	Coef	Std. Err	Z-value	P-value
t	1.4335	0.671	2.136	0.033
D_1	5246.4175	986.562	5.318	0.000
D_2	-1.51×10^4	1282.811	-11.768	0.000
D_5	6555.5779	1707.531	3.839	0.000
D_6	2968.1923	975.577	3.043	0.002
D_9	3089.7288	1441.089	2.144	0.032
D_{10}	-2567.9834	1045.753	-2.456	0.014
D_{11}	4478.0697	1108.205	4.041	0.000
M_1	1.774×10^4	975.335	18.185	0.000
M_2	1.833×10^4	968.295	18.929	0.000
M_3	1.970×10^4	1155.481	17.052	0.000
M_4	1.573×10^4	967.759	16.255	0.000
M_5	1.976×10^4	986.855	20.026	0.000
M_6	2.193×10^4	989.794	22.156	0.000
M_7	1.935×10^4	979.971	19.750	0.000
M_8	2.315×10^4	1168.357	19.814	0.000
M_9	2.074×10^4	1235.746	16.785	0.000
M_{10}	2.194×10^4	1256.814	17.454	0.000
M_{11}	2.083×10^4	1178.455	17.674	0.000
M_{12}	1.885×10^4	1308.294	14.405	0.000
Weekend	5549.7108	253.071	21.930	0.000
Double Date	8188.2223	455.239	17.987	0.000
MA(1)	0.4256	0.021	20.174	0.000
MA(2)	0.3950	0.024	16.338	0.000
MA(3)	0.2793	0.027	10.345	0.000
MA(4)	0.2021	0.028	7.191	0.000
MA(5)	0.1815	0.027	6.749	0.000
MA(6)	0.1331	0.029	4.628	0.000
MA(7)	0.1971	0.028	7.134	0.000
MA(8)	0.1301	0.027	4.779	0.000
MA(9)	0.1132	0.028	4.025	0.000
MA(10)	0.1016	0.025	4.107	0.000
σ^2	39.9479	2.467	8.22×10^6	0.000

Analisis Diagnostik Model ARIMAX

Table 5: Hasil Uji Ljung-Box Residual ARIMAX

Uji	P-Value (Prob Q)	Kesimpulan
Ljung-Box	0.93	White Noise

Model ARIMAX yang telah memenuhi asumsi signifikansi parameter, maka langkah selanjutnya adalah melakukan pemeriksaan diagnostik residual model ARIMAX menggunakan

independensi residual (*white noise*) dan normalitas residual. Berdasarkan Tabel 5, diperoleh nilai *p-value* sebesar $0,93 > 0,05$, sehingga gagal menolak H_0 . Dengan demikian, residual model ARIMAX telah bersifat *white noise*.

Table 6: Hasil Uji Normalitas Residual ARIMAX

Uji	P-Value	Kesimpulan
Jarque-Bera	0.00	Residual tidak berdistribusi normal

Pada Tabel 6, diperoleh nilai *p-value* sebesar $0,00 < 0,05$, sehingga H_0 ditolak. Hal ini menunjukkan bahwa residual model tidak berdistribusi normal. Meskipun demikian, tidak mengugurkan kelayakan model untuk peramalan. Sebagaimana dinyatakan oleh Hyndman & Athanasopoulos (2021) dalam penelitian (Septiansyah et al., 2025), meskipun normalitas tidak terpenuhi, model tetap layak digunakan untuk peramalan. Hal ini karena asumsi utama ARIMAX adalah residual yang bersifat *white noise* (acak dan tidak berkorelasi). Sebelum dilakukan evaluasi model, hasil peramalan ARIMAX yang masih berada pada skala transformasi Box-cox terlebih dahulu dikembalikan ke skala data asli menggunakan transformasi inverse Box-cox. Oleh karena itu, hasil peramalan perlu dikembalikan ke skala indeks pencarian *Google Trends* agar dapat dibandingkan secara langsung dengan data aktual.

3.2 Peramalan Tren Pencarian Kata Kunci “Wardah” Menggunakan Random Forest

Pada tahap ini dilakukan pemodelan peramalan tren pencarian kata kunci “Wardah” menggunakan metode *Random Forest*. Model dibangun menggunakan dataset yang telah melalui tahap pra-pemrosesan pada ARIMAX, termasuk pembentukan variabel *dummy* hari besar keagamaan (D1–D12), *dummy* bulan (M1–M12), *dummy weekend*, *double date*, serta fitur *time-lag*. *Time-lag* merepresentasikan nilai indeks pencarian pada beberapa hari sebelumnya yang digunakan sebagai prediktor untuk memodelkan nilai pada periode berikutnya (Amnur and Pramana, 2025)(Handayani and Qutub, 2025)(Moges et al., 2024). Variabel ini digunakan untuk membantu Random Forest menangkap pola historis pada data.

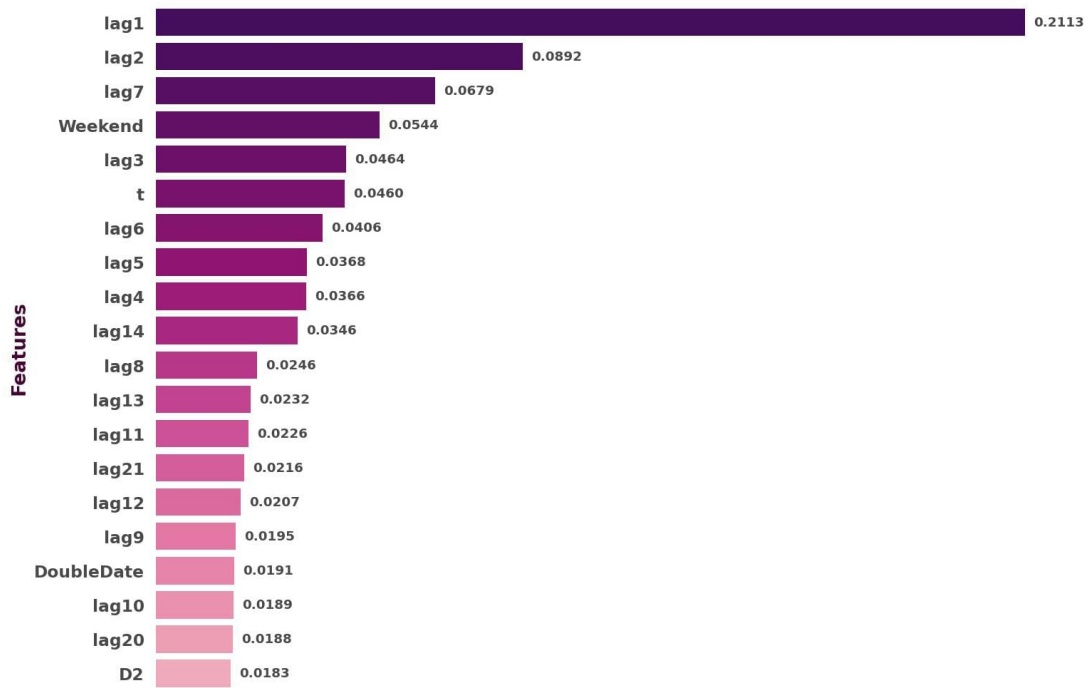
Berdasarkan hasil analisis ACF dan PACF, nilai autokorelasi dan parsial autokorelasi menunjukkan signifikansi pada beberapa lag awal hingga lag tertentu, sehingga fitur *time-lag* dibentuk secara berurutan dari lag 1 hingga 21. Pemilihan rentang ini bertujuan untuk menangkap ketergantungan data terhadap nilai historis secara lebih menyeluruh.

Hyperparameter Tuning Random Forest

Hasil *hyperparameter tuning* menggunakan *GridSearchCV* dengan *5-fold cross-validation*, diperoleh kombinasi *hyperparameter* terbaik yang digunakan dalam pembentukan model *Random Forest* dengan kombinasi parameter seperti yang disajikan pada Tabel 7 berikut.

Table 7: Ruang Pencarian dan Nilai Terbaik *Hyperparameter* Random Forest

<i>Hyperparameter</i>	Nilai yang Diuji	Nilai Terbaik
<i>n_estimators</i>	100, 200, 300	100
<i>max_depth</i>	10, 20, 30, None	30
<i>min_samples_split</i>	2, 5, 10	10
<i>min_samples_leaf</i>	1, 2, 4	1
<i>max_features</i>	auto, sqrt, log2	sqrt
<i>bootstrap</i>	True, False	False

Feature Importance Random ForestFigure 3: Nilai *Feature Importance*

Analisis *feature importance* dilakukan untuk mengidentifikasi kontribusi masing-masing variabel prediktor dalam memengaruhi hasil peramalan tren pencarian kata kunci “Wardah”. Hasil analisis pada Gambar 3 menunjukkan bahwa fitur *time-lag* memberikan kontribusi paling besar terhadap hasil peramalan. Variabel *lag 1* memiliki nilai *importance* tertinggi sebesar 0,2113, diikuti oleh *lag 2* sebesar 0,0892 dan *lag 7* sebesar 0,0679. Hasil tersebut menunjukkan bahwa indeks pencarian pada satu hingga tujuh hari sebelumnya memiliki pengaruh yang kuat dalam meramalkan tren pencarian kata kunci “Wardah”, sehingga mengindikasikan adanya ketergantungan temporal pada data.

Selain fitur *time-lag*, variabel tren (*t*) juga memberikan kontribusi yang cukup besar dengan nilai *importance* sebesar 0,0460. Hal ini menunjukkan bahwa perubahan tren pencarian dari waktu ke waktu turut memengaruhi hasil peramalan, sehingga informasi mengenai perkembangan pola pencarian secara keseluruhan dapat dimanfaatkan oleh model dalam menghasilkan

peramalan yang lebih akurat.

Variabel *dummy Weekend* memiliki nilai *importance* sebesar 0,0544, yang mengindikasikan adanya perbedaan pola pencarian antara akhir pekan dan hari kerja. Sementara itu, variabel *DoubleDate* memiliki nilai *importance* sebesar 0,0191 yang menunjukkan bahwa periode promosi *e-commerce* seperti 1.1, 2.2 hingga 12.12 turut memengaruhi peningkatan volume pencarian. Selain itu, variabel *dummy* hari raya Idul Fitri (D2) dengan nilai *importance* sebesar 0,0183 mengindikasikan bahwa momen hari besar keagamaan juga berkontribusi terhadap perubahan tren pencarian kata kunci "Wardah", meskipun pengaruhnya relatif lebih kecil dibandingkan fitur *time-lag* dan variabel tren (t).

3.3 Perbandingan Performa Peramalan Model ARIMAX dan *Random Forest*

Berdasarkan Tabel 8 disajikan hasil evaluasi performa model pada data *in-sample* dan *out-sample*, model *Random Forest* memberikan kinerja lebih baik dibandingkan ARIMAX. Pada data *out-sample*, *Random Forest* memperoleh RMSE sebesar 6,38, MAE sebesar 5,03, dan MAPE sebesar 6,16%, sedangkan ARIMAX memperoleh RMSE sebesar 7,71, MAE sebesar 6,17, dan MAPE sebesar 7,45%. Hal ini menunjukkan bahwa keunggulan *Random Forest* tidak hanya terjadi pada data pelatihan tetapi juga memiliki kemampuan generalisasi yang lebih baik dalam meramalkan data baru yang tidak digunakan pada proses pembentukan model. Secara keseluruhan, hasil evaluasi memperlihatkan bahwa *Random Forest* memberikan kinerja yang lebih stabil dalam meramalkan tren pencarian kata kunci "Wardah" pada periode *in-sample* maupun *out-sample*. Perbedaan performa ini juga menegaskan bahwa pendekatan berbasis algoritma *ensemble* seperti *Random Forest* lebih fleksibel dalam menangani pola data non-linear dan berubah-ubah, seperti pada data pencarian harian *Google Trends*.

Table 8: Hasil Perbandingan Evaluasi Model ARIMAX dan *Random Forest*

Model	Data	RMSE	MAE	MAPE (%)
ARIMAX	<i>In-Sample (Train)</i>	6,74	4,99	6,59
ARIMAX	<i>Out-Sample (Test)</i>	7,71	6,17	7,45
<i>Random Forest</i>	<i>In-Sample (Train)</i>	2,35	1,74	2,27
<i>Random Forest</i>	<i>Out-Sample (Test)</i>	6,38	5,03	6,16

Perbedaan kinerja tersebut berkaitan dengan karakteristik kedua pendekatan. Sebagai pendekatan statistik, ARIMAX memodelkan hubungan linier dalam deret waktu dengan mempertimbangkan struktur temporal data dan pengaruh variabel eksogen. Oleh karena itu, hasil peramalan ARIMAX dipengaruhi oleh pola ketergantungan temporal dan hubungan statistik pada data historis. Ketika data *Google Trends* mengalami perubahan yang bersifat tidak teratur atau fluktuasi yang relatif tajam, model ARIMAX kurang mampu mengikuti perubahan pola yang terjadi pada data. Sementara itu, *Random Forest* merupakan pendekatan *machine learning* yang memiliki fleksibilitas dalam mempelajari hubungan non-linier dan pola yang kompleks melalui kombinasi sejumlah *decision tree*. Karakteristik tersebut memungkinkan *Random Forest* menangkap variasi dan hubungan antarfitur yang lebih beragam. Dengan demikian, fleksibilitas *Random Forest* menjadi salah satu faktor yang mendukung kinerja model dalam memberikan hasil peramalan yang lebih baik dibandingkan ARIMAX, baik pada data *in-sample* dan *out-sample* untuk data pencarian kata kunci "Wardah".

Perbedaan karakteristik tersebut juga terlihat pada Gambar 4 dan Gambar 5, data aktual menunjukkan fluktuasi yang cukup tinggi dengan beberapa kenaikan dan penurunan se-

cara tajam. Pada periode *out-sample*, hasil peramalan ARIMAX cenderung lebih kaku sehingga pada beberapa titik belum dapat mengikuti perubahan data aktual secara optimal. Sebaliknya, hasil peramalan *out-sample Random Forest* terlihat lebih halus dan bergerak mendekati nilai aktual, terutama pada titik-titik ekstrem ketika tren mengalami kenaikan atau penurunan pencarian lebih cepat. Temuan ini memperlihatkan perbedaan karakteristik kedua pendekatan, yaitu ARIMAX lebih menekankan pemodelan berdasarkan struktur statistik deret waktu sedangkan *Random Forest* lebih fleksibel dalam mempelajari pola dan hubungan antarvariabel dari data historis. Dengan demikian, hasil Tabel 8, Gambar 4, dan Gambar 5 menunjukkan bahwa perbedaan pendekatan statistik dan *machine learning* berpengaruh terhadap karakteristik hasil peramalan. Pada penelitian ini, fleksibilitas model *Random Forest* dalam mempelajari pola data yang kompleks mampu menghasilkan nilai error yang lebih rendah dibandingkan model ARIMAX pada periode *in-sample* dan *out-sample*.

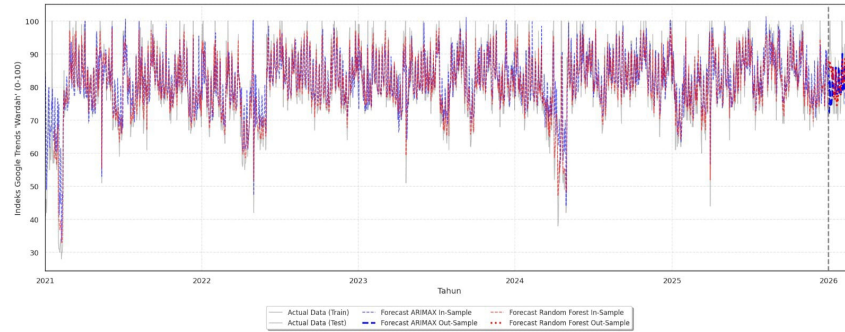


Figure 4: Grafik Perbandingan Model ARIMAX dan *Random Forest*

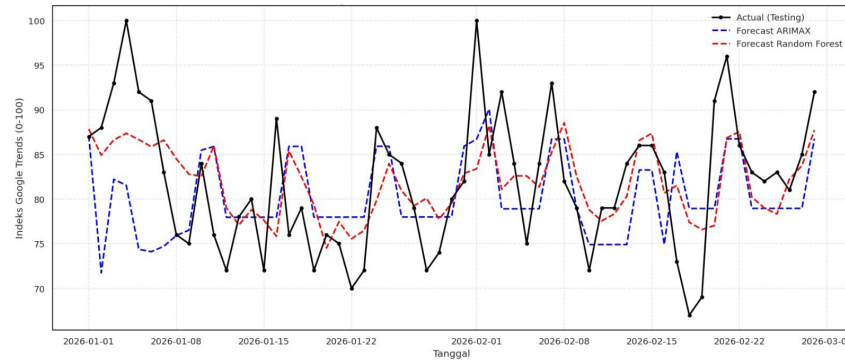


Figure 5: Grafik Perbandingan Hasil Peramalan *Out-Sample* Model ARIMAX dan *Random Forest*

4 Kesimpulan

Penelitian ini membandingkan performa model ARIMAX dan *Random Forest* dalam peramalan tren pencarian kata kunci "Wardah" menggunakan data *Google Trends*. Berdasarkan hasil evaluasi, *Random Forest* memberikan performa yang lebih baik dibandingkan ARIMAX pada data *in-sample* maupun *out-sample*, ditunjukkan oleh nilai RMSE, MAE, dan MAPE yang lebih rendah

serta kemampuan yang lebih baik dalam mengikuti pola fluktuasi data. Hasil ini menunjukkan bahwa model *Random Forest* lebih efektif dalam meramalkan perubahan tren pencarian kata kunci "Wardah" pada data penelitian. Hasil peramalan ini dapat dimanfaatkan sebagai informasi pendukung dalam mengoptimalkan perencanaan strategi promosi digital dan pengelolaan stok produk, dengan mempertimbangkan pola historis pencarian terutama menjelang akhir pekan (*weekend*), tanggal kembar (*double date*), dan Hari Raya Idul Fitri yang cenderung mengalami peningkatan minat pencarian masyarakat terhadap merek Wardah.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Program Studi Sarjana Sains Data Universitas Negeri Surabaya serta seluruh pihak yang telah memberikan dukungan dan masukan selama pelaksanaan penelitian ini.

References

- Amnur, M. A. and Pramana, S. (2025). Peramalan produksi perikanan laut di provinsi jawa tengah: Pendekatan statistik dan machine learning. *Seminar Nasional Official Statistics*, 2025(1):267–276.
- Chai, T. and Draxler, R. R. (2014). Root mean square error (rmse) or mean absolute error (mae)? – arguments against avoiding rmse in the literature. *Geoscientific Model Development*, 7(3):1247–1250.
- Compas (2025). Top 10 brand perawatan kecantikan di shopee indonesia kuartal i 2025: Siapa yang memimpin? Compas.
- Dani, A. T., Fauziyah, M., and Sandariria, H. (2023a). Forecasting the search trends of the keyword "sarung wadimor" in indonesia on google trends data using time series regression with calendar variation and arima box-jenkins. *Jurnal Matematika, Statistika dan Komputasi*, 19(3):447–459.
- Dani, A. T. R., Wahyuningsih, S., Putra, F. B., Fauziyah, M., Wigantono, S., Sandariria, H., A'yun, Q. Q., and Zen, M. A. (2023b). Aplikasi model arimax dengan efek variasi kalender untuk peramalan trend pencarian kata kunci "zalora" pada data google trends. *Inferensi*, 6(2):107–116.
- Du, B. (2026). Improving the accuracy of machine learning models in predicting evacuated tube solar collector through random forest replacement. *Thermal Science and Engineering Progress*, 69:104480.
- Ferdian, P. R., Kurniawan, R., Kosim, M., Salsabil, D. P., Anggraeni, R., and Maulana, R. A. (2025). Analisis perubahan perilaku konsumen terhadap umkm di cifest cikarang selatan selama periode libur: Studi kasus pembelian online vs. offline. *Jurnal Ilmiah Ekonomi, Akuntansi, dan Pajak*.
- GoodStats (2025a). 10 brand paket kecantikan terlaris di shopee indonesia q2 2025, mana favoritmu? GoodStats Data.
- GoodStats (2025b). 10 brand skincare terlaris di shopee indonesia 2025, mana favoritmu? GoodStats Data.
- Gujarati, D. N. (2006). *Basic Econometrics*. McGraw Hill, 4 edition.
- Handayani, D. N. and Qutub, S. (2025). Penerapan random forest untuk prediksi dan analisis kemiskinan. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(2):405–412.

- Hyndman, R. J. and Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. OTexts, Melbourne, Australia, 3rd edition.
- Kim, S. and Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3):669–679.
- Leilani, A., Armandari, M. O., and Ambarwati, N. S. S. (2023). Hubungan promosi tanggal kembar dengan perilaku konsumtif kosmetik pengguna marketplace.
- Moges, D. M., Virro, H., Kmoch, A., Cibin, R., Rohith, R. A. N., Martínez-Salvador, A., Conesa-García, C., and Uuemaa, E. (2024). Streamflow prediction with time-lag-informed random forest and its performance compared to swat in diverse catchments. *Water*.
- Oktaviansyah, M. R. and Wulansari, D. (2025). Analisis daya beli pada hari libur dan hari biasa terhadap penjualan di tujuh dua frozen food. *Jurnal Inovasi Manajemen dan Kewirausahaan*, 6(3).
- Prasojo, A. P. S. (2022). Pandemi covid-19 dan turnover ke pekerjaan informal: Analisis data google trends. *Jurnal Aplikasi Statistika Komputasi Statistik*, 14(1):49–62.
- Rahmawati, A., Nurawati, E., and Sugiyarto, T. (2024). Nowcasting hotel room occupancy rate using google trends index and online traveler reviews given lag effect with machine learning (case research: East kalimantan province). *Scientific Journal of Informatics*, 11(2):507–522.
- Risma, R. and Sahriman, S. (2020). Perbandingan estimasi metode kuadrat terkecil terboboti dan metode transformasi box-cox pada data heteroskedastisitas. *ESTIMASI: Journal of Statistics and Its Application*.
- Rizalde, F. A., Mulyani, S., and Bachtiar, N. (2021). Forecasting hotel occupancy rate in riau province using arima and arimax. In *Proceedings of the International Conference on Data Science and Official Statistics*, volume 2021, pages 578–589.
- Santika, E. F. (2024). Tren belanja produk kecantikan, diproyeksi tembus rp262 t pada 2025.
- Septiansyah, R., Zahra, H. N., Rahmadan, A. R. P., Simbolon, S. K., Sari, S. P., and Effendi, M. (2025). Modeling and forecasting rainfall in balikpapan using the arimax method: Case study of rainfall in balikpapan city from january to december 2024. *BESTARI BPS Kalimantan Timur*, 5(2).
- Shumway, R. H. and Stoffer, D. S. (2011). *Time Series Analysis and Its Applications*. Springer, New York.
- Subekti, S. T. and Rahayu (2025). Airlangga: Industri kosmetik catat kinerja positif, tumbuh 4,73
- Urrochman, M. Y., Asy'ari, H., and Hizham, F. A. (2025). Performance comparison of random forest regression and svr models in stock price prediction. *Jurnal Pilar Nusa Mandiri*, 21(1):16–23.
- Wei, W. W. S. (2006). *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Education, Inc., United States of America, 2 edition.
- Yin, L., Li, B., Li, P., and Zhang, R. (2023). Research on stock trend prediction method based on optimized random forest. *CAAI Transactions on Intelligence Technology*, 8(1):274–284.
- Zaghwan, A., Enshaiei, H., Majidiyan, H., and Amer, Y. (2026). Evaluating classical time-series models arima, sarima, arimax, and var for wave height forecasting. *Ocean Engineering*, 349:124098.
- Zeller, A. (2023). Google trends search volume explained. Exploding Topics.