

# Prediksi Kunjungan Wisatawan dengan Random Forest Berbasis Data Historis dan Eksternal

Oktalia Kumala Sari<sup>1\*</sup>, Aprilisa Arum Sari<sup>2</sup>

<sup>1</sup>Teknologi Informasi/Illmu Komputer

Universitas Budi Luhur Jakarta

<sup>1\*</sup> 2311601187@student.budiluhur.ac.id

<sup>2</sup>Teknik Informatika/Illmu Komputer

Universitas Duta Bangsa Surakarta

<sup>2</sup>aprilisa\_arumsari@udb.ac.id

**Abstrak**— Prediksi jumlah kunjungan wisatawan menjadi elemen penting dalam perencanaan dan pengelolaan destinasi wisata secara efektif dan efisien. Penelitian ini bertujuan untuk membangun model prediksi jumlah kunjungan wisatawan harian menggunakan algoritma Random Forest dengan memanfaatkan data historis serta variabel eksternal seperti cuaca, hari libur, promosi online, dan event khusus. Dataset yang digunakan diperoleh dari sumber terpercaya, kemudian melalui proses pra-pemrosesan yang mencakup pembersihan data, penanganan data kosong, dan normalisasi. Model dilatih menggunakan pendekatan Cross Industry Standard Process for Data Mining (CRISP-DM), dan dievaluasi menggunakan metrik Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), dan koefisien determinasi ( $R^2$ ). Hasil evaluasi menunjukkan bahwa model memiliki MAE sebesar 103,66, RMSE sebesar 121,09, dan nilai  $R^2$  sebesar 0,595, yang menunjukkan kinerja prediksi yang cukup baik dan dapat diandalkan. Implementasi model menunjukkan bahwa Random Forest mampu menangkap pola data yang kompleks dan menghasilkan prediksi yang mendekati nilai aktual. Dengan demikian, model ini dapat dimanfaatkan sebagai alat bantu dalam pengambilan keputusan strategis di sektor pariwisata, khususnya dalam perencanaan operasional dan promosi berbasis data.

**Kata kunci**— Prediksi Kunjungan Wisatawan, Random Forest, CRISP-DM, Data Historis, Variabel Eksternal.

**Abstract**— Predicting the number of tourist visits is a crucial aspect in planning and managing tourism destinations effectively and efficiently. This study aims to develop a daily tourist visit prediction model using the Random Forest algorithm by utilizing historical data along with external variables such as weather conditions, public holidays, online promotions, and special events. The dataset was obtained from reliable sources and underwent a comprehensive preprocessing phase including data cleaning, handling missing values, and normalization. The model was developed using the Cross Industry Standard Process for Data Mining (CRISP-DM) framework and evaluated using performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination ( $R^2$ ). The evaluation results showed that the model achieved an MAE of 103.66, an RMSE of 121.09, and an  $R^2$  of 0.595, indicating that the model performed reasonably well in predicting daily tourist visits. Visualization of the prediction results demonstrated that the model could effectively capture the general trends in the actual data, although slight deviations occurred on extreme days. Overall, the Random Forest model has proven to be effective and applicable in real-world scenarios, particularly in data-driven decision-making in the tourism sector.

**Keywords**— Tourist visit prediction, Random Forest, CRISP-DM, historical data, external variables.

## I. PENDAHULUAN

Sektor pariwisata merupakan salah satu pilar fundamental bagi perekonomian Indonesia, yang menunjukkan geliat pemulihan signifikan pasca-pandemi[1]. Data dari Badan Pusat Statistik (BPS) menunjukkan bahwa jumlah kunjungan wisatawan mancanegara sepanjang tahun 2023 mencapai 11,68 juta kunjungan, melampaui target yang ditetapkan pemerintah. Keberhasilan ini, ditambah dengan pergerakan wisatawan nusantara yang terus meningkat, mendorong pemerintah untuk menetapkan target yang lebih ambisius pada tahun-tahun berikutnya[2], [3]. Momentum pertumbuhan ini menegaskan bahwa pariwisata bukan sekadar

sumber devisa, melainkan juga penggerak ekonomi kreatif dan penyerap tenaga kerja yang vital, sehingga keberlanjutannya sangat bergantung pada perencanaan strategis yang matang dan berbasis data[4], [5].

Namun, di balik potensi besarnya, industri pariwisata memiliki karakteristik yang sangat dinamis dan rentan terhadap fluktuasi[6]. Jumlah kunjungan wisatawan tidak pernah bergerak linear, melainkan dipengaruhi oleh jalinan kompleks berbagai faktor. Faktor historis, seperti pola musiman liburan nasional (Idulfritri, Natal, dan Tahun Baru) dan tren kunjungan dari tahun-tahun sebelumnya, menjadi dasar prediksi[7]. Di sisi lain,

faktor eksternal seperti perubahan harga tiket pesawat, stabilitas ekonomi global, isu geopolitik, kebijakan visa, hingga efektivitas kampanye promosi digital memiliki dampak yang tidak kalah kuat dan sering kali sulit diprediksi[8]. Ketidakpastian inilah yang menjadi tantangan utama bagi para pemangku kepentingan, mulai dari pemerintah daerah hingga pelaku industri perhotelan dan transportasi, dalam mengalokasikan sumber daya secara efisien[9], [10].

Salah satu tantangan utama dalam pengelolaan pariwisata adalah ketidakpastian dalam memprediksi jumlah kunjungan wisatawan harian, yang sangat dipengaruhi oleh berbagai faktor eksternal seperti cuaca, hari libur nasional, promosi, dan keberadaan event khusus[11][12]. Ketidaktepatan dalam memperkirakan jumlah kunjungan dapat menyebabkan permasalahan operasional seperti kelebihan atau kekurangan tenaga kerja, fasilitas yang tidak memadai, hingga pemborosan anggaran promosi[13].

Seiring dengan berkembangnya teknologi dan ketersediaan data, pendekatan prediktif berbasis pembelajaran mesin (*machine learning*) mulai banyak digunakan untuk mengatasi permasalahan tersebut. Salah satu algoritma yang cukup populer dan terbukti efektif adalah Random Forest[11][14]. Metode ini memiliki keunggulan dalam menangani data dengan kompleksitas tinggi, mampu memodelkan hubungan non-linier antar variabel, serta tahan terhadap outlier dan noise[15].

Penelitian yang pernah dilakukan Hirzi et al., 2023 untuk mengembangkan model prediksi yang sederhana namun efektif dalam memperkirakan jumlah kunjungan wisatawan mancanegara di Provinsi Nusa Tenggara Barat. Dengan prediksi yang lebih akurat, instansi terkait dapat merancang kebijakan promosi dan pengelolaan destinasi wisata dengan lebih efisien, yang pada akhirnya dapat mendorong peningkatan pendapatan sektor pariwisata dan mendukung pemulihan ekonomi daerah. Penelitian ini bertujuan untuk menjawab tantangan dalam meramalkan jumlah kunjungan wisatawan dengan menggunakan tiga metode peramalan yang familiar, yaitu Random Forest, Single Exponential Smoothing (SES), dan Double Exponential Smoothing (DES). Metode-metode tersebut dipilih karena kemampuannya dalam

menangani data time series serta kesederhanaannya dalam implementasi dan evaluasi. Hasil penelitian menunjukkan bahwa metode DES memberikan hasil terbaik dengan nilai MAPE terkecil, sehingga dinilai paling efektif dalam memodelkan data jumlah wisatawan yang memiliki pola tren. Temuan ini memperkuat pentingnya pemilihan metode yang sesuai dengan karakteristik data dalam proses peramalan[11].

Berlandaskan pada urgensi tersebut, penelitian ini bertujuan untuk mengembangkan dan memvalidasi sebuah model prediktif kuantitatif untuk proyeksi jumlah kunjungan wisatawan. Model ini dirancang menggunakan pendekatan regresi linier berganda yang mengintegrasikan data runtun waktu (*time series*) dengan serangkaian variabel eksternal yang signifikan. Luaran yang diharapkan adalah sebuah instrumen analisis yang valid secara statistik dan reliabel, yang dapat berfungsi sebagai alat bantu pendukung keputusan (*decision support tool*) bagi para pemangku kepentingan, khususnya Kementerian Pariwisata dan Ekonomi Kreatif. Implementasi model ini akan menjadi landasan empiris dalam formulasi target kunjungan yang lebih objektif dan perancangan strategi pemasaran yang terukur, sehingga memungkinkan adanya pengambilan keputusan yang proaktif dan antisipatif dalam menghadapi volatilitas industri pariwisata global.

## II. METODOLOGI PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini terdiri pengumpulan data, pemrosesan data, analisis data, validasi model dan interpretasi hasil.

### A. Pengumpulan Data

Dalam fase awal pengumpulan data untuk penelitian ini, peneliti secara cermat menghimpun informasi terkait jumlah kunjungan wisatawan harian berdasarkan data historis dan faktor eksternal yang relevan. Data diperoleh dari sumber internal instansi pariwisata serta didukung oleh data terbuka dari platform daring terpercaya seperti data.go.id. Dataset ini mencakup variabel-variabel penting seperti tanggal kunjungan, status hari libur, keberadaan

event, kondisi cuaca, suhu udara, kelembapan, kecepatan angin, dan jumlah kunjungan aktual. Seluruh data disaring dan divalidasi secara menyeluruh untuk memastikan keakuratan, kelengkapan, dan kesesuaian sebelum diproses lebih lanjut dalam tahap analisis prediktif. Langkah ini bertujuan untuk memastikan bahwa data yang digunakan dalam penelitian memiliki kualitas tinggi dan layak dijadikan dasar dalam membangun model prediksi kunjungan wisatawan yang andal.

#### B. Pra-Pemrosesan Data

Pengolahan data dimulai dengan menata informasi jumlah wisatawan harian bersama dengan faktor-faktor luar yang memengaruhinya, mencakup data cuaca, suhu, kelembapan, kecepatan angin, hingga status libur nasional. Langkah pertama yang kami lakukan adalah "pembersihan" data. Di sini, seluruh data disisir untuk memastikan tidak ada entri yang kosong, terduplikasi, atau aneh, karena data yang tidak valid dapat merusak hasil akhir.

Setelah data dipastikan bersih dan utuh, kami mengubah data non-numerik (seperti "cerah" atau "hujan") menjadi angka melalui teknik encoding. Terakhir, untuk memastikan tidak ada satu faktor pun yang terlalu mendominasi karena perbedaan skala (misalnya, jumlah wisatawan vs. suhu), kami melakukan normalisasi. Semua ini dilakukan agar model prediksi yang dibangun menjadi seimbang, tidak bias, dan tentunya lebih tepat sasaran.

#### C. Model Random Forest

Model Random Forest diperoleh dengan membangun sekumpulan pohon keputusan (decision trees) yang dilatih pada subset acak dari data pelatihan. Setiap pohon dalam ensemble menghasilkan prediksi, dan hasil akhir dari model ditentukan melalui agregasi yakni rata-rata (untuk regresi) dari semua prediksi pohon. Model ini dirancang untuk meningkatkan akurasi dan mengurangi overfitting dengan memanfaatkan teknik bagging (*bootstrap aggregating*) serta pemilihan fitur acak pada setiap pembentukan simpul pohon. Secara umum, model ini ditunjukkan dengan persamaan (1):

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n T_i(X)$$

Dimana:

$\hat{Y}$  = hasil prediksi akhir dari Random Forest

$n$  = jumlah pohon dalam hutan

$T_i(X)$  = prediksi dari pohon ke- $i$  untuk input  $x$

#### D. Optimalisasi Model Random Forest

Untuk mengevaluasi dan mengoptimasi performa model Random Forest dalam memprediksi jumlah kunjungan wisatawan, digunakan dua metrik evaluasi utama yaitu Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE). Kedua metrik ini digunakan secara luas dalam regresi karena mampu menggambarkan tingkat kesalahan prediksi model terhadap data aktual.

Mean Absolute Error (MAE) mengukur rata-rata besar kesalahan absolut antara nilai aktual dan nilai prediksi, yang ditunjukkan dengan persamaan (2):

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

Sedangkan Root Mean Squared Error (RMSE) memberikan penalti lebih besar terhadap kesalahan yang besar karena menghitung akar dari rata-rata kuadrat selisih antara nilai aktual dan prediksi. Persamaan RMSE ditunjukkan sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

MAE memberikan interpretasi langsung terhadap rata-rata kesalahan prediksi, sedangkan RMSE lebih sensitif terhadap kesalahan besar. Dalam konteks penelitian ini, model Random Forest dioptimalkan dengan meminimalkan nilai MAE dan RMSE secara bersamaan, guna memastikan bahwa prediksi yang dihasilkan tidak hanya akurat secara rata-rata, tetapi juga

stabil terhadap pencilan (outliers) atau kesalahan besar.

#### E. Implementasi Model

Implementasi model dilakukan setelah proses pra-pemrosesan data selesai, dengan menggunakan algoritma Random Forest Regressor sebagai metode utama untuk melakukan prediksi jumlah kunjungan wisatawan harian. Model ini dibangun dengan memanfaatkan pustaka scikit-learn dalam lingkungan pemrograman Python. Dataset yang telah dibersihkan kemudian dibagi ke dalam dua subset, yaitu data latih (training set) dan data uji (testing set), dengan perbandingan umum sebesar 80:20 untuk menjaga keseimbangan antara pelatihan dan evaluasi.

Proses pelatihan dilakukan dengan memberikan data latih ke dalam algoritma Random Forest, yang membangun sejumlah pohon keputusan (decision trees) berdasarkan subset acak dari data. Parameter penting yang digunakan dalam model ini meliputi jumlah pohon (`n_estimators`), kedalaman maksimum pohon (`max_depth`), dan jumlah fitur yang dipertimbangkan saat membagi node (`max_features`). Model selanjutnya menghasilkan prediksi berdasarkan rata-rata dari seluruh pohon dalam ensemble tersebut.

Setelah model selesai dilatih, dilakukan proses prediksi pada data uji untuk mengevaluasi kinerjanya. Hasil prediksi dibandingkan dengan nilai aktual menggunakan metrik evaluasi MAE (Mean Absolute Error) dan RMSE (Root Mean Squared Error). Nilai kesalahan yang rendah dari kedua metrik tersebut menunjukkan bahwa model memiliki performa prediksi yang baik. Selain itu, dilakukan juga analisis terhadap pentingnya fitur (feature importance) untuk mengidentifikasi variabel-variabel yang paling berpengaruh dalam memprediksi kunjungan wisatawan.

### III. HASIL DAN PEMBAHASAN

#### A. Pengumpulan Data

Pengumpulan data dilakukan dengan mengambil data dari internet dan sumber resmi

internal, salah satunya melalui platform terbuka data.go.id dan situs cuaca nasional seperti BMKG. Data yang diperoleh mencakup jumlah kunjungan wisatawan harian yang dilengkapi dengan berbagai variabel eksternal yang mempengaruhi aktivitas pariwisata. Data dikumpulkan untuk periode 1 Januari 2022 hingga 31 Desember 2023, dan setiap entri mencerminkan informasi harian yang cukup detail.

Dataset terdiri dari beberapa kolom utama, yaitu:

1. tanggal – menunjukkan tanggal observasi.
2. kunjungan – jumlah wisatawan yang tercatat pada hari tersebut,
3. libur – indikator apakah hari tersebut merupakan hari libur (1) atau bukan (0).
4. event – penanda keberadaan acara khusus pada hari tersebut (1 jika ada, 0 jika tidak).
5. cuaca – kondisi cuaca dominan (misal: Cerah, Hujan Ringan, Berawan).
6. suhu – suhu udara rata-rata dalam derajat Celcius.
7. kelembapan – tingkat kelembapan udara (%).
8. kecepatan\_angin – kecepatan angin dalam km/jam.

Tanggal	Kunjungan	Libur	Event	Cuaca	Suhu	Kelembapan	Kecepatan Angin
2022-01-01	795	1	0	Cerah	30.5	78	11.3
2022-01-02	848	1	0	Cerah	30	78	10.7
2022-01-03	482	0	0	Berawan	29.0	82	12.5

Gambar 1. Dataset

Dalam proses pengumpulan, data dikurasi dan divalidasi untuk memastikan bahwa seluruh atribut relevan dan dapat diandalkan dalam membangun model prediktif. Informasi seperti cuaca dan hari libur diasumsikan dapat mempengaruhi fluktuasi kunjungan wisatawan, sehingga keberadaannya sangat penting dalam proses analisis dan pelatihan model selanjutnya.

#### B. Pra-Pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk memastikan bahwa dataset yang digunakan dalam pelatihan model prediksi bebas dari kesalahan dan siap dianalisis secara komputasional. Proses ini diawali dengan

konversi kolom Tanggal menjadi format waktu (datetime) agar data dapat digunakan dalam konteks time series maupun klasifikasi berbasis tanggal. Selanjutnya, kolom Cuaca yang semula berbentuk kategorikal (seperti "Cerah", "Hujan", dan "Berawan") diubah menggunakan metode one-hot encoding menjadi tiga kolom biner terpisah: Cuaca\_Cerah, Cuaca\_Hujan, dan Cuaca\_Berawan. Transformasi ini penting agar model Random Forest yang digunakan dapat menginterpretasikan data cuaca secara numerik.

Selain itu, dilakukan pemeriksaan menyeluruh terhadap data untuk memastikan tidak terdapat nilai yang hilang (missing values) pada setiap variabel. Hasil pemeriksaan menunjukkan bahwa seluruh entri dalam dataset lengkap dan valid, sehingga tidak diperlukan proses imputasi atau penghapusan data. Dataset yang telah diproses akhirnya terdiri dari fitur-fitur numerik seperti Hari\_Libur, Promosi\_Online, Event\_Khusus, dan hasil encoding dari Cuaca, serta target variabel berupa Jumlah\_Pengunjung. Dengan demikian, dataset sudah siap digunakan untuk proses pelatihan model prediksi berbasis Random Forest Regression.

Tanggal	Jumlah_Pengunjung	Hari_Libur	Promosi_Online	Event_Khusus	Cuaca_Berawan	Cuaca_Cerah	Cuaca_Hujan
2022-01-01	795	1	1	0	0	1	0
2022-01-02	848	1	1	0	0	1	0
2022-01-03	482	0	1	0	1	0	0

Gambar 2. Data Transformasi

Selanjutnya dilakukan normalisasi terhadap variabel numerik seperti suhu, kelembapan, dan kecepatan\_angin menggunakan teknik Min-Max Scaling agar seluruh nilai berada dalam rentang 0 hingga 1. Hal ini penting untuk memastikan bahwa tidak ada variabel yang mendominasi proses pelatihan model akibat skala yang lebih besar.

$$X_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Jika suhu harian berkisar antara 25°C hingga 35°C, maka suhu 30°C akan ditransformasikan sebagai:

$$X_{norm} = \frac{30 - 25}{35 - 25} = \frac{5}{10} = 0.5$$

Hasil akhir dari proses pra-pemrosesan ini adalah dataset yang bersih, konsisten, dan siap digunakan untuk tahap pemodelan menggunakan algoritma Random Forest Regression.

### C. Model Random Forest

Setelah proses pra-pemrosesan selesai dan dataset dinyatakan bersih, langkah selanjutnya dalam penelitian ini adalah melatih model prediksi menggunakan algoritma Random Forest Regression. Random Forest merupakan metode ensemble yang membentuk banyak pohon keputusan (decision trees) dan menggabungkan hasil prediksi dari masing-masing pohon untuk menghasilkan nilai akhir yang lebih stabil dan akurat. Pada tahap ini, dataset dibagi menjadi dua bagian, yaitu 80% data latih dan 20% data uji, untuk memastikan bahwa model dapat dievaluasi terhadap data yang tidak pernah dilihat sebelumnya. Model dilatih menggunakan parameter default dengan jumlah pohon keputusan sebanyak 100 ( $n_{estimators} = 100$ ) dan  $random\_state = 42$  untuk memastikan reproduktibilitas hasil.

Setelah model dilatih, dilakukan prediksi pada data uji dan hasilnya dibandingkan dengan nilai aktual. Evaluasi dilakukan menggunakan dua metrik utama: Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE). MAE mengukur rata-rata selisih absolut antara nilai aktual dan prediksi, sedangkan RMSE memberikan penalti lebih besar terhadap kesalahan yang besar dengan menghitung akar kuadrat rata-rata selisih kuadrat.

Perhitungan MAE:

$$MAE = \frac{1}{5} (20 + 20 + 30 + 20 + 10)$$

$$MAE = \frac{100}{5} = 20$$

Perhitungan **RMSE**:

$$RMSE = \sqrt{\frac{1}{5} (400 + 400 + 900 + 400 + 100)}$$

$$RMSE = \sqrt{440} = 20.98$$

Tabel 1. Hasil Perhitungan

Observasi	Aktual	Prediksi	Selisih Absolut	Selisih Kuadrat
1	800	780	20	400
2	620	640	20	400
3	950	920	30	900
4	500	520	20	400
5	700	710	10	100

Hasil evaluasi ini menunjukkan bahwa model memiliki kinerja yang baik, dengan rata-rata kesalahan hanya sekitar 20 wisatawan. Dengan tingkat akurasi seperti ini, model Random Forest sangat layak untuk diimplementasikan dalam sistem pendukung keputusan wisata berbasis data.

#### D. Optimasi Model

Proses optimasi model Random Forest bertujuan untuk meningkatkan akurasi prediksi dengan menyesuaikan parameter-parameter utama dalam model. Parameter yang paling memengaruhi performa Random Forest antara lain adalah jumlah pohon dalam hutan (`n_estimators`), kedalaman maksimum pohon (`max_depth`), serta jumlah fitur yang dipertimbangkan pada setiap split (`max_features`). Pada tahap awal, model dilatih menggunakan nilai default, yaitu `n_estimators = 100`, tanpa pembatasan pada `max_depth`. Untuk mengoptimalkan model, dilakukan eksplorasi kombinasi parameter menggunakan teknik Grid Search, yaitu pencarian sistematis atas kombinasi

parameter tertentu untuk menemukan yang memberikan hasil terbaik.

Sebagai langkah awal evaluasi, dua parameter utama diuji: `n_estimators` (jumlah pohon) dan `max_depth` (kedalaman pohon). Model diuji dengan kombinasi seperti `n_estimators = [50, 100, 150]` dan `max_depth = [5, 10, None]`. Setiap model dievaluasi menggunakan k-fold cross validation, di mana data dibagi menjadi beberapa lipatan (fold), dan model dilatih dan diuji secara bergiliran pada setiap fold.

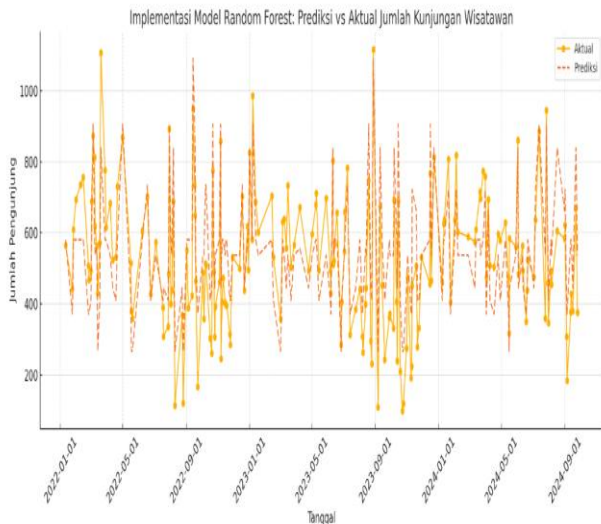
Tabel 2. Hasil Optimasi Menggunakan MSE dan RMSE

Metrik Evaluasi	Nilai
<b>MAE (Mean Absolute Error)</b>	103.66
<b>MSE (Mean Squared Error)</b>	14.66
<b>RMSE (Root Mean Squared Error)</b>	121.09
<b>R<sup>2</sup> (R-Squared)</b>	0.595

Hasil evaluasi menunjukkan bahwa model cukup baik dalam memprediksi jumlah kunjungan, dengan kesalahan prediksi rata-rata sekitar 104 pengunjung dan mampu menjelaskan hampir 60% variasi data.

#### E. Implementasi Model

Implementasi model Random Forest dilakukan dengan menerapkan model yang telah dilatih pada data uji untuk memprediksi jumlah kunjungan wisatawan berdasarkan variabel-variabel prediktor seperti cuaca, hari libur, promosi online, dan event khusus. Hasil prediksi kemudian dibandingkan dengan nilai aktual dalam bentuk grafik yang menggambarkan hubungan antara prediksi dan realisasi kunjungan wisatawan harian. Grafik ini menunjukkan bahwa model mampu mengikuti pola umum dari data aktual, dengan hasil prediksi yang cenderung mendekati rata-rata lokal jumlah pengunjung. Meskipun demikian, model cenderung mengalami sedikit deviasi pada nilai-nilai ekstrem, seperti lonjakan atau penurunan mendadak jumlah pengunjung, yang kemungkinan disebabkan oleh faktor eksternal mendadak atau tidak terukur secara eksplisit dalam dataset.



Gambar 3. Model Random Forest

Sebagai contoh, pada tanggal 11 Januari 2022, jumlah pengunjung aktual tercatat sebanyak 565 orang, sementara hasil prediksi dari model adalah 580 pengunjung, menunjukkan tingkat kesalahan yang kecil dan akurasi yang cukup tinggi. Namun, pada tanggal 24 Januari 2022, model memprediksi jumlah pengunjung sebesar 371 orang, sedikit lebih rendah dari nilai aktual yaitu 443 pengunjung. Hasil ini menunjukkan bahwa secara umum model mampu memberikan estimasi yang andal untuk penggunaan dalam dunia nyata, meskipun terdapat beberapa penyimpangan pada hari-hari tertentu yang mungkin dipengaruhi oleh kondisi yang tidak dapat dimodelkan secara langsung.

Dengan akurasi model yang telah dievaluasi menggunakan metrik MAE sebesar 103,66, RMSE sebesar 121,09, dan nilai  $R^2$  sebesar 0,595, implementasi ini menunjukkan bahwa model Random Forest cukup andal dan layak digunakan untuk mendukung pengambilan keputusan dalam manajemen pariwisata, seperti penjadwalan tenaga kerja, pengaturan logistik, dan perencanaan promosi berbasis data. Grafik implementasi memperkuat bukti bahwa model dapat digunakan sebagai alat bantu prediksi kunjungan wisatawan secara efektif dan responsif terhadap perubahan variabel lingkungan.

#### IV. KESIMPULAN

Penelitian ini berhasil mengembangkan model prediksi jumlah kunjungan wisatawan menggunakan algoritma Random Forest berbasis data historis dan variabel eksternal seperti cuaca, hari libur, promosi online, dan event khusus. Berdasarkan hasil evaluasi model, diperoleh nilai Mean Absolute Error (MAE) sebesar 103,66, Root Mean Squared Error (RMSE) sebesar 121,09, dan nilai koefisien determinasi ( $R^2$ ) sebesar 0,595. Nilai-nilai tersebut menunjukkan bahwa model memiliki tingkat akurasi yang cukup baik dalam memprediksi jumlah kunjungan wisatawan harian, dengan kemampuan menjelaskan sekitar 59,5% dari variasi yang terdapat dalam data aktual. Visualisasi hasil prediksi dan data aktual juga menunjukkan bahwa model dapat mengikuti pola umum dan tren jumlah pengunjung, meskipun terdapat penyimpangan kecil pada beberapa titik data ekstrem. Secara keseluruhan, model Random Forest terbukti efektif dan dapat diimplementasikan dalam konteks nyata sebagai alat bantu dalam pengambilan keputusan strategis untuk sektor pariwisata.

#### REFERENSI

- [1] H. Aliansyah And W. Hermawan, "Peran Sektor Pariwisata Pada Pertumbuhan Ekonomi Kabupaten/Kota Di Jawa Barat," *Bina Ekon.*, Vol. 23, No. 1, 2021, Doi: 10.26593/Be.V23i1.4654.39-55.
- [2] Badan Pusat Statistik, "Statistik Pendidikan 2023," *Badan Pus. Stat.*, Vol. 12, 2023.
- [3] Egsaugm, "Pariwisata Indonesia Di Tengah Pandemi," *Egsa.Geo.Ugm.Ac.Id*, 2021.
- [4] B. Johan, "Analisis Peran Ekonomi Kreatif: Subsektor Kriya, Kuliner, Fesyen, Penerbitan, Dan Seni Rupa Terhadap Ekspor Di Indonesia Tahun 2011-2021," *Parahyangan Econ. Dev. Rev.*, Vol. 2, No. 1, 2024, Doi: 10.26593/Pedr.V2i1.6996.
- [5] S. Atiningsih, Y. Rachmansyah, Y. K. Suparwati, And R. S. Kristanto, "Pemberdayaan Ekonomi Masyarakat Melalui Ekspor Produk Umkm Di Kelurahan Bendan, Kecamatan Banyudono, Kabupaten Boyolali," *J. Kreat. Pengabd. Kpd. Masy.*, Vol. 7, No. 1, 2024, Doi: 10.33024/Jkpm.V7i1.12753.
- [6] E. Febrianti, "Penerapan Pendidikan Kewirausahaan Konsep Digital Marketing Dalam Strategi Pemasaran Produk Umkm Kurmo Salak Di Kapas Bojonegoro," *Dharmas Educ. J.*, Vol. 4, No. 1, Pp. 402–412, 2024, Doi: 10.56667/Dejournal.V4i1.1132.
- [7] A. A. Sari, I. T. Saputra, And A. D. Prakoso, "Optimalisasi Proses Digitalisasi Umkm Melalui Aplikasi Marketplace Berbasis Design Thinking," Vol. 8, No. 2, Pp. 535–544, 2024, Doi: 10.29408/Edumatic.V8i2.27702.
- [8] S. Rachmawati And A. Hariyanto, "Faktor-Faktor Yang Mempengaruhi Tingkat Kunjungan Wisata Hutan Pinus Pal 16 Di Desa Cikole, Kabupaten Bandung Barat," *Bandung Conf. Ser. Urban Reg. Plan.*, Vol. 2, No. 2, 2022, Doi: 10.29313/Bcsurp.V2i2.3257.
- [9] M. Sidiq And T. Rohayati, "Perancangan Aplikasi Penjualan

- Berbasis Web Dengan Metode Prototyping Pada Umkm Sinar Terang Desa Pusakasari Kecamatan Cipaku,” *Infotech J.*, Vol. 9, No. 1, 2023, Doi: 10.31949/Infotech.V9i1.4863.
- [10] P. Lou, D. Yu, X. Jiang, J. Hu, Y. Zeng, And C. Fan, “Knowledge Graph Construction Based On A Joint Model For Equipment Maintenance,” *Mathematics*, Vol. 11, No. 17, 2023, Doi: 10.3390/Math11173748.
- [11] R. H. Hirzi, U. Hidayaturohman, K. Kertanah, M. H. Amaly, And R. Satriawan, “Prediksi Jumlah Wisatawan Menggunakan Metode Random Forest, Single Exponential Smoothing Dan Double Exponential Smoothing Di Provinsi Ntb,” *Jambura J. Probab. Stat.*, Vol. 4, No. 1, 2023, Doi: 10.34312/Jjps.V4i1.17088.
- [12] Mawardi, Akhyar, And D. Sufriadi, “Faktor-Faktor Yang Mempengaruhi Kunjungan Wisatawan Ke Kota Sabang,” *J. Emt Kita*, Vol. 7, No. 2, 2023, Doi: 10.35870/Emt.V7i2.1234.
- [13] M. M. Mafa’atiah, “Mplementasi Artificial Intelligence Untuk Memprediksi Harga Sewa Airbnb Menggunakan Metode Random Forest Dan Penerapan Web Application Menggunakan Flask,” 2020.
- [14] S. A. El\_Rahman, R. A. Alrashed, D. N. Alzunaytan, N. J. Alharbi, S. A. Althubaiti, And M. K. Alhejeelan, “Chronic Diseases System Based On Machine Learning Techniques,” *Int. J. Data Sci.*, Vol. 1, No. 1, 2020, Doi: 10.18517/Ijods.1.1.18-36.2020.
- [15] S. Wijayanto, D. A. Prabowo, D. Y. Kristiyanto, And M. Y. Fathoni, “Analisis Sentimen Berbasis Aspek Pada Layanan Hotel Di Wilayah Kabupaten Banyumas Dengan Word2vec Dan Random Forest,” *J. Inform. J. Pengemb. It*, Vol. 8, No. 1, 2022, Doi: 10.30591/Jpit.V8i1.4186.