

Analisis Kinerja Algoritma Machine Learning Dalam Prediksi Data Deret Waktu Wisatawan Mancanegara Ke Bali

Performance Analysis of Machine Learning Algorithms for Time Series Forecasting of International Tourist Arrivals to Bali

A.A Istri Candra Manika Dewi^{*1}, Vania Noverina², Azka WIdya Rahmawati³, Agung Widiyanto⁴, Atika Ratna Dewi⁵

^{1,2,3,4,5}Program Studi Sains Data, Universitas Telkom, Kampus Purwokerto

e-mail: *1candramanika@student.telkomuniversity.ac.id

Abstrak - Pariwisata Bali pascapandemi menghadapi fluktuasi ekstrem kedatangan wisatawan mancanegara. Kegagalan memprediksi pola deret waktu yang sarat anomali ini dapat menyebabkan inefisiensi alokasi sumber daya operasional. Oleh karena itu, penelitian ini membandingkan kinerja algoritma Linear Regression, Random Forest, dan Support Vector Regression (SVR) untuk merumuskan kerangka peramalan yang paling adaptif. Data historis bulanan kunjungan wisatawan dari tahun 1982 hingga 2025 diproses menggunakan rekayasa fitur ekstensif yang mencakup efek lag, pola musiman tahunan, serta indikator pandemi. Evaluasi komparatif metrik kesalahan dilakukan melalui strategi validasi bertahap: *chronological split*, *expanding-window cross-validation*, dan *final holdout 2024-2025*. Hasil evaluasi menunjukkan bahwa Linear Regression menjadi model terbaik secara keseluruhan yang mampu mengungguli metode benchmark dengan tingkat WAPE terendah sebesar 8,14 persen pada pengujian akhir. Walaupun Random Forest terbukti lebih tangguh mengenali fluktuasi nonlinier pada fase pemulihan dan SVR konsisten kurang optimal, ketiga model menurun drastis saat menghadapi anomali pandemi ekstrem. Penelitian ini menunjukkan bahwa kecocokan model dengan dinamika perubahan data dari waktu ke waktu lebih penting daripada sekadar kerumitan matematis algoritma dalam mendukung perencanaan strategis pariwisata di masa mendatang.

Kata kunci – *machine learning; peramalan; regresi linier; time series; wisatawan mancanegara*

Abstract - Post-pandemic Bali tourism faced extreme fluctuations in foreign tourist arrivals. The failure to predict these highly anomalous time series patterns could lead to an inefficient allocation of operational resources. Therefore, this study compared the performance of Linear Regression, Random Forest, and Support Vector Regression (SVR) algorithms to formulate the most adaptive forecasting framework. Monthly historical data of tourist visits from 1982 to 2025 were processed using extensive feature engineering covering lag effects, annual seasonal patterns, and pandemic indicators. A comparative evaluation using error metrics was conducted through a phased validation strategy: *chronological split*, *expanding-window cross-validation*, and a 2024–2025 final holdout. The evaluation results showed that Linear Regression was overall the best model capable of outperforming baseline methods with the lowest WAPE of 8.14 percent in the final testing. Although Random Forest proved more robust in recognizing nonlinear fluctuations during the recovery phase and SVR consistently underperformed, all three models experienced a drastic decline when facing extreme pandemic anomalies. This study demonstrates that the alignment of a model with the evolving dynamics of data over time is more important than

the mathematical complexity of the algorithm in supporting future strategic tourism planning.

Keywords – forecasting; linear regression; machine learning; time series; tourist arrivals

I. PENDAHULUAN

Sektor pariwisata merupakan elemen esensial penggerak ekonomi global dan nasional yang berkontribusi masif terhadap produk domestik bruto, di mana Provinsi Bali secara konsisten menjadi episentrum utama penyumbang devisa terbesar di Indonesia [1], [2]. Memasuki era pascapandemi, dinamika kunjungan wisatawan mancanegara ke Bali menunjukkan fase pemulihan pesat yang diwarnai oleh tingkat fluktuasi dan volatilitas ekstrem dari waktu ke waktu [3]. Pola deret waktu (*time series*) kedatangan turis saat ini tidak lagi bersifat linier, melainkan sangat rentan terhadap anomali musiman dan perubahan drastis preferensi perjalanan global [4]. Fenomena ketidakpastian ini menghadirkan permasalahan struktural yang mendesak; kegagalan dalam memprediksi lonjakan atau penurunan jumlah wisatawan secara presisi akan bermuara pada inefisiensi alokasi sumber daya operasional, beban kapasitas berlebih (*overcapacity*) pada infrastruktur publik di musim puncak, hingga hilangnya potensi pendapatan daerah saat musim sepi [5]. Oleh karena itu, kemampuan peramalan (*forecasting*) yang akurat bukan sekadar instrumen analitik, melainkan instrumen mitigasi risiko yang sangat vital bagi pemerintah daerah dan pelaku industri pariwisata.

Untuk menjawab tantangan peramalan pada data yang sarat akan fluktuasi dan *noise*, komputasi cerdas berbasis *Machine Learning* terbukti secara empiris jauh lebih tangguh dibandingkan metode parametrik konvensional yang sering gagal menangkap pola keacakan akibat asumsi linearitas dan stasioneritas yang kaku [6], [7]. Namun, literatur akademik terkini menegaskan bahwa tidak ada satu pun algoritma prediktif yang superior secara universal untuk semua distribusi data, sehingga evaluasi komparatif yang spesifik pada data riil menjadi sangat krusial [8]. Berangkat dari urgensi tersebut, penelitian ini memfokuskan analisis komparasi pada tiga algoritma dengan pendekatan komputasi yang kontras: *Linear Regression*, *Random Forest*, dan *Support Vector Regression* (SVR). *Linear Regression* diimplementasikan sebagai model garis dasar (*baseline*) karena pendekatan matematisnya yang transparan dalam membaca tren tren historis, meski para ahli sepakat kemampuannya terbatas pada data non-linear [9]. Sebagai komparasinya, *Random Forest* dievaluasi secara kritis karena mekanisme agregasi pohon keputusannya terbukti ampuh mereduksi varians dan mencegah *overfitting* pada data fluktuatif [10]. Sementara itu, SVR diandalkan karena arsitektur pemetaan *kernel* di ruang dimensi tinggi diakui sangat tangguh dalam meminimalkan margin kesalahan struktural pada deret waktu yang dipenuhi nilai pencilan (*outliers*) [11].

Tinjauan literatur membedah adanya kesenjangan riset (*gap analysis*) yang tajam; mayoritas studi komparasi algoritma prediktif kepariwisataan masih mengandalkan dataset prapandemi yang stabil, atau mengevaluasi varian algoritma secara terisolasi tanpa pengujian lintas arsitektur yang komprehensif pada volatilitas tingkat destinasi spasial seperti Bali [2], [12]. Kontribusi kebaruan (*novelty*) dari penelitian ini terletak pada uji evaluatif yang sejajar dan berimbang antara model parametrik konvensional (*Linear Regression*) dengan arsitektur algoritma adaptif tingkat lanjut (*Random Forest* dan SVR) secara berdampingan menggunakan deret waktu wisatawan mancanegara ke Bali terbaru di era pasca pandemi yang bergejolak. Berpijak pada celah penelitian tersebut, hipotesis utama studi ini menyatakan bahwa

algoritma *Random Forest* dan SVR akan secara empiris dan signifikan mengungguli performa *Linear Regression* dalam mereduksi metrik kesalahan prediksi. Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi, menganalisis, dan membandingkan performa ketiga algoritma tersebut guna merumuskan kerangka peramalan yang paling adaptif. Temuan esensial ini diharapkan dapat memperkaya diskursus sains data terapan di bidang pariwisata, sekaligus membekali pemangku kebijakan dengan rekomendasi berbasis bukti (*evidence-based*) guna menyusun strategi tata kelola destinasi yang tangguh terhadap ketidakpastian di masa depan.

II. TINJAUAN TEORI

Peramalan deret waktu merupakan metode yang digunakan untuk memprediksi nilai pada periode mendatang berdasarkan pola historis yang tersimpan dalam data. Pada sektor pariwisata, data kunjungan wisatawan umumnya memiliki tren jangka panjang, pola musiman, dan serta gangguan eksternal yang dapat menyebabkan perubahan pola secara tiba-tiba [13]. Karakteristik tersebut membuat proses prediksi menjadi lebih kompleks karena hubungan antarperiode sering bersifat nonlinier dan tidak selalu memenuhi asumsi stasioneritas yang digunakan pada metode statistik tradisional. Oleh karena itu, pendekatan *machine learning* semakin banyak digunakan karena cenderung lebih fleksibel dibanding metode statistik klasik dalam mempelajari pola kompleks dan menangkap hubungan nonlinear pada data deret waktu [14]. Penelitian ini membandingkan tiga algoritma machine learning, yaitu Linear Regression, Random Forest, dan Support Vector Regression (SVR). Linear Regression digunakan untuk memodelkan hubungan linear antara variabel prediktor dan target sehingga sering digunakan sebagai baseline model karena mudah diimplementasikan dan diinterpretasikan. Random Forest merupakan metode ensemble berbasis banyak pohon keputusan yang mampu menangkap hubungan nonlinear serta interaksi antarfitur dengan lebih baik. Sementara itu, SVR bekerja dengan mencari pemisahan terbaik pada ruang fitur menggunakan kernel, sehingga mampu memodelkan pola yang lebih kompleks pada data deret waktu [9], [13].

Sejumlah penelitian terdahulu menunjukkan efektivitas machine learning dalam peramalan permintaan wisata. Yu dan Chen [12] menunjukkan bahwa model machine learning dapat memanfaatkan data historis untuk menghasilkan prediksi permintaan wisata yang akurat. Peng et al. [9] menemukan bahwa Random Forest mampu memberikan performa yang kompetitif dalam prediksi kedatangan wisatawan. Selain itu, Bouhaddour et al. [11] menunjukkan bahwa model nonlinier seperti Random Forest memiliki performa yang lebih baik dibandingkan SARIMA pada data pariwisata yang dipengaruhi pandemi COVID-19. Penelitian oleh Li et al. [4] juga menegaskan bahwa pendekatan forecasting modern berbasis data internet dan machine learning semakin banyak digunakan karena mampu menangkap dinamika permintaan wisata yang kompleks dan berubah secara cepat. Pada konteks Bali, Waciko et al. [14] menemukan bahwa optimasi hiperparameter mampu meningkatkan akurasi model berbasis pohon dalam memprediksi kunjungan wisatawan mancanegara.

Berdasarkan kajian tersebut, masih terbatas penelitian yang membandingkan secara langsung Linear Regression, Random Forest, dan SVR pada data kunjungan wisatawan mancanegara ke Bali yang mencakup periode sebelum pandemi, saat pandemi, dan masa pemulihan. Padahal, ketiga model tersebut merepresentasikan pendekatan yang berbeda dalam menangani pola linier maupun nonlinier pada data deret waktu. Oleh karena itu, penelitian

ini dilakukan untuk mengevaluasi dan membandingkan kinerja ketiga algoritma tersebut menggunakan data historis kunjungan wisatawan mancanegara ke Bali untuk menentukan model yang paling sesuai untuk peramalan pada kondisi data yang memiliki volatilitas tinggi dan perubahan pola yang signifikan.

III. METODE

3.1 PENGUMPULAN DATA

Penelitian ini menggunakan data jumlah kunjungan wisatawan mancanegara ke Bali yang diperoleh dari Badan Pusat Statistik (BPS) Bali. Dataset terdiri atas data historis kunjungan wisatawan bulanan periode 1982 hingga 2025. Selain data jumlah wisatawan bulanan, penelitian juga memanfaatkan data agregat tahunan dan data kedatangan berdasarkan pintu masuk sebagai sumber informasi tambahan untuk mendukung proses eksplorasi data.

Tabel 1. Ringkasan Dataset

Dataset	Periode	Jumlah Data	Kegunaan
Bali vs Indonesia	1969 - 2025	57 tahun	Analisis tren makro
Wisman by Month	1982 - 2025	528 bulan	Dataset utama pemodelan
Wisman by Gate	2009-2025	204 bulan	Analisis pintu masuk

Tabel 1 menyajikan ringkasan dataset yang digunakan. Data bulanan digunakan sebagai dasar pengembangan model peramalan, sedangkan data tahunan dan data berdasarkan pintu masuk dimanfaatkan untuk mengidentifikasi tren makro, pola musiman, serta karakteristik kedatangan wisatawan ke Bali.

3.2 PRA-PEMROSESAN DATA DAN REKAYASA FITUR

Tahap pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum digunakan dalam pemodelan. Data yang masih tersimpan dalam format spreadsheet diekstraksi ke bentuk tabular yang konsisten, kemudian dilakukan pembersihan dengan menghapus karakter non numerik dan mengonversi nilai kosong menjadi nilai null. Seluruh atribut numerik selanjutnya diubah ke tipe data numerik dan diurutkan berdasarkan tahun serta bulan untuk mempertahankan urutan kronologis yang menjadi karakteristik utama data deret waktu. Data tahun 2026 tidak digunakan karena hanya tersedia sebagian periode sehingga berpotensi menimbulkan bias dalam proses pelatihan model.

Selanjutnya dilakukan rekayasa fitur untuk menangkap pola temporal pada data kunjungan wisatawan. Fitur yang dibentuk meliputi representasi waktu berupa Year, Month_Sin, dan Month_Cos untuk merepresentasikan pola musiman tahunan; indikator Covid_Intervention untuk menandai periode gangguan akibat pandemi; fitur lag berupa Arrivals_Lag_1, Arrivals_Lag_2, Arrivals_Lag_3, Arrivals_Lag_6, Arrivals_Lag_12, dan Arrivals_Lag_24; serta fitur statistik bergerak yang terdiri atas rolling mean, rolling median, dan rolling standard deviation pada beberapa jendela waktu. Selain itu, dibentuk pula fitur Arrivals_Trend_1_3 untuk merepresentasikan tren jangka pendek dan Arrivals_YoY_Change untuk menangkap perubahan kunjungan dibandingkan periode yang sama pada tahun sebelumnya.

3.3 PENGEMBANGAN MODEL

Penelitian ini membandingkan tiga algoritma machine learning, yaitu Linear Regression, Random Forest, dan Support Vector Regression (SVR). Linear Regression digunakan sebagai model dasar untuk memodelkan hubungan linier antara variabel prediktor dan target. Random Forest dipilih karena kemampuannya dalam menangkap hubungan nonlinier

melalui mekanisme ensemble berbasis banyak pohon keputusan. Sementara itu, SVR digunakan untuk memodelkan pola yang lebih kompleks melalui transformasi kernel.

Implementasi Random Forest menggunakan 100 pohon keputusan dengan `random_state` sebesar 42 untuk menjaga reproduktibilitas hasil. Pada SVR digunakan kernel radial basis function (RBF) dengan parameter `C` sebesar 100.000 dan `gamma` sebesar 0,1. Karena SVR sensitif terhadap skala data, proses pelatihan didahului dengan standarisasi menggunakan `StandardScaler` yang diintegrasikan melalui pipeline. Seluruh eksperimen dilakukan menggunakan pustaka `scikit-learn`.

3.4 PEMBAGIAN DATA DAN STRATEGI VALIDASI

Pada penelitian ini, data dibagi secara kronologis untuk menjaga urutan waktu yang menjadi karakteristik utama data deret waktu. Pembagian data dilakukan menggunakan `train-test split` dengan proporsi 70:30. Data pada periode awal digunakan sebagai data latih dan data pada periode yang lebih baru digunakan sebagai data uji. Pendekatan ini dipilih untuk menghindari data leakage yang dapat terjadi apabila data masa depan digunakan dalam proses pelatihan model.

Selain pengujian menggunakan `train-test split`, penelitian ini juga menerapkan metode *expanding-window-cross-validation* menggunakan `TimeSeriesSplit`. Ukuran data latih bertambah secara bertahap pada setiap fold, sedangkan data uji selalu berada setelah data latih secara kronologis. Strategi validasi ini lebih sesuai untuk data deret waktu karena mampu mensimulasikan kondisi prediksi pada dunia nyata, yaitu memprediksi nilai masa depan berdasarkan informasi yang tersedia pada masa sebelumnya. Hasil validasi digunakan untuk mengevaluasi konsistensi performa model pada berbagai periode waktu.

3.5 EVALUASI MODEL

Kinerja model dievaluasi menggunakan beberapa metrik untuk memperoleh gambaran performa yang komprehensif. Metrik yang digunakan meliputi Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Weighted Absolute Percentage Error (WAPE), Symmetric Mean Absolute Percentage Error (sMAPE), dan koefisien determinasi (R^2). Penggunaan beberapa metrik secara bersamaan dilakukan karena masing-masing memiliki karakteristik dan sudut pandang evaluasi yang berbeda.

MAE digunakan untuk mengukur rata-rata besarnya kesalahan prediksi dalam satuan jumlah wisatawan, sedangkan RMSE memberikan penalti yang lebih besar terhadap kesalahan prediksi dengan magnitudo tinggi. WAPE dan sMAPE digunakan untuk mengukur kesalahan dalam bentuk persentase sehingga lebih mudah diinterpretasikan dan dibandingkan antarperiode dengan skala kunjungan yang berbeda. Sementara itu, R^2 digunakan untuk menunjukkan proporsi variasi data aktual yang dapat dijelaskan oleh model.

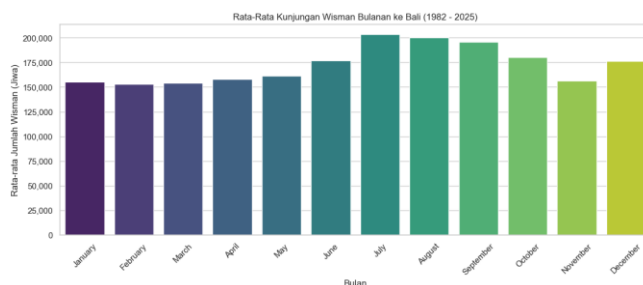
IV. HASIL DAN PEMBAHASAN

4.1 ANALISIS EKSPLORATIF DATA (EDA)

4.1.1 ANALISIS POLA MUSIMAN (SEASONALITY)

Analisis data bulanan periode 1982 hingga 2025 menunjukkan adanya pola musiman yang konsisten. Rata-rata kunjungan tertinggi terjadi pada Juli, Agustus, dan September yang bertepatan dengan musim liburan di belahan bumi utara, sedangkan Januari dan Februari menjadi periode dengan kunjungan terendah. Selisih antara bulan dengan kunjungan

tertinggi dan terendah mencapai hampir 33%. Perbandingan antara periode sebelum pandemi (2010–2019) dan setelah pemulihan (2022–2025) menunjukkan adanya perubahan pola musiman. Kenaikan terbesar terjadi pada Desember, diikuti Agustus, Oktober, dan September, sementara Februari mengalami sedikit penurunan. Hal ini mengindikasikan bahwa peningkatan kunjungan pascapandemi lebih terkonsentrasi pada periode puncak dan akhir tahun.

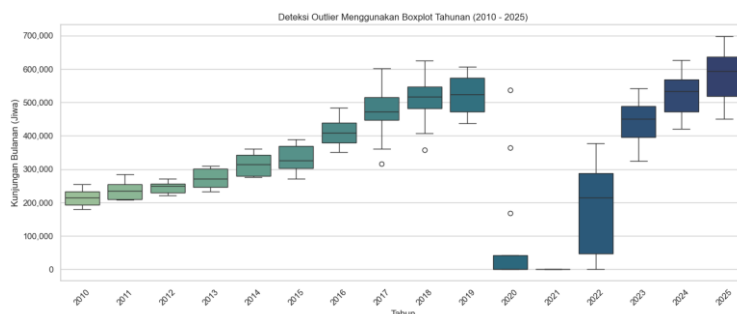


Gambar 1. Rata-rata Kunjungan Wisman Bulanan ke Bali (1982-2025)

Selain itu, intensitas pola musiman juga semakin kuat. Rasio puncak terhadap lembah meningkat dari 1,45 kali menjadi 1,71 kali, sedangkan koefisien variasi bulanan naik dari 12,2 persen menjadi 18,5 persen. Kondisi ini menunjukkan bahwa pola kunjungan pascapandemi menjadi lebih tajam sehingga lebih menantang untuk diprediksi oleh model. Pola musiman yang konsisten ini menjadi dasar penggunaan fitur lag-12, month_sin, dan month_cos pada tahap pemodelan. Secara konseptual, fitur tersebut dirancang untuk merepresentasikan pengaruh siklus tahunan sehingga model dapat memanfaatkan informasi musiman yang berulang dari tahun ke tahun.

4.1.4 ANALISIS OUTLIER DAN DAMPAK COVID-19

Pandemi COVID-19 menghasilkan outlier terbesar dalam seluruh rangkaian data. Jumlah kunjungan yang mencapai 536.611 jiwa pada Januari 2020 turun drastis menjadi 379 jiwa pada April 2020 dan bahkan mencapai nol kunjungan pada Juli serta Agustus 2021. Selama 21 bulan berturut-turut, jumlah kunjungan bulanan tidak pernah melebihi 379 jiwa. Dari perspektif statistik, kondisi tersebut menyebabkan distribusi data menjadi sangat ekstrem dengan tingkat kurtosis yang tinggi. Standar deviasi agregat sebesar 167.800 jiwa sebagian besar dipengaruhi oleh distorsi selama pandemi sehingga kurang merepresentasikan kondisi normal.



Gambar 2. Deteksi Outlier

Karakteristik data yang dipengaruhi pola musiman, perubahan periode, dan outlier ekstrem tersebut menunjukkan bahwa evaluasi model tidak dapat bergantung pada satu skema

validasi saja. Oleh karena itu, diperlukan beberapa pendekatan evaluasi untuk memperoleh gambaran kemampuan generalisasi model secara lebih komprehensif.

4.2 EVALUASI CHRONOLOGICAL

Berdasarkan hasil evaluasi *chronological split* 70:30, Linear Regression menunjukkan performa terbaik dengan nilai *test* MAE sebesar 75.341 jiwa, RMSE 93.604 jiwa, dan R^2 sebesar 0,7629. Namun, interpretasi hasil ini perlu dilakukan secara hati-hati karena data pengujian mencakup periode dengan karakteristik yang sangat berbeda, mulai dari kondisi normal sebelum pandemi, masa krisis COVID-19, hingga fase pemulihan. Oleh karena itu, satu nilai metrik agregat belum mampu menggambarkan kinerja model pada setiap periode waktu.

Tabel 2. Hasil Evaluasi Chronological Split 70:30

Model	Train MAE	Train RMSE	Test MAE	Test RMSE	Test WAPE (%)	Test sMAPE (%)	Test R^2
Linear Regression	7.382,82	10.773,05	75.341,03	93.603,71	20,89	48	0,76
Random Forest	2.890,62	4.579,17	168.699,87	200.782,57	46,77	76,38	-0,09
SVR	2.916,80	6.380,56	261.378,81	294.620,41	72,46	111,29	-1,35

Sebaliknya, Random Forest dan SVR menunjukkan perbedaan yang sangat besar antara kesalahan pada data pelatihan dan pengujian. Random Forest memiliki *train* MAE sebesar 2.891 jiwa, tetapi meningkat menjadi 168.700 jiwa pada data uji. Pola serupa juga terjadi pada SVR. Nilai R^2 Random Forest yang negatif mengindikasikan bahwa model tersebut bahkan lebih buruk dibandingkan prediksi menggunakan rata-rata data.

Temuan ini menunjukkan adanya indikasi *overfitting* pada Random Forest dan SVR terhadap pola historis pelatihan. Dengan demikian, evaluasi *chronological split* digunakan sebagai gambaran awal performa model dan bukan sebagai dasar utama dalam pemilihan model terbaik. Untuk memperoleh estimasi yang lebih stabil dan representatif, evaluasi selanjutnya dilakukan menggunakan expanding window cross validation.

4.3 EVALUASI EXPANDING-WINDOW CROSS-VALIDATION

Evaluasi menggunakan expanding-window cross-validation memberikan gambaran yang lebih reliabel karena model diuji pada beberapa konteks temporal dengan histori pelatihan yang terus bertambah. Pada skema ini, Linear Regression kembali menjadi model terbaik dengan rata-rata MAE sebesar 74.958 jiwa dan RMSE sebesar 86.989 jiwa. Nilai tersebut lebih baik dibandingkan Random Forest maupun SVR.

Tabel 3. Hasil Expanding-Window Cross-Validation (5-fold)

Model	CV MAE Mean	CV MAE Std	CV RMSE Mean	CV RMSE Std
Linear Regression	74.958,34	47.310,71	86.989,38	51.627,82
Random Forest	87.048,33	61.012,84	102.881,39	71.941,79
SVR	198.994,71	66.700,74	212.169,93	65.100,32

Standar deviasi RMSE yang relatif besar pada seluruh model menunjukkan adanya perbedaan tingkat kesulitan antar fold, terutama ketika data pengujian mencakup periode pandemi atau awal pemulihan. Random Forest memiliki variasi error tertinggi sehingga lebih sensitif terhadap perubahan komposisi data, sedangkan Linear Regression menunjukkan performa yang lebih konsisten.

SVR menghasilkan rata-rata MAE hampir 199.000 jiwa atau lebih dari dua kali lipat Linear Regression. Hasil ini mengindikasikan bahwa konfigurasi hiperparameter yang digunakan belum mampu menangkap pola global data secara optimal sehingga kurang sesuai untuk kasus peramalan ini.

4.4 EVALUASI FINAL HOLDOUT 2024–2025

4.4.1 METRIK AGREGAT DAN PERBANDINGAN DENGAN BENCHMARK

Pada evaluasi *final holdout*, Linear Regression menghasilkan MAE sebesar 45.057 jiwa, RMSE 56.574 jiwa, WAPE 8,14 persen, dan R^2 sebesar 0,4331. Dengan rata-rata kunjungan sekitar 550.000–600.000 wisatawan per bulan pada periode holdout, MAE sebesar 45.057 wisatawan setara dengan sekitar 7–8% dari volume aktual. Tingkat kesalahan tersebut menunjukkan bahwa model masih cukup akurat untuk mendukung perencanaan operasional, seperti penyediaan kapasitas hotel, transportasi, dan layanan imigrasi.

Tabel 4. Hasil Evaluasi Final Holdout (Januari 2024 – Desember 2025)

Model	Train MAE	Train RMSE	Test MAE	Test RMSE	WAPE (%)	sMAPE (%)	R^2
Linear Regression	14.713,32	24.538,56	45.056,97	56.573,77	8,14	8,53	0,4331
Random Forest	5.169,14	10.084,15	51.292,14	59.908,23	9,27	9,30	0,3643
SVR	8.129,07	17.033,69	105.757,07	127.805,12	19,11	20,89	-1,8932

Tabel 5. Benchmark Naive pada Final Holdout 2024-2025

Benchmark	MAE	RMSE	WAPE(%)	sMAPE (%)	R^2
Naive Lag-1	49.811,16	59.897,74	9	9,27	0,36
Seasonal Naive Lag-12	70.194	78.717	12,68	14,02	-0.1

Linear Regression juga berhasil melampaui kedua model benchmark. Dibandingkan *Naive Lag-1*, model ini menurunkan MAE sekitar 9,5 persen dan RMSE sebesar 5,6 persen. Sementara itu, performa Random Forest hampir setara dengan *Naive Lag-1*, sehingga nilai tambah prediktifnya relatif kecil. SVR secara konsisten memberikan hasil yang lebih buruk dibandingkan seluruh baseline.

Nilai R^2 yang lebih rendah dibandingkan evaluasi 70:30 tidak serta-merta menunjukkan penurunan kualitas model. Perbedaan tersebut dipengaruhi oleh karakteristik data holdout yang lebih stabil dengan variasi yang lebih kecil, sehingga nilai R^2 tidak dapat dibandingkan secara langsung antarperiode evaluasi yang berbeda.

4.4.2 KONSISTENSI ANTARTAHUN

Analisis per tahun menunjukkan bahwa model terbaik dapat berubah mengikuti karakteristik data. Evaluasi ini menggunakan model yang dilatih pada seluruh data pengembangan (Januari 1984-Desember 2023), kemudian diuji secara terpisah pada data tahun 2024-2025 untuk mengukur konsistensi performa antarperiode. Pada tahun 2024, Random Forest menghasilkan performa terbaik dengan RMSE sebesar 39.087 jiwa dan WAPE 6,68 persen. Hal ini mengindikasikan bahwa kemampuan menangkap hubungan nonlinier masih menguntungkan pada fase akhir pemulihan pariwisata.

Tabel 6. Performa Model per Tahun Holdout

Year	Model	MAE	RMSE	WAPE(%)	sMAPE (%)	R ²
2024	LR	41.183,18	50.701,49	7,8	8,09	0,35
2024	RF	35.276,96	39.086,97	6,7	6,71	0,62
2024	SVR	61.094,50	71.105,34	11,58	12,19	-0,27
2025	LR	48.930,76	61.891,37	8,5	8,98	0,36
2025	RF	67.307,32	75.167,81	11,62	11,88	0,05
2025	SVR	150.419,65	166.169,57	25,98	29,60	-3,6

Sebaliknya, pada tahun 2025 Linear Regression menjadi model terbaik dengan RMSE sebesar 61.891 jiwa. Error Random Forest meningkat secara tajam karena model cenderung mempelajari fluktuasi spesifik pada fase pemulihan sebelumnya sehingga kurang mampu melakukan generalisasi ketika jumlah kunjungan mencapai rekor baru.

Perubahan model terbaik dari Random Forest pada tahun 2024 menjadi Linear Regression pada tahun 2025 mengindikasikan bahwa karakteristik data terus mengalami evolusi selama fase pemulihan. Pada 2024, pola pemulihan masih mengandung unsur nonlinier yang cukup kuat sehingga menguntungkan Random Forest. Namun ketika pertumbuhan mulai stabil pada 2025, hubungan antarvariabel kembali menyerupai pola historis normal sehingga Linear Regression memperoleh keunggulan. Temuan ini menunjukkan bahwa tidak ada satu model yang selalu unggul pada seluruh kondisi. Kinerja model sangat dipengaruhi oleh perubahan karakteristik data atau periode waktu yang sedang berlangsung.

**Gambar 3.** Hasil Prediksi Seluruh Model

Gambar 6 menunjukkan bahwa Linear Regression menghasilkan kurva prediksi yang paling dekat dengan data aktual sepanjang periode holdout. Model mampu mengikuti arah tren kenaikan wisatawan tanpa menghasilkan deviasi yang terlalu besar. Sebaliknya, Random Forest cenderung menghasilkan prediksi yang lebih konservatif dan beberapa kali berada di bawah nilai aktual ketika terjadi lonjakan kunjungan. SVR menghasilkan kurva yang lebih halus (*over-smoothing*) sehingga gagal mengikuti perubahan yang terjadi secara cepat.

4.5 ANALISIS PERFORMA PER PERIODE WAKTU

Analisis berdasarkan periode waktu menunjukkan bahwa perilaku setiap algoritma berbeda sesuai kondisi data. Pada periode normal sebelum pandemi, Linear Regression memberikan hasil terbaik dengan WAPE 9,06 persen dan R² sebesar 0,75. Hal ini menunjukkan bahwa pola kunjungan pada kondisi stabil masih dapat dijelaskan secara memadai melalui hubungan

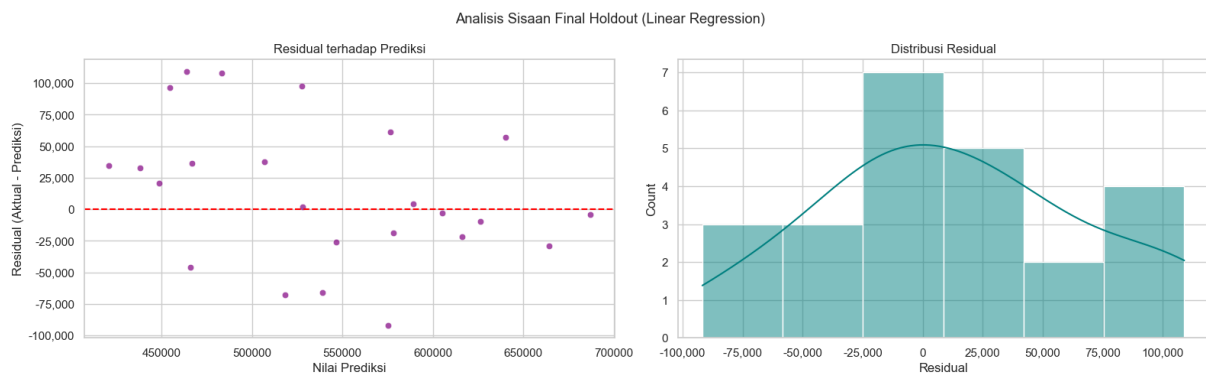
yang relatif linier.

Tabel 7. Performa Out-of-Fold per Periode Tahun

Periode Tahun		Model	N	MAE	RMSE	WAPE (%)	sMAPE (%)	R ²
Sebelum 19	COVID-19	LR	74	38.740,69	49.183,24	9,06	9,31	0,75
COVID-19 (Mar 2020-Dec 2021)		LR	22	120.288,80	147.964,80	1.572,65	194,80	-17
Setelah (2022-2023)	COVID-19	LR	24	145.076,48	153.737,61	46,87	76,14	0,15
Sebelum 19	COVID-19	RF	74	58.362,38	77.426,07	13,65	14,39	0,38
COVID-19 (Mar 2020-Dec 2021)		RF	22	211.170,42	245.679,89	2.760,82	194,46	-48,63
Setelah (2022-2023)	COVID-19	RF	24	61.718,11	70.628,09	19,94	42,83	0,82
Sebelum 19	COVID-19	SVR	74	183.163,07	204.282,34	42,84	53,63	-3,3
COVID-19 (Mar 2020-Dec 2021)		SVR	22	319.929,39	322.008,65	4.182,73	193,81	-84,26
Setelah (2022-2023)	COVID-19	SVR	24	136.952,13	150.168,06	44,24	63,72	0,19

Selama periode COVID-19, seluruh model mengalami penurunan performa yang sangat drastis dengan WAPE melebihi 1.500%. Kondisi ini mencerminkan keterbatasan pendekatan berbasis pola historis dalam menghadapi kejadian ekstrem yang tidak memiliki preseden dalam data, seperti penutupan perbatasan internasional secara mendadak.

Pada fase pemulihan 2022 hingga 2023, Random Forest justru menjadi model terbaik dengan R² sebesar 0,82. Pola pemulihan yang berlangsung cepat dan tidak linier lebih mudah ditangkap oleh pendekatan berbasis pohon keputusan. Temuan ini menunjukkan bahwa tidak terdapat algoritma yang unggul pada seluruh kondisi. Linear Regression mendominasi pada periode sebelum COVID-19, Random Forest lebih efektif pada fase setelah COVID-19, sedangkan seluruh model mengalami kesulitan selama pandemi. Dengan demikian, kesesuaian model terhadap karakteristik periode waktu menjadi faktor yang lebih menentukan dibandingkan pemilihan algoritma semata.



Gambar 4. Analisa Residual Model Terbaik

Gambar 7 menunjukkan bahwa residual model terbaik cenderung terdistribusi di sekitar garis nol tanpa pola sistematis yang kuat. Meskipun demikian, varians residual cenderung meningkat pada periode dengan jumlah kunjungan yang tinggi. Secara keseluruhan, hasil evaluasi menunjukkan bahwa Linear Regression memberikan keseimbangan terbaik antara akurasi, stabilitas, dan kemampuan generalisasi pada berbagai skenario pengujian.

V. KESIMPULAN DAN SARAN

Berdasarkan hasil evaluasi dan komparasi terhadap performa berbagai algoritma pembelajaran mesin dalam memprediksi deret waktu kunjungan wisatawan mancanegara ke Bali, dengan menerapkan rekayasa fitur *time series* yang ekstensif serta strategi validasi yang ketat (*chronological split 70:30*, *expanding-window cross-validation*, dan *final holdout 2024–2025*), diperoleh hasil bahwa Linear Regression merupakan model terbaik secara keseluruhan dengan nilai WAPE sebesar 8,14% dan RMSE sebesar 56.573,77 pada *final holdout*, serta mampu mengungguli kedua model pembandingan (*naive method*). Random Forest menunjukkan performa yang lebih baik dalam menangkap pola pada fase pemulihan pasca-pandemi, sedangkan SVR cenderung kurang optimal untuk kasus ini. Meskipun ketiga model mengalami kesulitan dalam memprediksi periode anomali ekstrem akibat pandemi COVID-19, Linear Regression menunjukkan kestabilan dan kemampuan generalisasi yang paling baik pada kondisi normal maupun periode pemulihan. Hasil penelitian ini menunjukkan bahwa pemilihan model peramalan perlu mempertimbangkan karakteristik periode waktu pada data, bukan hanya tingkat kompleksitas algoritma. Untuk penelitian selanjutnya, disarankan untuk menambahkan variabel eksternal seperti indikator ekonomi, tingkat hunian hotel, mobilitas wisatawan, atau kebijakan perjalanan internasional, serta mengeksplorasi model yang lebih kompleks seperti XGBoost, LightGBM, dan LSTM guna meningkatkan kemampuan prediksi terutama pada periode yang mengalami perubahan pola secara ekstrem.

VI. DAFTAR PUSTAKA

- [1] D. C. Wu, S. Zhong, J. Wu, and H. Song, "Tourism and Hospitality Forecasting With Big Data: A Systematic Review of the Literature," *Journal of Hospitality and Tourism Research*, vol. 49, no. 3, pp. 615–634, Mar. 2025, doi: 10.1177/10963480231223151.
- [2] D. T. Andariesta and M. Wasesa, "Machine learning models for predicting international tourist arrivals in Indonesia during the COVID-19 pandemic: a multisource Internet data approach," *Journal of Tourism Futures*, Jan. 2022, doi: 10.1108/JTF-10-2021-0239.

- [3] A. Dimitriadou, P. Gogas, and T. Papadimitriou, "Tourism and uncertainty: a machine learning approach," *Current Issues in Tourism*, vol. 28, no. 14, pp. 2278–2298, 2025, doi: 10.1080/13683500.2024.2370380.
- [4] X. Li, R. Law, G. Xie, and S. Wang, "Review of tourism forecasting research with internet data," Apr. 01, 2021, *Elsevier Ltd.* doi: 10.1016/j.tourman.2020.104245.
- [5] Y. Zhang, W. H. Tan, and Z. Zeng, "Tourism Demand Forecasting Based on a Hybrid Temporal Neural Network Model for Sustainable Tourism," *Sustainability (Switzerland)*, vol. 17, no. 5, Mar. 2025, doi: 10.3390/su17052210.
- [6] H. Fukuchi, W. Uehara, H. Kamata, and G. He, "COVID-19 policies and hoteliers' responses in Japan," *Annals of Tourism Research Empirical Insights*, vol. 3, no. 2, Nov. 2022, doi: 10.1016/j.annale.2022.100067.
- [7] R. X. Nie, C. Wu, and H. M. Liang, "Exploring Appropriate Search Engine Data for Interval Tourism Demand Forecasting Responding a Public Crisis in Macao: A Combined Bayesian Model," *Sustainability (Switzerland)*, vol. 16, no. 16, Aug. 2024, doi: 10.3390/su16166892.
- [8] S. Feng, "Job satisfaction, management sentiment, and financial performance: Text analysis with job reviews from indeed.com," *International Journal of Information Management Data Insights*, vol. 3, no. 1, Apr. 2023, doi: 10.1016/j.jjime.2023.100155.
- [9] L. Peng, L. Wang, X. Y. Ai, and Y. R. Zeng, "Forecasting Tourist Arrivals via Random Forest and Long Short-term Memory," *Cognit. Comput.*, vol. 13, no. 1, pp. 125–138, Jan. 2021, doi: 10.1007/s12559-020-09747-z.
- [10] B. Widodo, "Magnetohydrodynamics fluid flow passing through a sliced magnetic sphere influenced by mixed convection," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Mar. 2021. doi: 10.1088/1742-6596/1836/1/012042.
- [11] S. Bouhaddour, M. Sbihi, F. Guerouate, and C. Saadi, "Assessing Tourism Prediction Models: A Comparative Study of SARIMA, Random Forest, and LSTM, Considering Nonlinear Trends and the Influence of COVID-19," *Ingenierie des Systemes d'Information*, vol. 29, no. 6, pp. 2443–2454, Dec. 2024, doi: 10.18280/isi.290631.
- [12] N. Yu and J. Chen, "Design of Machine Learning Algorithm for Tourism Demand Prediction," *Comput. Math. Methods Med.*, vol. 2022, 2022, doi: 10.1155/2022/6352381.
- [13] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks," Aug. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/fi15080255.
- [14] K. J. Waciko, L. A. Susanti, M. Muayyad, and R. N. Fakhrurozi, "Forecasting Tourist Arrivals in Bali: A Grid Search-Tuned Comparative Study of Random Forest, XGBoost, and a Hybrid RF-XGBoost Model," *Inferensi*, vol. 8, no. 3, p. 251, Dec. 2025, doi: 10.12962/j27213862.v8i3.23334.