

Prediksi Jumlah Wisatawan Mancanegara Ke Indonesia Berdasarkan Kebangsaan Menggunakan SARIMA

Surya Halim ¹⁾ Bagus Mulyawan ²⁾ Manatap Dolok Lauro ³⁾

^{1) 2) 3)} Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jl. Letjen S. Parman No.1, Jakarta 11440 Indonesia
email: surya.535210020@stu.untar.ac.id ¹⁾, bagusm@fti.untar.ac.id ²⁾, manataps@fti.untar.ac.id ³⁾

ABSTRACT

Penelitian ini bertujuan untuk memprediksi jumlah wisatawan mancanegara ke Indonesia berdasarkan kebangsaan menggunakan model *Seasonal Autoregressive Integrated Moving Average (SARIMA)*. Data yang digunakan adalah kunjungan wisatawan dari berbagai negara ke beberapa kota utama di Indonesia. Hasil analisis menunjukkan bahwa model SARIMA (1,1,1)(1,1,1,12) memberikan performa terbaik untuk prediksi jumlah wisatawan dengan pola musiman yang konsisten. Model ini menunjukkan akurasi yang tinggi, terutama pada beberapa pasangan kota dan kebangsaan tertentu. Tiga pasangan dengan nilai Mean Absolute Error (MAE) terkecil adalah wisatawan asal Vietnam di Medan dengan MAE sebesar 85,24, wisatawan Australia di Surabaya dengan MAE 756,971, dan wisatawan Filipina di Yogyakarta dengan MAE 411,736. Hasil ini menunjukkan bahwa model SARIMA (1,1,1)(1,1,1,12) dapat menjadi pilihan yang efektif untuk memprediksi jumlah kunjungan wisatawan mancanegara di Indonesia, khususnya untuk kebangsaan dan kota yang memiliki pola kunjungan yang stabil.

Key words

SARIMA, prediksi, kebangsaan, MAE, Indonesia

1. Pendahuluan

Indonesia memiliki potensi dan sumber daya alam yang melimpah, namun belum sepenuhnya dimanfaatkan, terutama di sektor pariwisata [1]. Pengembangan sektor ini diharapkan dapat memberikan dampak positif bagi masyarakat, mengingat pariwisata adalah salah satu sektor nonmigas dengan pertumbuhan pesat yang berpotensi mendukung perekonomian nasional [2]. Keanekaragaman potensi dan kekayaan alam Indonesia memberikan peluang besar untuk pengembangan industri pariwisata [3].

Salah satu masalah utama yang dihadapi sektor pariwisata Indonesia adalah tingginya ketergantungan pada wisatawan dari negara-negara tertentu, seperti Tiongkok, Australia, dan Singapura. Ketergantungan ini menimbulkan risiko yang signifikan bagi industri pariwisata Indonesia karena adanya fluktuasi dalam jumlah kunjungan wisatawan dari negara-negara tersebut [4]. Jika terjadi perubahan kondisi ekonomi atau politik di

negara asal wisatawan, seperti resesi, ketidakstabilan politik, atau perubahan kebijakan perjalanan, dampaknya dapat langsung dirasakan oleh sektor pariwisata Indonesia [5].

Namun, pandemi COVID-19 yang mulai melanda pada awal 2020 telah memberikan dampak yang sangat besar terhadap sektor pariwisata global, termasuk di Indonesia. Pemerintah Indonesia memberlakukan pembatasan perjalanan internasional untuk menekan penyebaran virus, yang mengakibatkan penurunan drastis jumlah wisatawan mancanegara [6]. Menurut data yang dipublikasikan oleh Badan Pusat Statistik (BPS), pada awal tahun 2020 terjadi penurunan jumlah wisatawan mancanegara lebih dari 7,62 persen apabila dibandingkan jumlah kunjungan wisatawan mancanegara bulan Desember 2019 sebanyak 1,37 juta [7].

Selain penurunan jumlah wisatawan, pandemi juga telah mengubah perilaku wisatawan secara signifikan. Saat ini, wisatawan lebih memperhatikan faktor keamanan, kebersihan, dan kesehatan dalam memilih destinasi wisata. Hal ini mendorong industri pariwisata untuk beradaptasi dengan tren baru dan mengembangkan strategi yang lebih inovatif untuk menarik kembali wisatawan mancanegara [8].

Untuk mengatasi permasalahan ini, diperlukan solusi yang tepat dalam bentuk pemanfaatan data dan teknologi. Salah satu pendekatan yang dapat diimplementasikan adalah penggunaan model *Seasonal Autoregressive Integrate Moving Average (SARIMA)* untuk memprediksi jumlah wisatawan mancanegara. Model ini memungkinkan pemerintah dan industri pariwisata untuk memahami tren kunjungan wisatawan, mengantisipasi perubahan, dan merencanakan strategi yang lebih efektif [9].

“Comparison of Local Polynomial Regression and ARIMA in Predicting the Number of Foreign Tourist Visits to Indonesia”. Penelitian ini menggunakan dua metode, yaitu Local Polynomial Regression dan ARIMA, untuk memprediksi jumlah wisatawan mancanegara yang mengunjungi Indonesia [10]. Data yang digunakan mencakup periode Januari 2017 hingga Desember 2022, dan diperoleh dari Kementerian Pariwisata. Hasil penelitian menunjukkan bahwa ARIMA tidak memenuhi asumsi homoskedastisitas, sehingga hanya metode regresi polinomial lokal yang dianalisis lebih lanjut. Tingkat error yang dihasilkan menggunakan Local Polynomial

Regression diukur dengan nilai MAPE sebesar 1,43%, yang menunjukkan akurasi prediksi tinggi.

“Prediksi Kedatangan Wisatawan Mancanegara di Bali Menggunakan Algoritma Support Vector Machine dan Regresi Linear”. Penelitian ini bertujuan untuk memprediksi kedatangan wisatawan mancanegara ke Bali menggunakan algoritma Support Vector Machine (SVM) dan Regresi Linear [11]. Data yang digunakan mencakup statistik kunjungan wisatawan ke Bali, dan model prediksi dianalisis dengan menggunakan metode pembelajaran mesin. Penelitian ini menemukan bahwa metode SVM memberikan akurasi yang lebih baik dalam menangani data kompleks dibandingkan regresi linear.

“Can Google Trends (GT) Be Used to Predict Tourist Arrivals?: FB Prophet Machine Learning for Predicting Tourist Arrivals”. Penelitian ini memanfaatkan data dari Google Trends dan model FB Prophet untuk memprediksi kedatangan wisatawan [12]. Data diambil dari pencarian terkait pariwisata di Yogyakarta dari 2013 hingga 2023, yang kemudian dibandingkan dengan data aktual dari Badan Pusat Statistik. Hasil penelitian menunjukkan bahwa penggunaan kolom tambahan dalam dataset, seperti kolom *max*, dapat meningkatkan nilai akurasi (MAPE dan R2). Namun, penggunaan data Google Trends saja cenderung menghasilkan prediksi yang tidak stabil. Penelitian ini merekomendasikan kombinasi dataset untuk memperbaiki prediksi, terutama untuk periode waktu di bawah 200 hari.

“Backpropagation Algorithm Modeling to Predict the Number of Foreign Tourist Visits to Indonesia via Air Gates”. Penelitian ini menggunakan algoritma Backpropagation untuk memprediksi jumlah kunjungan wisatawan asing melalui pintu udara di Indonesia [13]. Data yang digunakan berasal dari Badan Pusat Statistik (BPS), dengan fokus pada kunjungan di Ngurah Rai Airport, Bali. Hasil prediksi menunjukkan tingkat akurasi sebesar 91,40% dengan nilai *Mean Squared Error* (MSE) sebesar 0,056. Algoritma Backpropagation terbukti cukup akurat dalam memprediksi jumlah kunjungan wisatawan, terutama di bandara utama Indonesia.

Dengan demikian, prediksi jumlah wisatawan berdasarkan kebangsaan dan destinasi ini sangat penting untuk mengoptimalkan strategi pemasaran dan promosi yang disesuaikan dengan tren dan preferensi wisatawan yang terus berubah. Implementasi sistem berbasis web ini tidak hanya berfungsi sebagai alat prediksi, tetapi juga sebagai panduan bagi para pemangku kepentingan untuk mengadaptasi strategi sesuai dengan perkembangan pasar wisatawan global.

2. Metode Penelitian

2.1 ADF (Augmented Dickey-Fuller)

Uji *Augmented Dickey-Fuller* (ADF) adalah salah satu metode statistik yang digunakan untuk menguji keberadaan unit root dalam data runtun waktu. Uji ini bertujuan untuk menentukan apakah data bersifat

stasioner, yaitu apakah karakteristik statistik seperti mean, varians, dan autokorelasi bersifat konstan sepanjang waktu. Keberadaan unit root mengindikasikan bahwa data bersifat non-stasioner, yang dapat memengaruhi validitas model runtun waktu, seperti SARIMA, yang mensyaratkan data stasioner untuk menghasilkan estimasi yang akurat dan stabil [14].

2.2 AutoCorrelation Function (ACF)

Autocorrelation adalah korelasi dari suatu deret waktu dengan nilai masa lalunya dan masa depannya sendiri. *AutoCorrelation Function* (ACF) mengukur korelasi antara nilai suatu deret waktu dengan nilai-nilai sebelumnya yang dipisahkan oleh lag tertentu. ACF digunakan untuk melihat orde *Moving Average* (MA) yaitu orde *q* dan orde *Q* untuk musiman. Berikut ini merupakan rumus dari *AutoCorrelation Function* (ACF) [15]:

$$P_k = \frac{\sum_{t=1}^{n-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^{n-1} (Z_t - \bar{Z})^2} \quad (1)$$

Dalam analisis deret waktu, P_k adalah koefisien autokorelasi pada lag ke- k , Z_t adalah nilai observasi pada waktu ke- t . Sedangkan \bar{Z} adalah rata-rata semua observasi dilambangkan sebagai n adalah total jumlah data.

2.3 Partial AutoCorrelation Function (PACF)

PACF atau *Partial Autocorrelation Function* mengukur korelasi parsial antara nilai deret waktu dengan nilai pada lag tertentu, setelah mengendalikan pengaruh lag-lag lainnya. PACF sering digunakan untuk menentukan orde model AR dalam analisis deret waktu. Pada grafik PACF, dapat mengidentifikasi lag di mana korelasi parsial berhenti signifikan, yang menunjukkan nilai p dalam model AR(p). Berikut ini merupakan rumus dari PACF.

$$\phi_{k,k} = p_k - \sum_{j=1}^{k-1} \phi_{k,j} p_{k+1-j} \quad (2)$$

Dalam analisis deret waktu, $\phi_{k,k}$ adalah ACF pada lag k , yaitu korelasi parsial yang diukur, p_k adalah Autokorelasi pada lag k . Sedangkan, $\sum_{j=1}^{k-1} \phi_{k,j} p_{k+1-j}$ adalah mengoreksi efek dari lag yang lebih pendek agar hanya tersisa korelasi parsial.

2.4 Autoregressive (AR)

Autoregressive (AR) adalah salah satu model dalam analisis deret waktu yang menyatakan bahwa nilai masa depan dari suatu variabel bergantung pada nilai-nilai masa lalu dari variabel tersebut. Dalam model ini, data masa lalu digunakan untuk memprediksi data masa depan berdasarkan hubungan linier antara lag data sebelumnya [16]. Model AR dinotasikan sebagai AR(p), di mana p adalah orde dari model atau jumlah lag yang digunakan. Persamaannya secara umum adalah sebagai berikut:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t \quad (3)$$

Dalam model AR, Y_t adalah nilai data saat ini, c adalah konstanta, $\phi_1, \phi_2, \dots, \phi_p$ adalah koefisien *autoregressive*, ϵ_t adalah *noise* atau *error*.

2.5 Moving Average (MA)

Moving Average (MA) merupakan model deret waktu yang memperkirakan nilai masa depan berdasarkan kesalahan prediksi (error) di masa lalu. Dalam model ini, kesalahan yang terjadi pada titik waktu sebelumnya digunakan untuk memprediksi nilai di masa kini. *Moving Average* (MA) dinotasikan sebagai MA(q), di mana q adalah orde dari model atau jumlah error lag yang digunakan [16]. Persamaan umumnya adalah:

$$Y_t = c + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (4)$$

Dalam model MA, Y_t adalah nilai data saat ini, c adalah konstanta, $\theta_1, \theta_2, \dots, \theta_q$ adalah koefisien *Moving Average*, ϵ_t adalah *noise* atau *error*. $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}$ adalah *error* di masa lalu.

2.6 Autoregressive Moving Average (ARMA)

Autoregressive Moving Average (ARMA) menggabungkan dua konsep, yaitu *Autoregressive* (AR) dan *Moving Average* (MA) untuk memprediksi nilai deret waktu. Model ini memperhitungkan hubungan antara nilai sebelumnya (komponen AR) dan kesalahan prediksi masa lalu (komponen MA) [16].

$$Y_t = c + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (5)$$

Dalam model ARMA, Y_t adalah nilai pengamatan pada waktu ke- t , ϵ_t adalah suatu proses *white noise* atau galat pada waktu ke- t , ϕ_0 adalah konstanta rata-rata, ϕ_i adalah koefisien AR non-musiman, $i = 1, 2, 3, \dots, p$. Sedangkan, θ_i adalah koefisien MA non-musiman, $i = 1, 2, 3, \dots, q$, p adalah orde AR non-musiman, q adalah orde MA non-musiman.

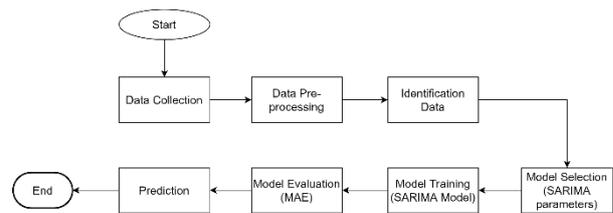
2.7 Seasonal Autoregressive Integrated Moving Average (SARIMA)

Model SARIMA adalah perkembangan dari model ARIMA yang lebih kuat, dan biasanya dinyatakan sebagai SARIMA(p,d,q)(P,D,Q)m. Model ini memungkinkan penggunaan variabel eksternal, dan ketika mempertimbangkan faktor eksternal, model peramalan disebut SARIMAX (p,d,q)(P,D,Q)m, di mana X mewakili faktor eksternal. Berbeda dengan model ARIMA, model SARIMA menambahkan empat parameter tambahan untuk menggambarkan aspek musiman dari deret waktu. Parameter p, q, dan d menunjukkan urutan *autoregressive*, urutan *differencing*, dan urutan *moving*

average untuk komponen tren, sementara P, D, dan Q menggambarkan urutan untuk komponen musiman. Untuk alur dari model SARIMA dapat dilihat pada Gambar 2.1. Parameter m menunjukkan jumlah langkah waktu dalam satu periode musiman [17]. Berikut ini merupakan rumus SARIMA secara umum:

$$\begin{aligned} \Phi_P(B^s)\Phi_p(B)(1-B^2)^d(1-B)^d X_t \\ = \theta_Q(B^s)\theta_q(B)Z_t \end{aligned} \quad (6)$$

$\Phi_P(B^s)$ dan $\Phi_p(B)$ adalah bagian *autoregressive* musiman dan non-musiman, yang menangkap ketergantungan musiman dan non-musiman dalam data. $(1-B^2)^d$ dan $(1-B)^d$ digunakan untuk menghilangkan tren musiman dan non-musiman. Sedangkan, $\theta_Q(B^s)$ dan $\theta_q(B)$ adalah komponen *moving average* musiman dan non-musiman, yang menangkap error musiman dan non-musiman. Z_t adalah error acak yang berdistribusi normal.



Gambar 1. Alur Model SARIMA

Dalam analisis deret waktu, model *Autoregressive* (AR) dan *Moving Average* (MA) merupakan dua komponen fundamental yang digunakan untuk menangkap pola data yang saling berkaitan dari waktu ke waktu. Kombinasi dari kedua model ini dikenal sebagai ARMA, yang memberikan fleksibilitas lebih dalam menangkap pola deret waktu yang lebih kompleks. Komponen-komponen ini juga menjadi dasar pengembangan model Seasonal ARIMA (SARIMA), yang dirancang untuk mengakomodasi pola musiman dalam deret waktu.

2.8 Mean Absolute Error (MAE)

MAE atau *Mean Absolute Error* merupakan sebuah evaluasi model yang digunakan untuk mengukur akurasi model prediksi. MAE menghitung rata-rata perbedaan absolut antara nilai prediksi dan nilai sebenarnya. Rumus dari MAE yaitu sebagai berikut [29].

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

3. Hasil Percobaan

Proses pengumpulan data diawali dengan mengakses laporan dan publikasi resmi melalui portal Kementerian Pariwisata. Setelah data dikumpulkan, dilakukan penyaringan untuk memastikan informasi yang relevan, seperti data dari negara-negara utama, yaitu Malaysia, Filipina, Singapura, Thailand, Vietnam, Tiongkok, India, dan Australia, dalam rentang waktu 2021 hingga 2024.

Uji *Augmented Dickey-Fuller* (ADF) dilakukan untuk menganalisis sifat stasioneritas pada data wisatawan mancanegara di berbagai kota di Indonesia. Stasioneritas merupakan salah satu asumsi penting dalam analisis deret waktu, terutama untuk memastikan bahwa data tidak memiliki pola tren atau musiman yang signifikan yang dapat memengaruhi hasil model prediktif. Melalui uji ADF ini, peneliti dapat menentukan apakah data sudah memenuhi kriteria stasioneritas atau memerlukan transformasi lebih lanjut, seperti differencing atau log transformation. Hasil uji ADF akan membantu dalam mempersiapkan data yang sesuai untuk pemodelan SARIMA, yang membutuhkan data yang stasioner untuk menghasilkan prediksi yang akurat.

Tabel 1. Uji Augmented Dickey-Fuller

Kota	Kebangsaan	ADF	p-value
DKI Jakarta	Malaysia	-0,154	0,944
	Filipina	-0,638	0,862
	Singapura	-0,458	0,9

	Thailand	-0,728	0,839
Bali	Malaysia	-0,311	0,924
	Filipina	-0,31	0,924
	Singapura	-1,493	0,537

	Thailand	-1,149	0,695
Surabaya	Malaysia	0,211	0,973
	Filipina	-0,074	0,952
	Singapura	-0,35	0,918

	Thailand	0,033	0,961
Medan	Malaysia	-0,677	0,853
	Filipina	-3,808	0,003
	Singapura	-1,291	0,633

	Thailand	-1,281	0,638
Yogyakarta	Malaysia	-1,022	0,745
	Filipina	-1,015	0,748
	Singapura	-2,871	0,049

	Thailand	-0,959	0,768

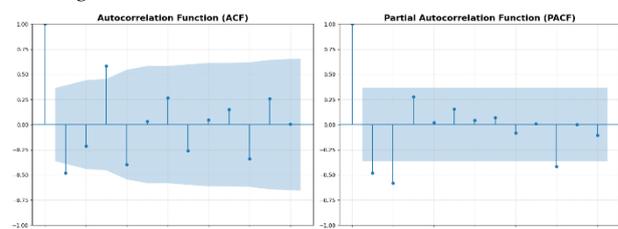
Batam	Malaysia	-0,687	0,85
	Filipina	-0,481	0,896
	Singapura	-1,53	0,519

	Thailand	-0,765	0,829

Hasil uji ADF menunjukkan bahwa sebagian besar seri data jumlah wisatawan di kota-kota Indonesia dari berbagai kebangsaan tidak stasioner. Data wisatawan dari negara tertentu di beberapa kota memang menunjukkan stasioneritas, seperti wisatawan dari Tiongkok di Surabaya dan Yogyakarta serta wisatawan dari Filipina di Medan. Namun, sebagian besar data menunjukkan adanya tren atau pola musiman, yang dapat mengindikasikan perubahan jumlah wisatawan sesuai dengan waktu atau musim tertentu.

Dalam *preprocessing data* untuk model *time series* seperti SARIMA, *differencing* digunakan untuk mengubah data menjadi stasioner, yang merupakan syarat penting untuk akurasi model. Proses *differencing* meliputi dua tahap, yaitu differencing non-musiman (*first differencing*) dan differencing musiman. Pertama, differencing non-musiman dilakukan dengan mengurangi nilai data pada suatu waktu dengan nilai data sebelumnya, sehingga tren linear dalam data dihilangkan. Kemudian, differencing musiman diterapkan dengan lag musiman 12 bulan (untuk data bulanan) guna menghilangkan pola musiman tahunan yang berulang.

Dalam analisis deret waktu, plot Autocorrelation Function (ACF) dan Partial Autocorrelation Function (PACF) merupakan alat penting untuk memahami pola autokorelasi dalam data. Kedua plot ini membantu dalam menentukan model yang sesuai dengan data, khususnya untuk model SARIMA yang memerlukan data yang stasioner. Plot ACF memberikan gambaran tentang kekuatan hubungan antara data pada waktu tertentu dengan nilai pada waktu sebelumnya, sehingga dapat membantu mengidentifikasi adanya pola musiman atau komponen *moving average* dalam data. Sementara itu, plot PACF menunjukkan korelasi parsial pada setiap lag dengan mengeluarkan pengaruh dari lag antara, yang berguna untuk menentukan adanya komponen *autoregressive*.



Gambar 2. Plot ACF dan PACF

Hasil analisis *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) untuk data jumlah wisatawan asal Malaysia di DKI Jakarta memberikan gambaran mengenai pola dan hubungan antar waktu dalam data tersebut. Plot ACF menunjukkan

bahwa terdapat korelasi yang cukup tinggi pada beberapa lag awal, terutama pada lag 1 hingga sekitar lag 3. Hal ini menunjukkan adanya keterkaitan yang cukup kuat antara jumlah wisatawan pada suatu bulan dengan jumlah wisatawan di bulan-bulan sebelumnya. Pola yang menurun secara perlahan setelah beberapa lag mengindikasikan adanya unsur tren atau musiman dalam data ini, yang bisa menunjukkan adanya kecenderungan musiman dalam kunjungan wisatawan Malaysia ke DKI Jakarta.

Di sisi lain, plot PACF memperlihatkan nilai yang signifikan pada lag 1 dan menurun secara cepat pada lag-lag berikutnya. Hal ini dapat diartikan bahwa data ini menunjukkan pola autoregresif yang signifikan pada lag pertama, di mana jumlah wisatawan pada suatu bulan memiliki hubungan yang cukup kuat dengan bulan sebelumnya setelah mengontrol pengaruh dari lag lainnya.

Berdasarkan pola ini, dapat disimpulkan bahwa model SARIMA mungkin cocok untuk data wisatawan Malaysia di DKI Jakarta, dengan komponen AR pada lag awal. Pola ACF yang signifikan pada beberapa lag pertama, dan PACF yang menonjol pada lag 1, mengindikasikan bahwa data ini memiliki karakteristik yang mungkin bisa ditangkap oleh model SARIMA dengan parameter *autoregressive* (p) yang rendah dan mungkin juga komponen *moving average* (q) yang kecil.

Tabel 2. Perbandingan Hasil Evaluasi Model SARIMA

Parameter	Kota	Kebangsaan	MAE	
(1,1,1) (1,1,1,12)	DKI Jakarta	Malaysia	30492,67	
		Filipina	7986,452	
		Singapura	6539,078	
		Thailand	4260,443	
		Vietnam	737,326	
		
		Tiongkok	84842,19	
		Bali	Malaysia	8499,525
	Bali	Filipina	34279,8	
		Singapura	54865,6	
		Thailand	3679,678	
		Vietnam	91136,32	
		
		Tiongkok	67343,74	
		Surabaya	Malaysia	4835,463
			Filipina	1052,446
	Singapura		13543,08	
	Thailand		1577,96	
	Vietnam		968,78	
	
	Australia		756,971	
	Medan		Malaysia	5751,997

(2,1,1) (1,1,1,12)	Yogyakarta	Filipina	561,461	
		Singapura	5302,396	
		Thailand	1393,771	
		Vietnam	85,24	
		
		Australia	743,401	
		Batam	Malaysia	22678,86
			Filipina	411,736
	Singapura		58738,7	
	Thailand		1047,4	
	Vietnam		177,348	
	
	Australia		1806,638	
	DKI Jakarta		Malaysia	24908,51
		Filipina	10968,84	
		Singapura	523415,7	
		Thailand	1698,392	
		Vietnam	21496,76	
		
		Tiongkok	1631,611	
		Bali	Malaysia	102744,2
	Filipina		8173,863	
	Singapura		54798,37	
	Thailand		14124,15	
	Vietnam		3752,839	
	
	Tiongkok		76102,15	
	Surabaya		Malaysia	26038,42
Filipina		12941,95		
Singapura		40661,66		
Thailand		29505,17		
Vietnam		90177,78		
...		...		
Australia		501489		
Medan		Malaysia	14447,5	
	Filipina	1157,771		
	Singapura	15118,74		
	Thailand	325,955		
	Vietnam	905,644		
		
	Australia	873,627		
	Medan	Malaysia	3573,693	
Filipina		535,602		
Singapura		2499,573		
Thailand		793,212		
	Vietnam	218,558		

	Yogyakarta	
		Australia	2281,464	
		Malaysia	17807,55	
		Filipina	385,878	
		Singapura	40326,31	
		Thailand	1734,095	
		Vietnam	112,861	
		
		Australia	2686,212	
	Batam	Malaysia	34672,6	
		Filipina	17903,1	
		Singapura	473388,5	
		Thailand	1590,745	
		Vietnam	15516,9	
		
		Australia	2875,836	
	(1,1,1) (2,1,1,12)	DKI Jakarta	Malaysia	24028,6
			Filipina	6446,075
Singapura			5870,133	
Thailand			4036,457	
Vietnam			731,605	
...			...	
Bali		Australia	22333,91	
		Malaysia	8374,981	
		Filipina	33441,13	
		Singapura	55899,31	
		Thailand	3680,263	
		Vietnam	97914,96	
Surabaya		
		Australia	29986,31	
		Malaysia	4258,924	
		Filipina	1051,477	
		Singapura	13644,54	
		Thailand	1581,044	
Medan		Vietnam	969,068	
		
		Australia	756,892	
		Malaysia	4208,319	
		Filipina	561,452	
		Singapura	5211,499	
Yogyakarta	Thailand	1393,612		
	Vietnam	85,974		
		
Yogyakarta	Australia	741,106		
	Malaysia	22732,38		
		Filipina	411,652	

		Singapura	58630,15
		Thailand	1048,186
		Vietnam	177,325
	
		Australia	1806,639
		Malaysia	24820,5
	Batam	Filipina	10973,89
		Singapura	523428,4
		Thailand	1698,49
		Vietnam	21529,02
	
		Tiongkok	1681,823

Hasil analisis *Mean Absolute Error* (MAE) untuk model SARIMA pada tabel menunjukkan perbedaan kinerja model dalam memprediksi jumlah wisatawan dari berbagai negara di beberapa kota besar di Indonesia. Pengujian dilakukan dengan menggunakan tiga parameter model SARIMA yang berbeda, yaitu (1,1,1)(1,1,1,12), (2,1,1)(1,1,1,12), dan (1,1,1)(2,1,1,12). Setiap model dievaluasi berdasarkan kemampuan prediksinya terhadap data wisatawan dari beberapa negara, seperti Malaysia, Filipina, Singapura, Thailand, Vietnam, Tiongkok, India, dan Australia, di berbagai kota seperti DKI Jakarta, Bali, Surabaya, Medan, Yogyakarta, dan Batam. Nilai MAE dihitung dengan membandingkan prediksi model terhadap data aktual dalam periode 12 bulan terakhir, yang digunakan sebagai data uji.

Pada model pertama, yaitu SARIMA (1,1,1)(1,1,1,12) memberikan hasil yang cukup beragam dalam hal MAE di berbagai kota dan kebangsaan. Nilai MAE untuk kota-kota besar seperti DKI Jakarta dan Batam cenderung tinggi pada beberapa kebangsaan, terutama untuk Tiongkok di DKI Jakarta (MAE sekitar 84.842,19) dan Singapura di Batam (MAE sekitar 523.416,7). Namun, model ini juga menunjukkan hasil MAE yang rendah untuk beberapa kebangsaan dan kota lainnya. Misalnya, MAE terendah di Surabaya untuk Australia (MAE sekitar 756,97) dan Medan untuk Vietnam (MAE sekitar 85,24). Hal ini menunjukkan bahwa model ini cukup efektif dalam menangani beberapa kota dan kebangsaan tertentu tetapi kurang konsisten dalam performa di kota-kota lainnya.

Pada model kedua, SARIMA (2,1,1)(1,1,1,12), menunjukkan perbedaan yang signifikan dalam hasil MAE dibandingkan model pertama. Model ini cenderung menghasilkan nilai MAE yang lebih tinggi pada kebanyakan kota dan kebangsaan. Misalnya, di DKI Jakarta untuk Malaysia, MAE tercatat sebesar 102.744,22, dan untuk Singapura di Batam, MAE tercatat sebesar 473.388,49. Model ini tidak menunjukkan hasil yang baik secara konsisten dan cenderung memberikan kesalahan prediksi yang lebih besar dibandingkan model pertama. Namun, terdapat beberapa kota seperti Medan dan Surabaya yang menunjukkan hasil yang lebih baik

untuk kebangsaan tertentu dengan MAE yang relatif rendah, seperti Medan untuk Filipina (MAE sekitar 535,60) dan Surabaya untuk Filipina (MAE sekitar 1.157,77).

Pada model ketiga, yaitu SARIMA (1,1,1)(2,1,1,12), memberikan hasil yang lebih bervariasi di setiap kota dan kebangsaan, tetapi secara umum menunjukkan performa yang cukup kompetitif. MAE pada beberapa kebangsaan dan kota menunjukkan nilai yang lebih rendah dibandingkan dengan model pertama dan kedua. Misalnya, DKI Jakarta untuk Vietnam (MAE sekitar 731,61) dan Surabaya untuk Australia (MAE sekitar 756,89). Namun, ada beberapa kasus di mana model ini juga memberikan MAE yang tinggi, seperti Singapura di Batam (MAE sekitar 523.428,43). Model ini lebih baik dalam beberapa kota dengan kebangsaan tertentu tetapi memiliki hasil yang kurang optimal untuk beberapa negara lain.

Berdasarkan analisis nilai *Mean Absolute Error* (MAE) dari ketiga model SARIMA, SARIMA (1,1,1)(1,1,1,12) cenderung memiliki performa yang lebih baik secara keseluruhan, dengan MAE yang relatif lebih rendah di banyak kota dan kebangsaan, terutama untuk beberapa kota besar dan kebangsaan yang sering muncul dalam data. Model kedua, SARIMA (2,1,1)(1,1,1,12), umumnya memberikan hasil yang kurang baik dengan MAE yang lebih tinggi, sehingga kurang disarankan untuk digunakan. Model ketiga, SARIMA (1,1,1)(2,1,1,12), memiliki beberapa kasus di mana performanya baik, tetapi secara keseluruhan tidak konsisten.

Untuk memvisualisasikan hasil prediksi model *Seasonal ARIMA* (SARIMA), berikut disajikan grafik yang menampilkan data historis serta prediksi jumlah wisatawan asal Malaysia di kota DKI Jakarta. Grafik ini memberikan gambaran bagaimana model SARIMA menangkap pola musiman dan tren dari data masa lalu, sekaligus menghasilkan prediksi untuk periode mendatang. Pada grafik, data historis wisatawan asal Malaysia ditampilkan dengan garis biru, sementara hasil prediksi untuk periode ke depan ditampilkan dengan garis merah. Area abu-abu di sekitar garis prediksi menunjukkan interval kepercayaan, yang menggambarkan rentang ketidakpastian dari hasil prediksi model. Visualisasi ini dapat membantu dalam mengevaluasi akurasi model SARIMA dengan parameter (1,1,1)(1,1,1,12) dalam meramalkan jumlah wisatawan asal Malaysia di DKI Jakarta.



Gambar 3. Grafik Prediksi Kunjungan Wisatawan Malaysia Di DKI Jakarta

4. Kesimpulan

Penelitian ini menganalisis data jumlah wisatawan mancanegara di beberapa kota besar di Indonesia menggunakan model *Seasonal ARIMA* (SARIMA) untuk menemukan model terbaik dalam memprediksi jumlah wisatawan berdasarkan kebangsaan. Tiga parameter SARIMA diuji: (1,1,1)(1,1,1,12), (2,1,1)(1,1,1,12), dan (1,1,1)(2,1,1,12), dengan evaluasi akurasi menggunakan *Mean Absolute Error* (MAE) pada data uji 12 bulan terakhir.

Hasil analisis menunjukkan bahwa model SARIMA (1,1,1)(1,1,1,12) memberikan performa terbaik secara keseluruhan, terutama untuk data dengan pola musiman yang stabil, seperti wisatawan asal Vietnam di Medan dan Australia di Surabaya. Hal ini menunjukkan bahwa model SARIMA(1,1,1)(1,1,1,12) mampu menangkap pola musiman yang konsisten, sehingga dapat diandalkan untuk prediksi kunjungan wisatawan dari negara-negara dengan tren kunjungan yang stabil sepanjang tahun.

Dengan menggunakan model SARIMA yang tepat, perencanaan dan strategi pemasaran untuk sektor pariwisata dapat lebih terarah, terutama dalam menyusun strategi promosi dan alokasi sumber daya yang disesuaikan dengan jumlah prediksi wisatawan dari negara-negara tertentu. Prediksi ini dapat membantu pemerintah Indonesia dan pelaku industri pariwisata untuk memaksimalkan potensi kunjungan wisatawan asing dan mempersiapkan fasilitas serta infrastruktur yang sesuai dengan kebutuhan wisatawan dari berbagai negara.

REFERENSI

- [1] S. Paludi, "SETAHUN PANDEMI COVID-19 DAN DAMPAKNYA TERHADAP INDUSTRI PARIWISATA INDONESIA," *Equilibrium: Jurnal Penelitian Pendidikan dan Ekonomi*, vol. 19, no. 01, 2022, doi: 10.25134/equi.v19i01.4337.
- [2] Muchammad Satrio Wibowo and L. A. Belia, "Partisipasi Masyarakat Dalam Pengembangan Pariwisata Berkelanjutan," *Jurnal Manajemen Perhotelan dan Pariwisata*, vol. 6, no. 1, pp. 25–32, Mar. 2023, doi: 10.23887/jmpp.v6i1.58108.
- [3] Tiara Dita Puspita and Vivaldy Ismail, "Analisis Strategi Pengembangan Digital Tourism Sebagai Promosi Pariwisata," *Gemawisata: Jurnal Ilmiah Pariwisata*, vol. 19, no. 1, 2023, doi: 10.56910/gemawisata.v19i1.262.
- [4] R. E. Caraka *et al.*, "Connectivity, sport events, and tourism development of Mandalika's special economic zone: A perspective from big data cognitive analytics," *Cogent Business & Management*, vol. 10, no. 1, p. 2183565, Dec. 2023, doi: 10.1080/23311975.2023.2183565.
- [5] D. A. D. Nasution, E. Erlina, and I. Muda, "Dampak Pandemi COVID-19 terhadap Perekonomian Indonesia," *Jurnal Benefita*, vol. 5, no. 2, 2020, doi: 10.22216/jbe.v5i2.5313.
- [6] A. Rusmini, S. Tinggi, and P. Ambarrukmo, "GAMBARAN DAMPAK PANDEMI COVID-19 TERHADAP DESTINASI DAN PARIWISATA DI INDONESIA."
- [7] D. T. Anggarini, "UPAYA PEMULIHAN INDUSTRI PARIWISATA DALAM SITUASI PANDEMI COVID -

- 19,” *Jurnal Pariwisata*, vol. 8, no. 1, 2021, doi: 10.31294/par.v8i1.9809.
- [8] S. Saadah, K. Shaleh, D. Arwaty, F. Sukmawati, R. F. Mulyawan, and D. Nababan, “Analisis Sektor Industri Pariwisata Yang Terdampak Covid – 19 Dan Upaya Pemulihan Ekonomi Indonesia Dari Sektor Pariwisata,” *Jesya*, vol. 6, no. 1, 2023, doi: 10.36778/jesya.v6i1.914.
- [9] T. Nguyen-Da, Y. M. Li, C. L. Peng, M. Y. Cho, and P. Nguyen-Thanh, “Tourism Demand Prediction after COVID-19 with Deep Learning Hybrid CNN–LSTM— Case Study of Vietnam and Provinces,” *Sustainability* 2023, Vol. 15, Page 7179, vol. 15, no. 9, p. 7179, Apr. 2023, doi: 10.3390/SU15097179.
- [10] B. S. Pratama, A. F. Suryono, N. Auliyah, and N. Chamidah, “COMPARISON OF LOCAL POLYNOMIAL REGRESSION AND ARIMA IN PREDICTING THE NUMBER OF FOREIGN TOURIST VISITS TO INDONESIA,” *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 1, 2024, doi: 10.30598/barekengvol18iss1pp0043-0052.
- [11] L. Tugas Akhir, “Prediksi Kedatangan Wisatawan Mancanegara di Bali Menggunakan Algoritma Support Vector Machine dan Regresi Linear”, Accessed: Aug. 28, 2024. [Online]. Available: <https://lib.mercubuana.ac.id>
- [12] Indra Gunawan, Dwi Purnomo Putro, and Adhika Pramita Widyassari, “Can Google Trends(GT) be used to predict tourist arrivals?: FB Prophet Machine Learning(ML) for Predicting Tourist Arrivals,” *INTERNATIONAL CONFERENCE ON DIGITAL ADVANCE TOURISM, MANAGEMENT AND TECHNOLOGY*, vol. 1, no. 1, 2023, doi: 10.56910/ictmt.v1i1.57.
- [13] F. Astuti, R. Dwi Becti, A. Arianita, B. Keliat, T. I. Sebo, and A. History, “Backpropagation algorithm modeling to predict the number of foreign tourist visits to indonesia via air gates ARTICLE INFO ABSTRACT,” *Desimal: Jurnal Matematika*, vol. 6, pp. 337–348, 2023, doi: 10.24042/djm.
- [14] D. R. Lyngkhai, S. B. Singh, R. Singh, and H. Tyngkan, “Trend Analysis of Milk Production in India,” *Asian Journal of Dairy and Food Research*, vol. 41, no. 2, 2022, doi: 10.18805/ajdfr.DR-1789.
- [15] F. E. Mokorimban, N. Nainggolan, Y. A. R. Langi, and K. Kunci, “Penerapan Metode Autoregressive Integrated Moving Average (ARIMA) dalam Model Intervensi Fungsi Step terhadap Indeks Harga Konsumen di Kota Manado,” *d’Cartesian: Jurnal Matematika dan Aplikasi*, vol. 10, no. 2, pp. 91–99, Sep. 2021, doi: 10.35799/DC.10.2.2021.34969.
- [16] K. R. A. Muslihin and B. N. Ruchjana, “Model Autoregressive Moving Average (ARMA) untuk Peramalan Tingkat Inflasi di Indonesia,” *Limits: Journal of Mathematics and Its Applications*, vol. 20, no. 2, p. 209, Jul. 2023, doi: 10.12962/limits.v20i2.15098.
- [17] J. Liu, Z. Zhao, Y. Zhong, C. Zhao, and G. Zhang, “Prediction of the dissolved gas concentration in power transformer oil based on SARIMA model,” *Energy Reports*, vol. 8, 2022, doi: 10.1016/j.egy.2022.03.020.

Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara.

Manatap Dolok Lauro, S.Kom., M.M.S.I., memperoleh gelar S.Kom. dari Universitas Tarumanagara. Kemudian memperoleh gelar M.M.S.I. dari Universitas Bina Nusantara. Saat ini aktif sebagai dosen Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara.

Surya Halim, mahasiswa Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Tarumanagara

Bagus Mulyawan, S.Kom., M.M., memperoleh gelar S.Kom. dari Universitas Gunadarma. Kemudian memperoleh gelar MM. dari Universitas Budi Luhur. Saat ini aktif sebagai dosen tetap