

# Prediksi Wisatawan Mancanegara di Indonesia Menggunakan Metode SARIMAX dengan Efek Variasi Kalender Libur Nasional

Desya Neydi Putri Pakaya<sup>1\*</sup>, Novianita Achmad<sup>2</sup>, Isran K. Hasan<sup>1</sup>, Djihad Wunggul<sup>2</sup>, and Siti Nurmardia Abdussamad<sup>1</sup>

<sup>1</sup>Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Gorontalo, Gorontalo, Indonesia

<sup>2</sup>Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Gorontalo, Gorontalo, Indonesia

## Abstrak

Fluktuasi jumlah kunjungan wisatawan mancanegara seringkali menghasilkan nilai *outlier* yang dapat mengganggu akurasi model peramalan. Penelitian ini menggunakan pendekatan boxplot untuk mendekripsi *outlier*, diikuti dengan transformasi Logaritma Natural (*ln*) sebagai langkah penanganan. Metode *Seasonal Autoregressive Integrated Moving Average with Exogenous Variables* (SARIMAX) diterapkan dengan mempertimbangkan tiga variabel eksogen yang menunjukkan efek variasi kalender Libur Nasional berupa Hari Raya Nyepi, Hari Raya Idul Fitri dan libur akhir tahun. Hasil analisis menunjukkan bahwa ketiga variabel tersebut berpengaruh positif terhadap peningkatan jumlah kunjungan wisatawan mancanegara, di mana Hari Raya Nyepi memberikan kontribusi terbesar dibandingkan dua periode libur lainnya. Model 2 (0,1,1)(1,0,1)[12] terpilih sebagai model yang paling optimal berdasarkan hasil evaluasi terhadap beberapa model yang telah dibandingkan. Model ini menunjukkan performa yang sangat baik, ditunjukkan oleh nilai *Mean Absolute Percentage Error* (MAPE) sebesar 3.75% yang mengindikasikan bahwa model memiliki akurasi prediksi yang sangat tinggi. Sehingga model SARIMAX efektif dalam memodelkan dan memprediksi jumlah kunjungan wisatawan mancanegara di Indonesia.

**Kata Kunci:** Libur Nasional; *Outlier*; SARIMAX; Variasi Kalender; Wisatawan Mancanegara

## Abstract

Fluctuations in the number of foreign tourist arrivals often produce outlier values that can interfere with the accuracy of the forecasting model. This study uses a boxplot approach to detect outliers, followed by Natural Logarithm (*ln*) transformation as a treatment step. The Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) method is applied by considering three exogenous variables that show the effect of variations in the National Holiday calendar in the form of Nyepi Day, Idul Fitri Day and year-end holidays. The results of the analysis show that the three variables have a positive effect on the increase in the number of foreign tourist arrivals, where Nyepi Day makes the largest contribution compared to the other two holiday periods. Model 2 (0,1,1)(1,0,1)[12] was selected as the most optimal model based on the evaluation results of several models that have been compared. This model shows excellent performance, indicated by the Mean Absolute Percentage Error (MAPE) value of 3.75% which indicates that the model has very high prediction accuracy. So that the SARIMAX model is effective in modeling and predicting the number of foreign tourist visits in Indonesia.

\*Corresponding author. E-mail: [desyaneydiii@gmail.com](mailto:desyaneydiii@gmail.com)

**Keywords:** National Holidays; Outliers; SARIMAX; Calendar Variation; Foreign Tourists

Copyright © 2025 by Authors, Published by JRMM Group. This is an open access article under the CC BY-SA License (<https://creativecommons.org/licenses/by-sa/4.0>)

---

## 1 Pendahuluan

Potensi pariwisata di Indonesia sangatlah besar karena setiap wilayah memiliki keunikannya masing-masing, yang dapat memberikan manfaat serta keuntungan bagi Indonesia [1]. Keberhasilan sektor pariwisata di negara ini bergantung pada kekayaan alam dan keunikan budaya yang dimiliki, sehingga menjadi daya tarik tersendiri bagi wisatawan asing untuk mengunjunginya. Jumlah wisatawan mancanegara yang berkunjung ke Indonesia terus menunjukkan peningkatan setiap tahun, dengan Indonesia tetap menempati peringkat keempat di antara negara-negara ASEAN lainnya [2].

Berdasarkan data dari Badan Pusat Statistik jumlah kunjungan wisatawan mancanegara di Indonesia menunjukkan adanya fluktuasi. Penurunan drastis terjadi pada tahun 2020 hingga tahun 2021 dengan jumlah kunjungan yang sangat rendah akibat pembatasan perjalanan selama pandemi yang hampir menghentikan arus wisatawan. Hingga tahun 2022 sampai pertengahan 2024, jumlah kunjungan perlamban meningkat kembali, dipengaruhi oleh libur panjang dan pemulihian aktivitas pariwisata.

Fluktuasi kunjungan wisatawan mancanegara menimbulkan tantangan serius bagi sektor pariwisata di Indonesia, yang kini dihadapkan pada fenomena *overtourism*. *Overtourism* terjadi pada periode-periode tertentu dengan konsentrasi kunjungan wisatawan yang mengalami fluktuasi, seperti Hari Raya Nyepi, Hari Raya Idul Fitri dan libur akhir tahun. Perubahan pola kunjungan ini tidak hanya dipengaruhi oleh faktor eksternal seperti pandemi saja, tetapi juga sangat erat kaitannya dengan efek variasi kalender atau *holiday variation*, yang mempengaruhi waktu kedatangan wisatawan.

Kondisi fluktuasi ini dapat menyebabkan munculnya nilai pencilan (*outlier*) dalam data. *Outlier* merupakan nilai yang menyimpang secara signifikan dari pola distribusi data pada umumnya, sehingga dapat mengganggu estimasi parameter statistik dan dapat menurunkan kinerja model prediktif yang dipakai [3]. Salah satu pendekatan yang umum digunakan untuk mendeteksi *outlier* adalah melalui visualisasi boxplot. Kemudian dilakukan penanganan menggunakan transformasi Logaritma Natural ( $\ln$ ), yang berfungsi untuk menstabilkan varians dan mengurangi pengaruh nilai ekstrim. Setelah proses ini dilakukan, langkah selanjutnya memilih model peramalan yang mampu menggambarkan pola data.

Salah satu model yang umum digunakan dalam peramalan adalah *Autoregressive Integrated Moving Average* (ARIMA). ARIMA terbagi menjadi dua jenis, yaitu ARIMA non-musiman dan ARIMA musiman [4]. ARIMA musiman biasa dikenal sebagai *Seasonal ARIMA* (SARIMA), dengan notasi model yang dituliskan sebagai  $\text{SARIMA}(p, d, q)(P, D, Q)_S$  [5]. Namun, model ini memiliki keterbatasan karena hanya dapat diterapkan pada data *time series* tunggal dan tidak mempertimbangkan variabel eksogen yang mempengaruhi data.

Sebagai pengembangan dari SARIMA, metode *Seasonal Autoregressive Integrated Moving Average with Exogenous Variables* (SARIMAX) dirancang untuk mengatasi keterbatasan dengan memasukkan variabel eksogen ke dalam model. Beberapa penelitian sebelumnya telah membahas peramalan jumlah kunjungan wisatawan mancanegara dengan berbagai metode. Rais dkk [6] menggunakan ARIMA dan menemukan bahwa kunjungan tertinggi diperkirakan terjadi pada Agustus 2021, sedangkan terendah pada Desember 2020. Hendayanti dan Nurhidayati [7] membandingkan SARIMA dan SVR untuk kunjungan wisatawan ke Bali, dengan hasil bahwa SARIMA memberikan akurasi lebih tinggi. Sementara itu, Andika dkk [8] menerapkan metode *Fuzzy Time Series Chen* untuk Provinsi Aceh dan mencatat peningkatan kunjungan pada Agustus

2022 pasca pandemi.

Sebagian keterbatasan dari metode-metode di atas dalam memperkirakan jumlah kunjungan wisatawan mancanegara adalah tidak mempertimbangkan faktor eksternal yang dapat mempengaruhi pola data. Salah satunya dengan mempertimbangkan Libur Nasional, berupa Hari Raya Nyepi, Hari Raya Idul Fitri dan libur akhir tahun, yang seringkali memicu lonjakan atau penurunan kunjungan wisatawan. Penelitian ini mengintegrasikan variabel *dummy* pada Libur Nasional ke dalam model SARIMAX serta menangani keberadaan *outlier* yang kerap mengganggu akurasi hasil peramalan. Penelitian ini juga berkontribusi dalam penyusunan strategi pengembangan sektor pariwisata, agar dapat mengoptimalkan layanan kepada wisatawan, meningkatkan daya tarik destinasi di Indonesia, serta mendukung pengambilan keputusan yang lebih tepat dalam pengelolaan sektor pariwisata. Oleh karena itu, metode ini menjadi pilihan yang tepat untuk memprediksi jumlah kunjungan wisatawan mancanegara di Indonesia.

## 2 Metode

Metodologi dalam penelitian ini disusun secara sistematis untuk mendukung tujuan analisis dan peramalan. Pendekatan yang digunakan mencakup penentuan data, variabel yang relevan, serta tahap analisis menggunakan metode SARIMAX. Penjabaran setiap tahap metodologi dijelaskan pada bagian berikut.

### 2.1 Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari situs resmi Badan Pusat Statistik (BPS) Indonesia, mengenai jumlah kunjungan wisatawan mancanegara di Indonesia dari tahun 2020 hingga Agustus 2024, yang disajikan dalam bentuk tabel statistik.

### 2.2 Variabel Penelitian

Variabel yang digunakan terdiri atas variabel dependen, yaitu jumlah kunjungan wisatawan mancanegara (Y), serta tiga variabel independen berupa variabel *dummy* Libur Nasional Hari Raya Nyepi (X1), Hari Raya Idul Fitri (X2), dan libur akhir tahun (X3).

### 2.3 *Outlier*

*Outlier* adalah suatu amatan yang memiliki nilai sangat berbeda atau jauh dari mayoritas nilai lainnya dalam suatu kumpulan data. Keberadaan *outlier* tidak hanya menyebabkan distribusi data menjadi tidak normal, tetapi juga dapat mempengaruhi hasil analisis terutama dalam pengambilan keputusan penelitian [9]. Jika dalam dataset mengandung *outlier*, maka diperlukan metode diagnostik untuk mendeteksi keberadaannya serta melakukan langkah penanganan yang sesuai.

#### 1. Deteksi *Outlier*

Salah satu pendekatan yang sering diterapkan untuk mendeteksi *outlier* adalah melalui visualisasi boxplot. Sebuah data disebut sebagai *outlier* jika nilainya terletak di luar area dan *whisker* pada boxplot, yaitu lebih tinggi dari Q3 ditambah 1.5 kali IQR atau lebih rendah dari Q1 dikurangi 1.5 kali IQR. Q1 mengacu pada kuartil bawah, Q3 kuartil atas, dan IQR merupakan selisih antara Q3 dan Q1 [10].

#### 2. Transformasi Logaritma Natural (*ln*)

Transformasi *ln* digunakan untuk menangani permasalahan data yang mengandung *outlier* atau memiliki distribusi yang tidak normal. Menurut Lestari dan Pabulo [11], transformasi *ln* digunakan sebagai langkah untuk melakukan analisis ketika ditemukan *outlier* dalam data. Hal ini disebabkan oleh sifat fungsi logaritma yang memperkecil perbedaan antara

nilai-nilai besar dan kecil, sehingga nilai ekstrim tidak lagi memberikan pengaruh yang terlalu besar terhadap hasil analisis.

## 2.4 Autoregressive Integrated Moving Average (ARIMA)

ARIMA digunakan untuk meramalkan data univariat dengan hanya mengandalkan pola historis dari data itu sendiri tanpa melibatkan variabel independen lainnya. Berikut persamaan model ARIMA [12]:

### 1. Autoregressive (AR)

AR merupakan komponen utama dalam model ARIMA yang diidentifikasi sebagai orde  $p$  atau dapat dituliskan sebagai ARIMA( $p, 0, 0$ ).

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t$$

### 2. Moving Average (MA)

MA dinotasikan sebagai MA( $q$ ), di mana  $q$  merepresentasikan orde model MA, yang juga dikenal sebagai ARIMA( $0, 0, q$ ).

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

### 3. Autoregressive Moving Average (ARMA)

Model ARMA menggabungkan dua komponen utama yaitu, AR dan MA. Model ini dinyatakan dalam bentuk ARIMA ( $p, 0, q$ ).

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

### 4. Autoregressive Integrated Moving Average (ARIMA)

ARIMA digunakan untuk menganalisis dan memprediksi data *time series*, terutama ketika data tersebut tidak bersifat konstan. Jika data tidak stasioner, maka perlu dilakukan *differencing* sebelum membangun model ARIMA. Berikut persamaan model ARIMA [13] :

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)a_t$$

## 2.5 Seasonal ARIMA (SARIMA)

SARIMA dimanfaatkan untuk menganalisis data deret waktu dengan mempertimbangkan pola musiman sebagai pengembangan dari model ARIMA [14]. Model ini memiliki tingkat akurasi yang tinggi dalam melakukan prediksi jangka pendek dan tidak memperhitungkan variabel independen dalam proses peramalannya. Berikut persamaan model SARIMA [13]:

$$\Phi_P(B^S)\phi_p(B)(1-B)^d(1-B^S)^D Y_t = \theta_q(B)\Theta_Q(B^S)a_t$$

Model SARIMA dipilih berdasarkan perbandingan nilai *Akaike Information Criterion* (AIC). Model dengan nilai AIC paling rendah dianggap sebagai model terbaik. AIC adalah estimator yang mengukur kualitas relatif model untuk suatu kumpulan data.

## 2.6 SARIMAX (Efek Variasi Kalender)

Metode ini dinyatakan dalam  $(p, d, q)(P, D, Q)S(X)$  dengan  $X$  sebagai vektor dari variabel eksogen. Dalam penelitian ini, variabel eksogen yang dimaksud adalah variabel *dummy* yang berfungsi untuk mewakili kategori atau kelompok dari suatu variabel kualitatif atau kategorikal. Variabel *dummy* ini biasanya diberi nilai 0 dan 1. Model regresi dengan variabel *dummy* secara umum dinyatakan sebagai berikut [15]:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \dots + \beta_k X_{k,t} + \omega_t \quad (1)$$

Dengan  $\beta_0$  sebagai konstanta (*intercept*),  $\beta_1, \beta_2, \dots, \beta_k$  sebagai koefisien regresi untuk variabel *dummy*,  $X_{1,t}, X_{2,t}, \dots, X_{k,t}$  sebagai variabel *dummy* efek variasi kalender, serta  $\omega_t$  sebagai residual dari variabel *dummy*. Deret residual  $\omega_t$  dapat dinyatakan dalam bentuk SARIMA sebagai berikut [16]:

$$\omega_t = \frac{\theta_q(B) \Theta_Q(B^S)}{\phi_p(B) \Phi_P(B^S) (1 - B)^d (1 - B^S)^D} a_t \quad (2)$$

Persamaan model SARIMAX secara umum dapat dinyatakan sebagai berikut:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \dots + \beta_k X_{k,t} + \left( \frac{\theta_q(B) \Theta_Q(B^S)}{\phi_P(B) \Phi_P(B^S) (1 - B)^d (1 - B^S)^D} a_t \right) \quad (3)$$

## 2.7 Identifikasi Model

Identifikasi ini bertujuan untuk memeriksa kestasioneran data dan menganalisis plot ACF dan PACF, yang masing-masing dapat dijelaskan sebagai berikut:

### 1. Stasioner

Stasioner adalah kondisi di mana data tidak mengalami perubahan ekstrim, sehingga tetap berada di sekitar nilai rata-rata yang konsisten. Dalam proses peramalan deret waktu, penting untuk memastikan bahwa data telah memenuhi kriteria stasioneritas [17]. Untuk menangani data yang tidak stasioner terhadap nilai tengah, proses *differencing* dapat diterapkan untuk mencapai stasioneritas rata-rata.

### 2. Plot ACF dan PACF

Koefisien ACF dan PACF juga digunakan untuk mengevaluasi kestasioneran data serta memberikan petunjuk dalam pemilihan model awal.

## 2.8 Pemeriksaan Diagnostik

Evaluasi diperlukan untuk memeriksa adanya autokorelasi dan memastikan bahwa residual telah memenuhi kriteria *white noise*. Menurut Fadliani dkk [18] pemeriksaan diagnostik terdiri atas dua tahapan, yakni pengujian signifikansi parameter dan evaluasi kesesuaian model. Uji kesesuaian model mencakup pengujian terhadap asumsi *white noise* pada residual dengan memanfaatkan uji statistik *Ljung-Box*, sebagai berikut:

### 1. Uji Signifikansi Parameter Model

Pengujian ini dilakukan untuk mengevaluasi model sementara yang telah ditetapkan. Tahapan ini bertujuan untuk menilai kelayakan model dalam proses analisis selanjutnya [19]. Dalam model ARIMA( $p, d, q$ ), salah satu indikator model yang baik adalah parameter yang signifikan berbeda dari nol, yang menandakan parameter tersebut memiliki peranan penting dalam menggambarkan pola data [18].

### 2. Uji Kesesuaian Model

Uji kesesuaian model dilakukan melalui pendekatan pengujian residual *white noise*. Uji *White Noise* atau uji independensi bertujuan untuk memeriksa apakah residual tidak memiliki korelasi dengan residual lainnya, sehingga residual harus bersifat independen. Residual dianggap memenuhi asumsi *white noise* jika nilai *p-value* lebih besar dari 0.05, yang menunjukkan bahwa residual tidak memiliki pola tertentu dan tidak berkorelasi [20].

## 2.9 Pemilihan Model Terbaik

Pemilihan model terbaik dilakukan dengan mempertimbangkan nilai *Akaike Information Criterion* (AIC). Nilai ini digunakan untuk mengevaluasi kesesuaian model dengan data yang ada. AIC

berfungsi untuk mengidentifikasi model yang dapat menjelaskan data dengan efisiensi jumlah parameter yang tepat. Berikut bentuk umum AIC [18]:

$$AIC = 2k - 2\ell$$

## 2.10 Evaluasi Model

Proses pembentukan model peramalan diawali dengan pemisahan data menjadi dua bagian, yakni data *in-sample* dan *out-sample*. Data *in-sample* berfungsi untuk menyusun atau melatih model, sementara data *out-sample* dimanfaatkan dalam tahap evaluasi untuk menilai hasil prediksi. Salah satu parameter yang umum digunakan untuk menilai kualitas model adalah *Mean Absolute Percentage Error* (MAPE), yang dapat dihitung menggunakan rumus berikut [21]:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \quad (4)$$

Dengan  $y_t$  sebagai nilai aktual pada waktu ke- $t$ , dan  $\hat{y}_t$  sebagai nilai prediksi pada waktu ke- $t$ . MAPE menunjukkan rata-rata kesalahan persentase antara nilai aktual dan prediksi. Semakin kecil nilai MAPE, maka semakin tinggi tingkat keakuratan model peramalan yang dibangun.

## 3 Hasil dan Pembahasan

Bagian ini menyajikan hasil dari seluruh tahapan analisis yang telah dilakukan. Setiap langkah dianalisis untuk melihat pola hubungan, dan pengaruh variabel terhadap jumlah wisatawan mancanegara.

### 3.1 Deskripsi Statistik

Deskripsi statistik digunakan untuk menyajikan dan meringkas data agar informasinya dapat diidentifikasi dengan mudah. Hasil deskripsi statistik disajikan pada tabel berikut ini.

**Tabel 1:** Deskripsi statistik

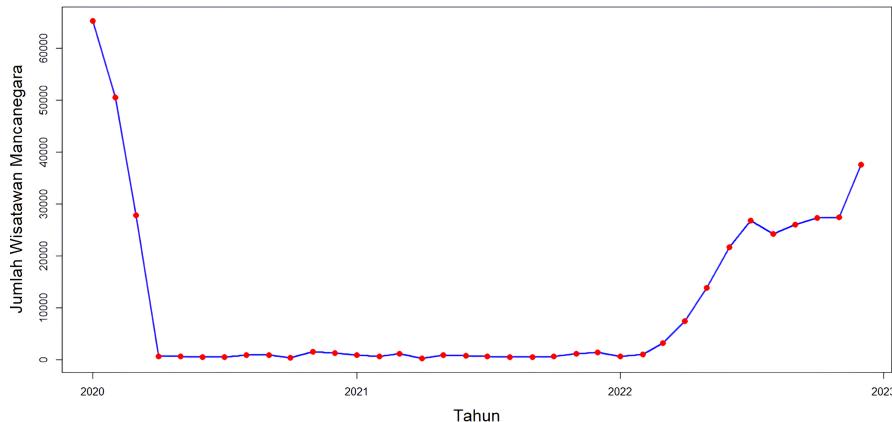
Tahun	Mean	Median	Minimum	Maksimum
2020	12.597	930.000	419.000	65.337
2021	829.000	744.000	263.000	1.487
2022	18.135	22.994	699.000	37.587
2023	34.958	35.293	25.344	45.550
2024	46.848	42.571	35.180	75.260

Tahun 2021, tercatat jauh lebih rendah dibandingkan tahun lainnya. Penurunan ini secara langsung dipengaruhi oleh kebijakan pembatasan perjalanan global yang diberlakukan selama pandemi COVID-19, menyebabkan mobilitas wisatawan mengalami penurunan drastis hingga hampir terhenti sepenuhnya. Sebaliknya, tahun 2024 mencatat nilai tertinggi, menunjukkan pemulihan sektor pariwisata setelah periode pandemi serta peningkatan dibandingkan tahun-tahun sebelumnya.

### 3.2 Identifikasi Data *In-Sample*

Proses identifikasi data diawali dengan pemisahan data menjadi dua bagian, yakni data *in-sample* dan *out-sample*. Data *in-sample* mencakup periode tahun 2020 hingga tahun 2022, yang digunakan untuk melatih model dan menguji kemampuannya dalam mempelajari pola historis. Sementara itu, data *out-sample* mencakup periode tahun 2023 hingga Agustus 2024,

yang digunakan untuk mengevaluasi kinerja model dalam melakukan prediksi. Sebaran data *in-sample* dapat diamati melalui plot *time series* berikut.



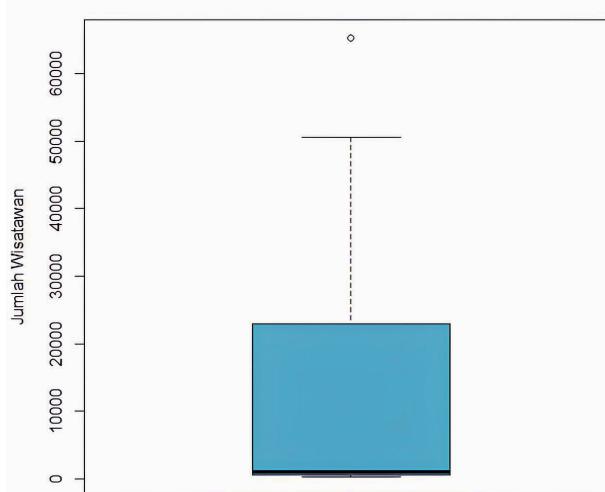
**Gambar 1:** Plot jumlah kunjungan wisatawan mancanegara di Indonesia (2020-2022)

Bisa dilihat dari **Gambar 1**, jumlah wisatawan mancanegara di Indonesia mengalami perubahan drastis selama periode 2020 hingga 2022. Terlihat adanya penyimpangan dalam data terutama pada awal tahun 2020, di mana jumlah wisatawan jauh lebih tinggi dibandingkan periode setelahnya. Fenomena ini menunjukkan kemungkinan adanya *outlier* yang dapat mempengaruhi hasil analisis. Oleh karena itu, diperlukan deteksi untuk mengidentifikasi keberadaan *outlier* dan menentukan penanganan yang tepat.

### 3.3 Deteksi dan Penanganan *Outlier*

#### 3.3.1 Deteksi *Outlier*

*Outlier* adalah suatu amatan yang memiliki nilai sangat berbeda atau jauh dari mayoritas nilai lainnya dalam suatu kumpulan data [9]. Keberadaan *outlier* dapat berdampak pada analisis terutama dalam model peramalan, karena dapat menyebabkan bias dalam estimasi parameter dan dapat meningkatkan kesalahan prediksi. Oleh karena itu, identifikasi keberadaan *outlier* menjadi langkah penting yang harus dilakukan, salah satunya dengan menggunakan pendekatan statistik seperti boxplot.



**Gambar 2:** Boxplot jumlah kunjungan wisatawan mancanegara

Berdasarkan **Gambar 2**, distribusi jumlah wisatawan menunjukkan adanya *outlier*. Garis

horizontal di dalam kotak merepresentasikan median, yang menunjukkan bahwa sebagian besar data cenderung memiliki nilai yang relatif rendah. Garis putus-putus (*whiskers*) menggambarkan rentang data yang masih dalam batas normal, sedangkan titik di atas *whiskers* menunjukkan keberadaan *outlier* yang berada jauh di atas Q3. Hal ini menandakan adanya bulan dengan jumlah wisatawan yang jauh lebih tinggi dibandingkan periode lainnya, sehingga transformasi diterapkan sebagai langkah penanganan.

### 3.3.2 Transformasi Logaritma Natural (*ln*)

Transformasi *ln* digunakan untuk mengatasi distribusi data yang tidak normal, terutama ketika data memiliki *skewness* positif yang tinggi dan rentang nilai yang sangat luas. *Skewness* ini menunjukkan bahwa terdapat beberapa nilai ekstrim yang secara signifikan lebih besar dibandingkan nilai lainnya. Untuk mengurangi dampak nilai ekstrim tersebut, transformasi *ln* diterapkan sebelum proses pemodelan terhadap variabel dependen, yaitu jumlah kunjungan wisatawan mancanegara, dengan tujuan untuk mengecilkan perbedaan antar nilai besar, menstabilkan varians, serta mengurangi efek *outlier* [11].

**Tabel 2:** Transformasi logaritma natural (*ln*)

Bulan	Data Aktual	Transformasi ( <i>ln</i> )
Jan-2020	65.337	11.087314
Feb-2020	50.600	10.831707
Mar-2020	27.795	10.232611
Apr-2020	625.000	6.437752
Mei-2020	648.000	6.473891
:	:	:
Jul-2024	37.315	10.525540
Agu-2024	51.132	10.854370

## 3.4 Uji Stasioneritas

Uji stasioneritas dilakukan pada data non-musiman maupun musiman guna memastikan bahwa keduanya memenuhi kriteria stasioneritas. Salah satu metode yang digunakan adalah uji *Augmented Dickey-Fuller* (ADF).

**Tabel 3:** Uji stasioner non-musiman

Tahap Analisis	ADF Statistic	p-value	Kesimpulan
Data Kunjungan	-1.6986	0.6904	Tidak stasioner
<i>Differencing</i> ke-1	-4.0223	0.0203	Stasioner

Hasil uji stasioner non-musiman menunjukkan bahwa data kunjungan memiliki nilai ADF *Statistic* sebesar -1.6986 dengan *p-value* sebesar 0.6904, sehingga data belum bersifat stasioner karena *p-value* melebihi tingkat signifikansi (0.05). Setelah dilakukan proses *differencing* sebanyak satu kali, nilai ADF *Statistic* meningkat menjadi -4.0223 dan *p-value* turun menjadi 0.0203. Karena nilai *p-value* berada di bawah (0.05), maka data sudah memenuhi syarat stasioneritas terhadap nilai tengah. Untuk memastikan bahwa data juga stasioner terhadap pola musiman, dilakukan uji stasioner pada data musiman.

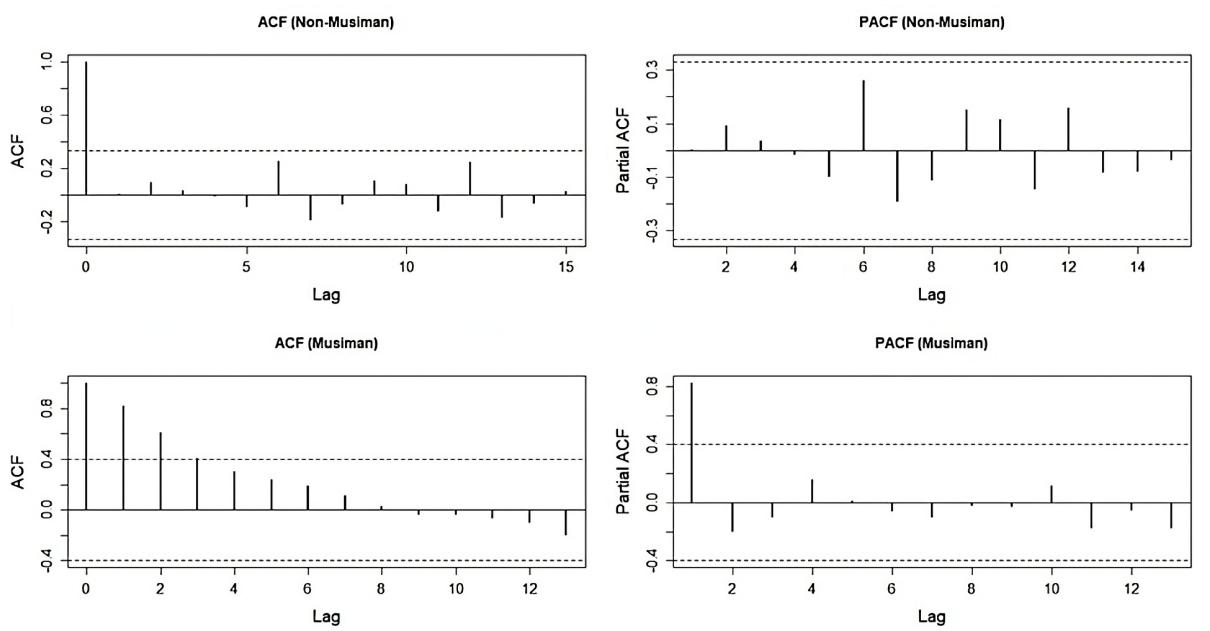
**Tabel 4:** Uji stasioner musiman

Tahap Analisis	ADF Statistic	p-value	Kesimpulan
<i>Differencing</i> Musiman	-3.76	0.0386	Stasioner

Hasil uji stasioner musiman menunjukkan bahwa setelah dilakukan *differencing*, nilai ADF *Statistic* menjadi -3.76 dengan *p-value* sebesar 0.0386. Karena *p-value* lebih kecil dari *alpha* (0.05), maka data telah memenuhi kriteria stasioner.

### 3.5 Identifikasi Plot ACF dan PACF

Setelah melakukan uji stasioneritas pada data, langkah berikutnya adalah menganalisis plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF). ACF dan PACF digunakan untuk mengevaluasi kestasioneran data serta memberikan petunjuk dalam pemilihan model awal [18]. Melalui analisis ini, dapat ditentukan parameter *p* dan *q* yang optimal dalam membangun model.



**Gambar 3:** Plot acf dan pacf pada data musiman dan non-musiman

Plot ACF untuk data non-musiman menunjukkan adanya satu *lag* yang melewati batas signifikansi yaitu *lag* 0, sedangkan plot PACF tidak menunjukkan *lag* yang signifikan karena seluruh garis *lag* berada di dalam batas. Sementara pada data musiman, plot ACF menunjukkan pola penurunan secara perlahan dengan tiga *lag* yang signifikan yaitu *lag* 0, *lag* 1 dan *lag* 2, sedangkan pada plot PACF hanya *lag* 1 yang melewati batas signifikansi. *Lag* yang signifikan ini dapat digunakan sebagai dasar dalam pembentukan model SARIMA. Beberapa model yang dapat dibentuk antara lain Model 1 (0,1,0)(1,0,1)[12], Model 2 (0,1,1)(1,0,1)[12], Model 3 (1,1,0)(1,0,1)[12], dan Model 4 (1,1,1)(1,0,1)[12].

### 3.6 Uji Signifikansi Parameter

Diketahui bahwa beberapa model masih mengandung parameter yang tidak signifikan. Artinya tidak semua parameter dalam model memberikan pengaruh terhadap hasil peramalan. Untuk menentukan model terbaik di antara model-model yang telah dibangun, dilakukan perbandingan berdasarkan nilai *Akaike Information Criterion* (AIC). AIC digunakan sebagai indikator evaluasi performa model, di mana model dengan nilai AIC terkecil dianggap sebagai model terbaik. Hasil perbandingan nilai AIC dapat dilihat pada tabel dibawah ini.

**Tabel 5:** Estimasi parameter

Model	Parameter	Estimasi	S.E	p-value	Kesimpulan
Model 1	SAR(1)	0.6629	0.3098	0.0323	Signifikan
	SMA(1)	-1.0000	0.3848	0.0093	Signifikan
	Xreg1	0.9431	0.2532	0.0002	Signifikan
	Xreg2	0.3234	0.2534	0.2003	Tidak Signifikan
	Xreg3	0.3593	0.2532	0.1553	Tidak Signifikan
Model 2	MA(1)	0.2512	0.1660	0.1309	Tidak Signifikan
	SAR(1)	0.6651	0.3175	0.0365	Signifikan
	SMA(1)	-1.0000	0.4023	0.0128	Signifikan
	Xreg1	0.9329	0.2183	<0.0001	Signifikan
	Xreg2	0.3791	0.2172	0.0802	Tidak Signifikan
	Xreg3	0.2735	0.2449	0.2623	Tidak Signifikan
Model 3	AR(1)	0.2425	0.1678	0.1485	Tidak Signifikan
	SAR(1)	0.7029	0.3122	0.0244	Signifikan
	SMA(1)	-1.0000	0.4017	0.0128	Signifikan
	Xreg1	0.9504	0.2221	<0.0001	Signifikan
	Xreg2	0.3779	0.2225	0.0894	Tidak Signifikan
	Xreg3	0.2660	0.2535	0.2941	Tidak Signifikan
Model 4	AR(1)	0.0478	0.7658	0.9506	Tidak Signifikan
	MA(1)	0.2062	0.7508	0.7830	Tidak Signifikan
	SAR(1)	0.6727	0.3441	0.0505	Tidak Signifikan
	SMA(1)	-1.0000	0.4021	0.0129	Signifikan
	Xreg1	0.9365	0.2258	<0.0001	Signifikan
	Xreg2	0.3802	0.2182	0.0814	Tidak Signifikan
	Xreg3	0.2704	0.2508	0.2809	Tidak Signifikan

**Tabel 6:** Perbandingan nilai AIC

Model	df	AIC
Model 1 (0,1,0)(1,0,1)[12]	6	89.45233
Model 2 (0,1,1)(1,0,1)[12]	7	89.40298
Model 3 (1,1,0)(1,0,1)[12]	7	89.47362
Model 4 (1,1,1)(1,0,1)[12]	8	91.39882

Berdasarkan **Tabel 6**, model 2 dipilih sebagai model terbaik di antara keempat model yang dibandingkan karena memiliki nilai AIC terendah, yaitu 89.40298. Namun, model ini masih mengandung parameter yang tidak signifikan, yaitu MA(1), Xreg2 dan Xreg3. Meskipun tidak signifikan secara statistik, variabel Xreg2 dan Xreg3 secara konseptual mewakili variasi kalender Libur Nasional yang berpotensi mempengaruhi kunjungan wisatawan, sehingga tetap digunakan untuk melihat apakah variabel tersebut berpotensi meningkatkan atau menurunkan jumlah kunjungan wisatawan. Pengujian lebih lanjut tetap dilakukan untuk memastikan kesesuaian model terbaik [4].

### 3.7 Uji Kesesuaian Model

Uji kesesuaian model bertujuan untuk memastikan bahwa model yang dipilih memenuhi asumsi yang diperlukan dalam analisis deret waktu. Pengujian dilakukan melalui pendekatan uji residual *white noise*.

### 3.7.1 Uji Residual White Noise

Evaluasi residual *white noise* membantu memastikan bahwa residual dari model bersifat acak dan tidak menunjukkan pola autokorelasi. Dengan kata lain, residual yang bersifat *white noise* menunjukkan bahwa model telah menangkap seluruh informasi yang terkandung dalam data. Dalam penelitian ini, pengujian dilakukan menggunakan *Ljung-Box Test* yang diterapkan melalui *software R Studio*.

**Tabel 7:** Hasil Uji Ljung-Box Test

Statistik Uji	X-squared	p-value	Kesimpulan
<i>Ljung-Box Test</i>	6.5034	0.8886	<i>White Noise</i>

Hasil uji *Ljung-Box Test* menunjukkan nilai statistik uji sebesar 6.5034 dengan *p-value* sebesar 0.8886. Karena *p-value* lebih besar dari taraf signifikansi (0.05), maka tidak terdapat cukup bukti untuk menolak  $H_0$ . Dengan demikian, residual dari Model 2 dapat dikatakan bersifat *white noise*, yang berarti tidak terdapat autokorelasi yang signifikan di dalam residual.

## 3.8 Pemodelan Akhir SARIMAX

Model akhir yang diperoleh dari proses estimasi dapat dituliskan sebagai berikut:

$$Y_t = 0.9329 \cdot h_t + 0.3791 \cdot l_t + 0.2735 \cdot v_t + \omega_t$$

$$\omega_t = \frac{(1 + 0.2512B)(1 - 1.0000B^{12})}{(1 - B)(1 - 0.6651B^{12})} \alpha_t$$

$$Y_t = 0.9329 \cdot h_t + 0.3791 \cdot l_t + 0.2735 \cdot v_t + \left( \frac{(1 + 0.2512B)(1 - 1.0000B^{12})}{(1 - B)(1 - 0.6651B^{12})} \alpha_t \right)$$

Model menunjukkan bahwa jumlah kunjungan wisatawan dipengaruhi oleh tiga faktor eksternal, yakni Hari Raya Nyepi ( $h_t$ ), Hari Raya Idul Fitri ( $l_t$ ), dan libur akhir tahun ( $v_t$ ). Berdasarkan koefisien dari ketiga variabel tersebut, Hari Raya Nyepi memberikan pengaruh paling besar terhadap peningkatan jumlah kunjungan wisatawan yakni sebesar 0.9329, sedangkan Hari Raya Idul Fitri dan libur akhir tahun tetap menunjukkan peningkatan meskipun pengaruhnya relatif lebih kecil dibandingkan Hari Raya Nyepi. Sementara dari sisi komponen stasioner, model ini menggunakan *differencing* non-musiman  $(1 - B)$  dan *differencing* musiman tahunan  $(1 - 0.6651B^{12})$ . Sementara itu, *Moving Average* pada data musiman diperoleh dari  $(1 + 0.2512B)$  dan komponen musiman tambahan  $(1 - 1.0000B^{12})$ . Sehingga bentuk akhir dari model SARIMAX sebagai berikut:

$$(1 - B)(1 - 0.6651B^{12})Y_t = 0.9329 \cdot h_t + 0.3791 \cdot l_t + 0.2735 \cdot v_t + (1 + 0.2512B)(1 - 1.0000B^{12})\alpha_t$$

Persamaan tersebut menggambarkan bahwa model SARIMAX tidak hanya mempertimbangkan pengaruh tiga variabel eksogen, yaitu Hari Raya Nyepi, Hari Raya Idul Fitri dan libur akhir tahun, tetapi juga menangkap pola musiman dan tren jangka panjang melalui komponen *differencing*.

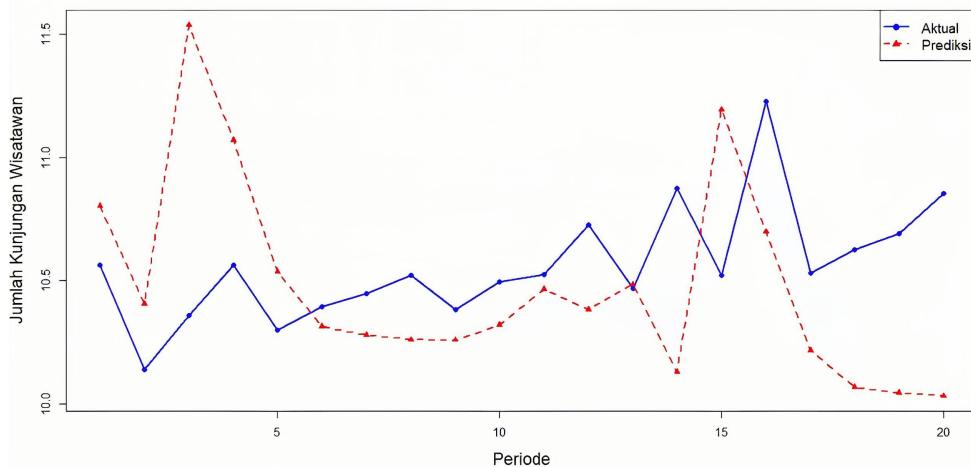
## 3.9 Evaluasi Model

Evaluasi terhadap hasil peramalan dilakukan dengan menghitung nilai *Mean Absolute Percentage Error* (MAPE), yang digunakan untuk mengukur tingkat keakuratan model dalam memprediksi

data. Semakin kecil nilai MAPE yang dihasilkan, maka semakin baik model tersebut dalam menghasilkan prediksi. Sehingga, perhitungan nilai MAPE sebagai berikut:

$$MAPE = \frac{1}{20} \sum_{t=1}^{20} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 = 3.75\%$$

Nilai MAPE yang diperoleh sebesar 3.75%. Nilai ini menunjukkan bahwa rata-rata tingkat kesalahan relatif antara nilai aktual dan nilai prediksi hanya sebesar 3.75%. Menurut Nur Hazimah Latief [4] MAPE di bawah 10% termasuk kategori akurasi prediksi tinggi. Hal ini mengindikasikan bahwa model yang digunakan mampu memprediksi data dengan tingkat kesalahan yang sangat rendah. Berikut ditampilkan perbandingan plot hasil prediksi dengan data aktual, untuk melihat sejauh mana model dapat mereplikasi pola pergerakan jumlah kunjungan wisatawan secara visual.



Gambar 4: Perbandingan plot hasil prediksi vs aktual

## 4 Kesimpulan

Hasil pembahasan menunjukkan bahwa model terbaik untuk meramalkan jumlah kunjungan wisatawan mancanegara di Indonesia, setelah mempertimbangkan efek variasi kalender Libur Nasional pada data yang mengandung *outlier* adalah Model 2 (0,1,1)(1,0,1)[12]. Peramalan menggunakan model ini menunjukkan bahwa jumlah kunjungan wisatawan mancanegara di Indonesia dipengaruhi oleh tiga faktor eksternal, yaitu Hari Raya Nyepi, Hari Raya Idul Fitri, dan libur akhir tahun, dengan koefisien sebesar 0.9329 (*ht*), 0.3791 (*lt*), dan 0.2735 (*vt*). Perubahan pada ketiga variabel tersebut memiliki kontribusi langsung terhadap jumlah kunjungan wisatawan, dengan pengaruh terbesar berasal dari Hari Raya Nyepi (*ht*). Akurasi model yang dianalisis juga tergolong tinggi, ditunjukkan oleh nilai MAPE sebesar 3.75%, yang menunjukkan bahwa model mampu memberikan estimasi dengan tingkat kesalahan relatif yang rendah.

Penelitian selanjutnya disarankan untuk dapat menggunakan data dengan cakupan waktu yang lebih panjang serta menambahkan lebih banyak variabel eksogen ke dalam model lainnya guna meningkatkan akurasi dan ketepatan model.

## Pernyataan Kontribusi Penulis (CRedit)

**Desya Neydi Putri Pakaya:** Konseptualisasi, Metodologi, Perangkat Lunak, Analisis Formal, Investigasi, Kurasi Data, Penulisan–Draf Awal, Penulisan–Telaah dan Penyuntingan, Visualisasi. **Novianita Achmad:** Konseptualisasi, Metodologi, Penulisan–Draf Awal, Penulisan–Telaah dan Penyuntingan, Supervisi. **Isran K. Hasan:** Konseptualisasi, Metodologi, Perangkat

Lunak, Validasi, Penulisan–Draf Awal, Penulisan–Telaah dan Penyuntingan, Supervisi. **Djihad Wungguli:** Metodologi, Penulisan–Draf Awal, Penulisan–Telaah dan Penyuntingan, Supervisi. **Siti Nurmaidia Abdussamad:** Metodologi, Penulisan–Draf Awal, Penulisan–Telaah dan Penyuntingan, Supervisi.

## Deklarasi Konflik Kepentingan

Penulis menyatakan tidak ada konflik kepentingan.

## Ketersediaan Data

Seluruh data beserta kode program yang digunakan dalam penelitian ini tersedia secara terbuka dan dapat diakses oleh umum. Data yang digunakan mencakup jumlah kunjungan wisatawan mancanegara di Indonesia yang bersumber dari situs resmi Badan Pusat Statistik (BPS) Indonesia. Dataset ini tidak memuat informasi bersifat pribadi. Dokumentasi lengkap dan tautan untuk mengakses data tersedia dalam repositori Badan Pusat Statistik (BPS) Indonesia.

## Daftar Pustaka

- [1] C. Aponno, “Kontribusi sektor pariwisata dan pertumbuhan ekonomi di provinsi maluku,” *Jurnal Ekonomi, Sosial & Humaniora*, vol. 2, no. 05, pp. 111–118, 2020. [Available online](#).
- [2] V. Reza, “Pariwisata halal dalam pengembangan ekonomi indonesia,” *Jurnal An-Nahl*, vol. 7, no. 2, pp. 106–112, 2020. DOI: [10.54576/annahl.v7i2.20](https://doi.org/10.54576/annahl.v7i2.20)
- [3] L. Santoso and P. Priyadi, “Mengoptimalkan proses pembersihan data dalam analisis big data menggunakan pipeline berbasis ai,” *Elkom: Jurnal Elektronika dan Komputer*, vol. 17, no. 2, pp. 657–666, 2024. DOI: [10.51903/elkom.v17i2.2311](https://doi.org/10.51903/elkom.v17i2.2311)
- [4] N. H. Latief, I. Setiawan, et al., “Peramalan curah hujan di kota makassar dengan menggunakan metode sarimax,” *Statistika*, vol. 22, no. 1, pp. 55–63, 2022. DOI: [10.29313/statistika.v22i1.990](https://doi.org/10.29313/statistika.v22i1.990)
- [5] R. Wahyudi, M. Kom, H. Nambo, et al., *Kombinasi Metode Sarima dengan PARCD untuk Prediksi Kasus Covid-19*. Zahira Media Publisher, 2021, vol. 1.
- [6] A. N. Rais, R. Rousyati, I. J. Thira, D. N. Kholidah, N. Purwati, and Y. M. Kristania, “Evaluasi metode forecasting pada data kunjungan wisatawan mancanegara ke indonesia,” *EVOLUSI: Jurnal Sains Dan Manajemen*, vol. 8, no. 2, 2020. DOI: [DOI:10.31294/evolusi.v8i2.8971](https://doi.org/10.31294/evolusi.v8i2.8971)
- [7] N. P. N. Hendayanti and M. Nurhidayati, “Perbandingan metode seasonal autoregressive integrated moving average (sarima) dengan support vector regression (svr) dalam memprediksi jumlah kunjungan wisatawan mancanegara ke bali,” *Jurnal Varian*, vol. 3, no. 2, pp. 149–162, 2020. DOI: [10.30812/varian.v3i2.668](https://doi.org/10.30812/varian.v3i2.668)
- [8] F. Andika, S. Dayanti, F. Dewi, et al., “Peramalan jumlah kunjungan wisatawan mancanegara ke provinsi aceh menggunakan fuzzy time series chen,” *JURNAL ILMIAH MATEMATIKA DAN TERAPAN*, vol. 20, no. 1, pp. 15–24, 2023. DOI: [10.22487/2540766X.2023.v20.i1.16168](https://doi.org/10.22487/2540766X.2023.v20.i1.16168)
- [9] P. R. Sihombing, S. Suryadiningrat, D. A. Sunarjo, and Y. P. A. C. Yuda, “Identifikasi data outlier (pencilan) dan kenormalan data pada data univariat serta alternatif penyelesaiannya,” *Jurnal Ekonomi Dan Statistik Indonesia*, vol. 2, no. 3, pp. 307–316, 2022. DOI: <http://dx.doi.org/10.11594/jesi.02.03.07>

- [10] M. Syukron, R. Santoso, and T. Widiharih, “Perbandingan metode smote random forest dan smote xgboost untuk klasifikasi tingkat penyakit hepatitis c pada imbalance class data,” *Jurnal Gaussian*, vol. 9, no. 3, pp. 227–236, 2020. DOI: [10.14710/j.gauss.9.3.227-236](https://doi.org/10.14710/j.gauss.9.3.227-236)
- [11] N. A. Lestari and A. M. A. Pabulo, “Pengaruh rasio solvabilitas, rasio likuiditas dan rasio profitabilitas terhadap nilai perusahaan,” *Jurnal Ilmiah Manajemen, Ekonomi, & Akuntansi (MEA)*, vol. 7, no. 3, pp. 1249–1266, 2023. DOI: [10.31955/mea.v7i3.3530](https://doi.org/10.31955/mea.v7i3.3530)
- [12] H. Annisa Khoiri, *Analisis deret waktu univariat*, 2023. [Available online](#).
- [13] M. I. Rizki and T. A. Taqiyuddin, “Penerapan model sarima untuk memprediksi tingkat inflasi di indonesia,” *Jurnal Sains Matematika dan Statistika*, vol. 7, no. 2, pp. 62–72, 2021. DOI: [10.24014/jsms.v7i2.13168](https://doi.org/10.24014/jsms.v7i2.13168)
- [14] Muryanto, “Pemodelan arima dan sarima untuk peramalan indeks harga konsumen kota balikpapan,” *BESTARI: Buletin Statisistik dan Aplikasi Terkini*, vol. 1, no. 2, pp. 11–28, 2021. [Available online](#).
- [15] H. Hayati, N. N. Debarajaya, and S. Martha, “Prediksi data jumlah penumpang kereta dengan efek variasi kalender pada model sarimax,” *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 10, no. 4, 2021. DOI: [10.26418/bbimst.v10i04.49536](https://doi.org/10.26418/bbimst.v10i04.49536)
- [16] Ď. Peter and P. Silvia, “Arima vs. arimax—which approach is better to analyze and forecast macroeconomic time series,” in *Proceedings of 30th international conference mathematical methods in economics*, vol. 2, 2019, pp. 136–140. [Available online](#).
- [17] S. Nurwahdania and W. Sulistijanti, “Prediksi produksi beras di provinsi jawa timur menggunakan musiman autoregressive terintegrasi moving average dengan metode exogenous input (sarimax),” *EDUSAINTEK*, vol. 4, 2020. [Available online](#).
- [18] I. Fadliani, I. Purnamasari, and W. Wasono, “Peramalan dengan metode sarima pada data inflasi dan identifikasi tipe outlier (studi kasus: Data inflasi indonesia tahun 2008-2014),” *Jurnal Statistika Universitas Muhammadiyah Semarang*, vol. 9, no. 2, pp. 109–116, 2021. DOI: [10.26714/jsunimus.9.2.2021.109-116](https://doi.org/10.26714/jsunimus.9.2.2021.109-116)
- [19] R. Agustin, “Peramalan data intermiten menggunakan metode autoregressive integrated moving average dan neural network (arima-nn),” *PhD diss., Institut Teknologi Sepuluh Nopember*, 2018. [Available online](#).
- [20] A. Dirgantara, A. L. Jinda, M. F. Rinanda, I. M. Nur, and F. Fauzi, “Peramalan inflasi di kota surakarta menggunakan model autoregressive integrated moving average (arima),” *EDUSAINTEK*, vol. 4, 2020. [Available online](#).
- [21] A. S. Kinanta, I. F. Amri, S. A. W. Sari, M. Al Haris, I. M. Sidqi, and M. F. Choirudin, “Predict of average humidity in central java clomatology station using the sarimax method (seasonal autoregressive integrated moving average with exogenous variables) with average temperature as the exogenous variable,” *PARAMETER: Jurnal Matematika, Statistika dan Terapannya*, vol. 3, no. 02, pp. 93–106, 2021. [Available online](#).