

# Seasonal Decomposition dan Transformasi Logaritma Sebagai Preprocessing Data Pada Prediksi Kunjungan Wisatawan Menggunakan Seasonal Auto-Regressive Integrated Moving average

Asih Saputri<sup>1</sup>, Eva Yulia Puspaningrum<sup>2\*</sup>, Rr. Ani Dijah Rahajoe<sup>3</sup>

<sup>1,3</sup> Informatika/Fakultas Ilmu Komputer/UPN Veteran Jawa Timur

[121081010053@student.upnjatim.ac.id](mailto:121081010053@student.upnjatim.ac.id)

[evapuspaningrum.if@upnjatim.ac.](mailto:evapuspaningrum.if@upnjatim.ac.)

[anidijah.if@upnjatim.ac.](mailto:anidijah.if@upnjatim.ac.)

**Abstrak**— Penelitian ini bertujuan untuk menganalisis peran *Seasonal Decomposition* dan transformasi logaritma sebagai tahapan *Preprocessing* data dalam meningkatkan kualitas prediksi kunjungan wisatawan menggunakan model *Seasonal Auto-Regressive Integrated Moving average* (SARIMA). Data yang digunakan berupa data bulanan kunjungan wisatawan Kabupaten Tegal periode 2015–2024 yang diperoleh dari laporan statistik pariwisata Jawa Tengah dan Dinas Pemuda, Olahraga, dan Pariwisata Kabupaten Tegal. Permasalahan utama pada data adalah adanya *missing value*, *varians* yang tidak stabil, serta ketidakstasioneran deret waktu yang dapat menurunkan akurasi model peramalan. Metode *Seasonal Decomposition* diterapkan untuk menangani *missing value* dengan mempertimbangkan pola musiman, sedangkan transformasi logaritma natural digunakan untuk menstabilkan *varians* data. Selanjutnya, uji stasioneritas dilakukan menggunakan uji *Augmented Dickey–Fuller* (ADF) pada data asli, data hasil *seasonal decomposition* dan data hasil transformasi logaritma. Hasil pengujian menunjukkan bahwa data asli belum memenuhi asumsi stasioneritas, terutama pada tingkat signifikansi 1%. Namun, setelah dilakukan transformasi logaritma natural, data menjadi stasioner pada tingkat signifikansi 5% dan 10%. Temuan ini menunjukkan bahwa kombinasi *seasonal decomposition* dan transformasi logaritma efektif sebagai tahap *preprocessing* untuk meningkatkan karakteristik stasioneritas data. Dengan demikian, penerapan *preprocessing* yang tepat sangat penting sebelum pemodelan SARIMA agar hasil estimasi dan prediksi kunjungan wisatawan menjadi lebih akurat dan dapat diandalkan sebagai dasar pengambilan keputusan dalam pengelolaan dan pengembangan sektor pariwisata daerah.

**Kata Kunci**— SARIMA, *seasonal decomposition*, transformasi logaritma, stasioneritas, kunjungan wisatawan

## I. PENDAHULUAN

Sektor pariwisata merupakan salah satu pendorong utama pertumbuhan ekonomi daerah karena mampu memberikan kontribusi terhadap pendapatan daerah, menciptakan lapangan kerja, serta mendorong perkembangan sektor pendukung lainnya [1]. Dalam upaya meningkatkan kualitas pengelolaan pariwisata, diperlukan perencanaan yang didasarkan pada informasi yang akurat mengenai jumlah kunjungan wisatawan di masa mendatang [2]. Oleh karena itu, peramalan jumlah

kunjungan wisatawan menjadi salah satu aspek penting yang dapat membantu pemerintah maupun pelaku industri pariwisata dalam menyusun strategi pengembangan destinasi, pengalokasian sumber daya, serta penyediaan fasilitas pendukung yang sesuai dengan kebutuhan wisatawan [3].

Data kunjungan wisatawan umumnya berbentuk deret waktu (*time series*) yang memiliki karakteristik tren dan pola musiman. Pola musiman dapat muncul akibat pengaruh hari libur nasional, musim liburan sekolah, maupun kegiatan atau acara tertentu yang diselenggarakan secara berkala. Salah satu metode yang banyak digunakan untuk memodelkan data deret waktu yang memiliki pola musiman adalah *Seasonal Auto-Regressive Integrated Moving average* (SARIMA) [4]. Model SARIMA mampu mengakomodasi hubungan antar data pada periode sebelumnya sekaligus mempertimbangkan pengaruh pola musiman yang berulang dalam interval waktu tertentu.

Meskipun demikian, keberhasilan model SARIMA sangat dipengaruhi oleh kualitas data yang digunakan. Dalam praktiknya, data kunjungan wisatawan sering kali menghadapi berbagai permasalahan, seperti adanya *missing value*, *varians* data yang tidak stabil, serta ketidakstasioneran data [5]. Keberadaan *missing value* dapat mengurangi representasi pola sebenarnya dalam data, sedangkan *varians* yang tidak stabil dapat menyebabkan model kesulitan dalam menangkap hubungan antarperiode secara konsisten. Selain itu, ketidakstasioneran data dapat menyebabkan parameter model menjadi tidak stabil sehingga menurunkan akurasi hasil peramalan [6].

Beberapa penelitian sebelumnya menunjukkan bahwa tahap *preprocessing* memiliki peran penting dalam meningkatkan kualitas data sebelum dilakukan pemodelan. Salah satu pendekatan yang dapat digunakan untuk menangani *missing value* pada data deret waktu adalah *seasonal decomposition* [7]. Metode ini memanfaatkan informasi tren dan pola musiman yang terkandung dalam data sehingga nilai yang hilang dapat diestimasi dengan mempertahankan karakteristik asli data. Dibandingkan metode imputasi sederhana seperti rata-rata atau interpolasi linear, *seasonal decomposition* lebih sesuai digunakan pada data yang memiliki pola musiman karena mampu mempertahankan struktur temporal yang ada.

Selain penanganan *missing value*, stabilisasi *varians* juga diperlukan untuk meningkatkan kualitas data. Salah satu teknik yang umum digunakan adalah transformasi logaritma natural. Transformasi ini mampu mengurangi perbedaan skala antar nilai data, menekan pengaruh nilai ekstrem, serta membuat distribusi data menjadi lebih homogen. Dengan *varians* yang lebih stabil, proses identifikasi dan estimasi parameter pada model SARIMA dapat dilakukan dengan lebih baik sehingga menghasilkan model yang lebih representatif terhadap pola data [8].

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan penggunaan *seasonal decomposition* dan transformasi logaritma sebagai tahapan *preprocessing* data sebelum pemodelan SARIMA pada prediksi kunjungan wisatawan. Penggunaan kedua metode ini diharapkan mampu mengatasi permasalahan *missing value* dan ketidakstabilan *varians* sehingga karakteristik data menjadi lebih mendekati kondisi stasioner yang dibutuhkan oleh model SARIMA. Selanjutnya, efektivitas tahapan *preprocessing* tersebut dievaluasi melalui pengujian stasioneritas menggunakan *Augmented Dickey-Fuller* (ADF) untuk mengetahui sejauh mana kualitas data dapat ditingkatkan sebelum digunakan dalam proses peramalan [9].

## II. METODE

Metode SARIMA (*Seasonal Auto-Regressive Integrated Moving average*) merupakan pengembangan dari model ARIMA yang dirancang khusus untuk menangani data deret waktu dengan pola musiman, seperti kunjungan wisatawan Indonesia yang sering fluktuatif akibat faktor libur nasional, musim liburan, atau *event* tahunan [10]. SARIMA dinotasikan sebagai SARIMA(p,d,q)(P,D,Q)s, di mana komponen non-musiman (p,d,q) menggambarkan autoregresi, integrasi (diferensiasi), dan *moving average*, sementara (P,D,Q)s menangkap pola musiman dengan periode s (misalnya, 12 untuk data bulanan) [11]. Dalam prediksi kunjungan wisatawan, SARIMA efektif karena mampu memodelkan tren jangka panjang sekaligus siklus musiman, seperti peningkatan kunjungan saat musim kemarau atau hari raya sehingga menghasilkan ramalan yang akurat dalam pengambilan keputusan atau kebijakan. Pentingnya stasioneritas data tidak bisa diabaikan karena stasioneritas berarti mean, *varians*, dan *kovarians* data konstan sepanjang waktu yang menjadi prasyarat utama model SARIMA agar parameter stabil dan prediksi *reliable*, menghindari *spurious regression*, atau kesalahan interpretasi pola acak sebagai tren nyata [12].

*Preprocessing* data menjadi kunci untuk mencapai stasioneritas tersebut terutama pada data kunjungan wisatawan yang sering mengandung *missing value* akibat pandemi atau kesalahan pengumpulan data [13]. Tahap awal adalah *seasonal decomposition* untuk menangani *missing value* pada komponen *seasonal* yang dapat diimputasi dengan rata-rata musiman dari periode serupa. Selanjutnya, transformasi logaritma natural diterapkan untuk menstabilkan *varians* pada data dengan tren eksponensial, seperti lonjakan kunjungan wisatawan yang proporsional meningkat, sehingga distribusi menjadi lebih simetris dan mendekati normal. Proses ini sering diikuti diferensiasi orde 1 atau musiman untuk menghilangkan tren

dan siklus. Pendekatan ini memastikan data stasioner dan tervalidasi via uji ADF sehingga model SARIMA dapat diidentifikasi optimal melalui ACF/PACF plot, dan menghasilkan *forecast* presisi tinggi untuk mendukung pengambilan keputusan di sektor pariwisata Indonesia [14].

## III. HASIL PEMBAHASAN

### A. Pengumpulan Data

Data yang digunakan adalah data kunjungan wisatawan Kabupaten Tegal yang bersumber dari Laporan Statistik Pariwisata Jawa Tengah mulai tahun 2015 hingga 2023 ditambah dengan data dari Dinas Disporapar Kabupaten Tegal untuk tahun 2024 [15].

### B. Konversi Tipe Data

Dalam dataset, kolom "bulan" pada awal masih berupa tipe data object, sehingga memerlukan konversi ke format datetime. Transformasi ini krusial karena model Seasonal ARIMA (SARIMA), sebagai pendekatan statistik untuk meramalkan data time series yang mengandung pola musiman dan tren, mensyaratkan input berupa datetime atau indeks waktu yang eksplisit. Kebutuhan tersebut timbul dari esensi SARIMA yang memproses data berdasarkan sekuens temporal serta siklus periodik.

### C. Penanganan Missing value

Data yang hilang atau tidak tersedia dalam dataset, yang dikenal sebagai *missing value*, sering kali muncul akibat kesalahan pengukuran, proses entri data, atau keterbatasan akses informasi. Pada dataset ini, kolom jumlah kunjungan wisatawan mencakup beberapa nilai berupa 0, NaN, maupun sel kosong. Untuk mengatasinya, diterapkan metode *seasonal decomposition* yang secara efektif mengisi *missing value* dengan memanfaatkan pola musiman inheren pada data deret waktu—pendekatan ini sangat sesuai untuk time series berpola musiman. Prosesnya melibatkan pengelompokan data berdasarkan siklus bulanan, diikuti perhitungan komponen tren melalui *moving average* (contohnya periode 3 bulan), sebelum nilai hasil tersebut diimputasikan ke posisi data yang hilang guna mempertahankan integritas temporal data.

TABEL I  
DATA KUNJUNGAN WISATAWAN

Bulan	Jumlah Kunjungan Wisatawan
2015-09-01	51.976
2016-09-01	64.012
2017-09-01	NaN
2018-09-01	98.721
2019-09-01	93.211
2020-09-01	43.815
2021-09-01	29.125
2022-09-01	10.697
2023-09-01	42.291

Bulan	Jumlah Kunjungan Wisatawan
2024-09-01	72.818

Pada rentang tersebut misalnya, terdapat nilai NaN atau *missing value* pada bulan September 2017. Untuk hal ini yang perlu dilakukan adalah menghitung nilai rata-rata dari bulan september terlebih dahulu.

$$\mu_{\text{Sept}} = \frac{51976 + 64012 + 98721 + 93211 + \dots + 64012}{11} = 56.296 \quad (1)$$

Dari nilai rata-rata ini kemudian dicari nilai komponen tren dengan menggunakan *Moving average* nilai rata-rata data bulan agustus dan oktober.

$$\text{Tren} = \frac{46.723 + \text{NaN} + 66}{3} = \frac{46.723 + 66}{2} = 23.395 \quad (2)$$

Selanjutnya diterapkan konsep *Seasonal Decomposition* di mana Data Asli = Tren × Musiman × Residu dengan nilai residu dapat dianggap 1 jika tidak ada komponen residu yang cocok. Berdasarkan hal tersebut, nilai *missing value* dapat dilakukan pergantian dengan menggunakan perhitungan berikut:

$$\begin{aligned} \text{September 2017} &= \text{Tren} \times \frac{\text{Rata-rata September}}{\text{Rata-rata seluruh data}} \times \text{Residu} \\ &= 23.395 \times \frac{56.296}{68.742} \times 1 = 23.395 \times 0,82 \times 1 = 19.184 \quad (3) \end{aligned}$$

Nilai 19.184 didapat untuk menggantikan nilai *missing value* pada bulan September 2017 berdasarkan pola tren dan musimannya. Proses ini dilakukan terus hingga seluruh nilai *missing value* berhasil diisi dengan data baru.

#### D. Transformasi Data

Sebelum melakukan transformasi data, kita perlu melihat distribusi data melalui statistik deskriptif apakah telah seragam atau belum.

Statistik	Nilai
Count	120
Mean	60039,515
Standar Deviasi	34094,489
Min	66
25%	36901,454
50%	58459,000
75%	76791,000
Max	159443,000

Data kunjungan wisatawan menunjukkan disparitas skala yang mencolok, terlihat dari jarak minimum ke maksimum yang mencapai lebih dari 2.400 kali lipat, mengindikasikan ketidakseimbangan ekstrem antar nilai. Standar deviasi yang besar, melebihi rata-rata sebesar 34.094, semakin menegaskan variabilitas tinggi dan distribusi data yang tidak homogen. Walaupun kuartil menggambarkan mayoritas data terkonsentrasi antara Q1 dan Q3, keberadaan outlier ekstrem tetap menjadikan data tidak seragam secara keseluruhan, sehingga transformasi atau standarisasi data menjadi langkah esensial.

Transformasi data merupakan proses mengonversi skala, distribusi, maupun struktur data agar lebih optimal untuk analisis, guna meningkatkan kualitas masukan model dan menghasilkan prediksi yang akurat. Dalam konteks ini, transformasi logaritma natural (ln) diterapkan, di mana jika  $y$  adalah data asli maka  $y = \ln(x)$ . Logaritma natural didefinisikan sebagai logaritma berbasis  $e$ , dengan nilai konstan

$e \approx 2,71828$ . Sebagai ilustrasi, transformasi untuk data Januari 2015 dapat dihitung dengan rumus tersebut.

$$\ln(\text{Januari 2015}) = \ln(68505) \approx 11.134676 \quad (4)$$

TABEL II  
DATA SETELAH TRANSFORMASI

Bulan	Data Asli	Data setelah Transformasi
2015-01-01	68.505	11,13466201
2015-02-01	26.986	10,20307349
2015-03-01	34.914	10,46064317
2015-04-01	35.996	10,49116310
2015-05-01	61.241	11,02257218
2015-06-01	43.426	10,67881362
2015-07-01	202.420	12,21810003
2015-08-01	62.103	11,03654958
2015-09-01	51.976	10,85853735
2015-10-01	54.512	10,90617614
...	...	...
2024-11-01	50.838	10,79673481
2024-12-01	112.733	11,11623178

Dari hasil transformasi ini dapat dilihat bahwa data sudah seragam pada rentang 10,5 sampai 12 di mana rentang nilai dianggap sempit pada distribusi data yang sebagian besar.

#### E. Uji Stasioneritas

Pada dataset kemudian akan dilakukan uji stasioneritas baik pada data sebelum dilakukan *preprocessing* maupun setelah seluruh proses *Preprocessing*. Untuk menguji stasioneritas data tanpa intercept atau dengan regresi sederhana, kita dapat menggunakan rumus perhitungan untuk mencari koefisien  $\gamma$ :

$$\gamma = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad (4)$$

Di mana  $\gamma$  nantinya digunakan untuk menghitung nilai  $t$ -statistic yang menggunakan nilai  $\gamma$  dan standar error pada rumus berikut  $t = \gamma / SE$ . Nilai Standard Error (SE) sendiri didapat dari perhitungan nilai residual dengan lag 1 data ke- $n$  pada deret waktu seperti pada rumus berikut:

$$SE = \sqrt{\frac{\sum e_i^2}{n-1} + \sum (x_i - \bar{x})^2} \quad (5)$$

Hasil dari perhitungan nilai  $t$ -statistic atau nilai kritis lalu dicocokkan dengan tabel panduan nilai kritis ADF Tester untuk menentukan stasioneritas data. Berdasarkan perhitungan yang dilakukan, karena tidak adanya intercept atau tren yang termasuk di dalamnya, maka model yang digunakan akan mengacu pada Model 2 dengan ketentuan sebagai berikut:

TABEL III  
DATA PANDUAN ADF TESTER

N	1%	5%	10%
25	-4.375	-3.943	-3.589
50	-4.152	-3.791	-3.495
100	-4.052	-3.722	-3.452
250	-3.995	-3.683	-3.427
500	-3.977	-3.670	-3.419
>500	-3.963	-3.660	-3.413

Dari perhitungan yang telah dilakukan diperoleh hasil sebagai berikut:

TABEL IV  
DATA HASIL UJI STATIONERITAS

Nilai	ADF Tester	p-value	Hasil
Data Asli	-3,1924	-0,0859	Tidak Stasioner
<i>Seasonal Decomposition</i>	-8,2663	-0,0000	Stasioner
Logaritma Natural	-6,5498	-0,0000	Stasioner

## IV. KESIMPULAN

Berdasarkan hasil pengujian stasioneritas menggunakan uji Augmented Dickey–Fuller (ADF), dapat disimpulkan bahwa karakteristik stasioneritas data kunjungan wisatawan sangat dipengaruhi oleh bentuk transformasi data yang digunakan. Pengujian dilakukan pada tiga tingkat signifikansi, yaitu 1%, 5%, dan 10%, dengan membandingkan nilai statistik ADF dari data asli, hasil *seasonal decomposition*, serta data yang telah ditransformasi menggunakan logaritma natural terhadap nilai kritis masing-masing tingkat signifikansi.

Pada tingkat signifikansi 1%, seluruh bentuk data yang diuji, baik data asli, data hasil *seasonal decomposition*, maupun data logaritma natural, menunjukkan nilai statistik ADF yang lebih besar dibandingkan nilai kritis. Hal ini mengindikasikan bahwa pada tingkat kepercayaan yang sangat ketat, data belum memenuhi kriteria stasioneritas dan masih mengandung unit root. Dengan demikian, hipotesis nol yang menyatakan bahwa data tidak stasioner belum dapat ditolak pada tingkat signifikansi 1%.

Namun, pada tingkat signifikansi 5% dan 10%, hasil yang diperoleh menunjukkan perbedaan yang signifikan. Nilai statistik ADF untuk data logaritma natural berada di bawah nilai kritis pada kedua tingkat signifikansi tersebut, sehingga hipotesis nol dapat ditolak. Hal ini menandakan bahwa setelah dilakukan transformasi logaritma natural, data dapat dianggap stasioner pada tingkat signifikansi 5% dan 10%. Kondisi serupa juga terlihat pada data hasil *seasonal decomposition*, yang menunjukkan kecenderungan stasioner pada tingkat signifikansi yang lebih longgar.

## V. KESIMPULAN

Secara keseluruhan, hasil pengujian ini menunjukkan bahwa data kunjungan wisatawan pada bentuk aslinya belum bersifat stasioner, terutama pada tingkat signifikansi yang ketat. Transformasi logaritma natural terbukti efektif dalam meningkatkan kestabilan *varians* dan mengurangi pengaruh tren, sehingga memungkinkan data memenuhi asumsi stasioneritas pada tingkat signifikansi 5% dan 10%. Temuan ini mengindikasikan bahwa transformasi data merupakan langkah penting sebelum penerapan model peramalan time series, seperti SARIMA agar hasil estimasi dan prediksi yang dihasilkan lebih akurat dan dapat diandalkan.

## REFERENSI

- [1] Disporapar Kab. Tegal, "Rancangan Akhir Rencana Kerja," Kab. Tegal, 2025.
- [2] A. R. Pamungkas, "Smart Tourism sebagai pengungkit dalam pemulihan pariwisata pasca pandemi," 18 Oktober 2023. [Online]. Available: [https://tegalkab.go.id/news/view/artikel/smart\\_torism\\_sebagai\\_pengu ngkit\\_dalam\\_pemulihan\\_pariwisata\\_pasca\\_pandemi\\_2023110909455\\_4](https://tegalkab.go.id/news/view/artikel/smart_torism_sebagai_pengu_ngkit_dalam_pemulihan_pariwisata_pasca_pandemi_2023110909455_4).
- [3] W. Alwi, K. Nurfadilah and Munira, "Penerapan Metode SARIMA untuk Peramalan Jumlah Pengunjung Wisata Taman Nasional Bantimurung Bulusaraung Maros," *Journal of Mathematics: Theory and Applications*, pp. 1-7, 2021.
- [4] A. N. Azizah, Penerapan Hybrid Seasonal Autoregressive Integrated Moving Average – Long Short Term Memory (Sarima-Lstm) dalam Meramalkan Curah Hujan di Bogor, Universitas Pendidikan Indonesia, 2023.
- [5] M. S. U. Buku Ajar, "Memahami Konsep Stasioneritas Data Deret Waktu (Time Series)," [Online]. Available: <https://www.mobilestatistik.com/stasioneritas-data-deret-waktu/>.
- [6] N. Lestari and N. Wahyuningsih, "Peramalan Kunjungan Wisata dengan Pendekatan Model SARIMA (Studi kasus : Kusuma Agrowisata)," *JURNAL SAINS DAN SENI ITS Vol. 1, No. 1*, pp. 29-33.
- [7] Z. R. Firdhani, Penerapan Metode Seasonally Decomposed Missing Value Imputation pada Pemodelan Hybrid Machine Learning untuk Peramalan Kualitas Udara di Kota Surabaya, Surabaya: Institut Teknologi Sepuluh Nopember, 2023.
- [8] F. Mu'minin, Fauziah and A. Gunaryati, "Prediksi Kunjungan Wisatawan Mancanegara Melalui Pintu Udara Menggunakan ARIMA, Glnnet, dan Prophet," *Techno.COM*, pp. 149-156, 1 Februari 2022.
- [9] Urfah, "Deteksi Stasioneritas Dalam Time Series," 2023. [Online]. Available: <https://www.datalearns247.com/deteksi-stasioneritas-dalam-time-series-201>.
- [10] N. P. Riau, Analisis Deret Waktu Musiman dengan Memperhatikan Variasi Kalender Pada Jumlah Wisatawan Domestik di Bali Menggunakan Model Hybrid Jaringan Saraf Tiruan, Padang: FMIPA - Universitas Andalas, 2025.
- [11] A. K. Dubey, A. Kumar, V. García-Díaz and A. K. Sharma, "Study and Analysis of SARIMA and LSTM in Forecasting Time Series Data," *Sustainable Energy Technologies and Assessments* 47, 2021.
- [12] 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, April 2020.
- [13] G. Christie, D. Hatidja and R. Tumilaar, "Penerapan Metode SARIMA dalam Model Intervensi Fungsi Step untuk Memprediksi Jumlah Pegunjung Objek Wisata Londa," *Jurnal Ilmiah Sains*, pp. 96-103, Oktober 2022.
- [14] G. Chhabra, "Comparison of Imputation Methods for Univariate Time Series," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. Volumen: 11, no. Issue: 2s, pp. 286-292, 28 January 2023.
- [15] B. Jateng, "BUKU STATISTIK PARIWISATA JAWA TENGAH 2024," Disbudparekraf Jawa Tengah, Semarang, 2024.