

Prediksi Jumlah Pengunjung Wisata Guci Menggunakan Algoritma *Random Forest*

Azkiyatul Maulida^{1*}, Bambang Irawan², dan Nur Ariesanto Ramdhan³

¹ Program Studi Teknik Informatika, Universitas Muhadi Setiabudi; azkiyatulmaulida2@gmail.com

² Program Studi Teknik Informatika, Universitas Muhadi Setiabudi; bambangumus@gmail.com

³ Program Studi Teknik Informatika, Universitas Muhadi Setiabudi; ariesantoramdhan@gmail.com

* Penulis Korespondensi : Azkiyatul Maulida

Abstract: Tourism is one of the strategic sectors that plays an important role in increasing regional revenue, making effective planning supported by accurate information essential. One of the key pieces of information required is the prediction of visitor numbers as a basis for developing tourism destination management strategies. This study aims to develop a prediction model for the number of visitors to Guci Tourism using the Random Forest algorithm based on historical data. The research employed the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, which consists of six stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The dataset used in this study consisted of 60 monthly historical records of Guci Tourism from January 2021 to December 2025, obtained from the Department of Youth, Sports, and Tourism of Tegal Regency. The input variables were year and month, while the number of visitors was used as the target variable. The prediction model was developed using the Random Forest Regressor algorithm with an 80:20 split between training and testing data. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). The experimental results showed that the model achieved an MAE of 25,950.41, an RMSE of 32,835.06, and an R^2 value of 0.9624. These results indicate that the Random Forest algorithm is capable of producing visitor predictions that are close to the actual values with good predictive accuracy. Therefore, the Random Forest algorithm can be considered an alternative approach for developing visitor prediction models in the tourism sector.

Keywords: Data Mining; Random Forest; Visitor Prediction; Guci Tourism; Machine Learning.

Abstrak: Pariwisata merupakan salah satu sektor yang berperan penting dalam meningkatkan pendapatan daerah sehingga memerlukan perencanaan yang didukung oleh informasi yang akurat. Salah satu informasi yang dibutuhkan adalah prediksi jumlah pengunjung sebagai dasar dalam penyusunan strategi pengelolaan destinasi wisata. Penelitian ini bertujuan membangun model prediksi jumlah pengunjung Wisata Guci menggunakan algoritma Random Forest berdasarkan data historis. Penelitian menggunakan metodologi Cross Industry Standard Process for Data Mining (CRISP-DM) yang meliputi business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Dataset yang digunakan merupakan data historis bulanan Wisata Guci periode Januari 2021 hingga Desember 2025 yang diperoleh dari Dinas Pemuda, Olahraga, dan Pariwisata Kabupaten Tegal sebanyak 60 data. Variabel masukan yang digunakan terdiri atas tahun dan bulan, sedangkan jumlah pengunjung digunakan sebagai variabel keluaran. Model dibangun menggunakan algoritma Random Forest Regressor dengan pembagian data pelatihan sebesar 80% dan data pengujian sebesar 20%. Evaluasi model dilakukan menggunakan metrik Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan koefisien determinasi (R^2). Hasil penelitian menunjukkan bahwa model memperoleh nilai MAE sebesar 25.950,41, RMSE sebesar 32.835,06, dan R^2 sebesar 0,9624. Nilai tersebut menunjukkan bahwa algoritma Random Forest mampu menghasilkan prediksi jumlah pengunjung yang mendekati data aktual dengan tingkat akurasi yang baik. Oleh karena itu, algoritma Random Forest dapat digunakan sebagai alternatif metode dalam membangun model prediksi jumlah pengunjung pada sektor pariwisata.

Received June 26, 2026

Revised June 29, 2026

Accepted July 21, 2026

Available: July 25, 2026

Published: July 30, 2026



Hak cipta: © 2026 oleh penulis.

Diserahkan untuk kemungkinan publikasi akses terbuka

berdasarkan syarat dan ketentuan lisensi Creative Commons

Attribution (CC BY SA) (

<https://creativecommons.org/licenses/by-sa/4.0/>)

Kata kunci: Data Mining; Random Forest; Prediksi; Jumlah Pengunjung; Wisata Guci.

1. Pendahuluan

Pariwisata merupakan salah satu sektor strategis yang berkontribusi terhadap pertumbuhan ekonomi, peningkatan pendapatan daerah, serta penciptaan lapangan kerja. Selain menjadi sumber penerimaan daerah, sektor pariwisata juga mendorong berkembangnya berbagai sektor pendukung, seperti transportasi, perhotelan, perdagangan, dan usaha mikro, kecil, dan menengah (UMKM). Seiring meningkatnya jumlah wisatawan dan kompleksitas pengelolaan destinasi wisata, pengambilan keputusan tidak lagi dapat dilakukan hanya berdasarkan pengalaman atau perkiraan, tetapi perlu didukung oleh analisis data yang akurat. Oleh karena itu, pemanfaatan data historis untuk menghasilkan informasi prediktif menjadi salah satu pendekatan yang semakin penting dalam mendukung pengelolaan destinasi wisata secara efektif, efisien, dan berbasis data [1].

Kabupaten Tegal merupakan salah satu daerah di Provinsi Jawa Tengah yang memiliki potensi pariwisata cukup besar, dengan Wisata Guci sebagai destinasi unggulannya. Kawasan wisata yang berada di lereng Gunung Slamet ini dikenal memiliki sumber air panas alami, panorama pegunungan, serta berbagai wahana rekreasi yang menjadi daya tarik wisatawan dari berbagai daerah. Tingginya jumlah kunjungan wisatawan menjadikan Wisata Guci sebagai salah satu penyumbang penting bagi sektor pariwisata Kabupaten Tegal. Kondisi tersebut menuntut pengelola untuk mampu merencanakan kebutuhan fasilitas, sumber daya, serta strategi pengelolaan yang adaptif agar kualitas pelayanan kepada wisatawan tetap terjaga.

Berdasarkan data yang diperoleh dari Dinas Pemuda, Olahraga, dan Pariwisata Kabupaten Tegal, jumlah kunjungan wisatawan ke Wisata Guci selama periode 2021–2025 menunjukkan pola yang berfluktuasi. Fluktuasi tersebut dianalisis menggunakan data historis bulanan yang menjadi dasar dalam proses pembentukan model prediksi. Pada tahun 2021 jumlah kunjungan tercatat sebanyak 395.803 pengunjung, kemudian meningkat signifikan menjadi 724.358 pengunjung pada tahun 2022. Selanjutnya, jumlah kunjungan menurun menjadi 679.888 pengunjung pada tahun 2023, kembali meningkat menjadi 741.591 pengunjung pada tahun 2024, dan kembali mengalami penurunan menjadi 706.583 pengunjung pada tahun 2025. Perubahan jumlah kunjungan tersebut menunjukkan bahwa pola kedatangan wisatawan tidak bersifat tetap, melainkan dipengaruhi oleh berbagai faktor, seperti musim liburan, hari besar nasional, tren wisata, serta kebijakan pemerintah yang menyebabkan terjadinya fluktuasi dari waktu ke waktu. Kondisi ini menyulitkan pengelola dalam merencanakan kebutuhan fasilitas, alokasi sumber daya, penyusunan strategi promosi, serta estimasi potensi pendapatan pada periode berikutnya [2]. Oleh karena itu, diperlukan suatu pendekatan yang mampu memanfaatkan data historis untuk menghasilkan prediksi jumlah pengunjung yang lebih akurat sebagai dasar dalam mendukung proses pengambilan keputusan [3].

Untuk mengatasi permasalahan tersebut, diperlukan suatu metode yang mampu mengolah data historis menjadi informasi yang mendukung proses prediksi. Salah satu pendekatan yang banyak digunakan untuk menemukan pola tersembunyi dari data historis adalah data mining [4]. Pendekatan ini memungkinkan proses ekstraksi informasi dari data dalam jumlah besar sehingga dapat dimanfaatkan untuk menghasilkan model prediksi yang mendukung pengambilan keputusan berbasis data [5]. Dalam bidang pariwisata, data mining dapat dimanfaatkan untuk menganalisis pola kunjungan wisatawan berdasarkan data historis sehingga mampu menghasilkan prediksi jumlah pengunjung pada periode mendatang. Informasi hasil prediksi tersebut dapat digunakan sebagai dasar dalam menyusun strategi pengelolaan destinasi, meningkatkan kualitas pelayanan, mengoptimalkan pemanfaatan sumber daya, serta mendukung perencanaan operasional yang lebih efektif dan berbasis data. Dalam penelitian ini, proses pengembangan model prediksi mengikuti tahapan Cross Industry Standard Process for Data Mining (CRISP-DM), yang meliputi business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Penerapan CRISP-DM bertujuan untuk memastikan setiap tahapan penelitian dilakukan secara sistematis sehingga model yang dihasilkan memiliki proses pengembangan yang terstruktur dan mudah dievaluasi pada setiap tahap penelitian.

Salah satu algoritma machine learning yang banyak digunakan untuk menyelesaikan permasalahan prediksi adalah Random Forest. Algoritma ini merupakan metode ensemble learning yang membangun sejumlah decision tree dan menggabungkan hasil prediksi dari setiap pohon untuk menghasilkan model yang lebih stabil dan akurat [6]. Random Forest memiliki beberapa keunggulan, antara lain mampu menangani hubungan nonlinier antarvariabel, mengurangi risiko overfitting, serta memberikan performa yang baik pada data dengan karakteristik yang kompleks. Berdasarkan karakteristik tersebut, algoritma Random Forest dipilih karena mampu menghasilkan model prediksi yang stabil, memiliki kemampuan generalisasi yang baik, serta efektif dalam memodelkan hubungan antara variabel waktu dengan jumlah pengunjung berdasarkan data historis.

Beberapa penelitian sebelumnya telah menerapkan algoritma Random Forest untuk menyelesaikan berbagai permasalahan prediksi pada berbagai bidang. Penelitian yang dilakukan oleh Oktalia Kumala Sari dan Aprilisa Arum Sari mengembangkan model prediksi kunjungan wisatawan menggunakan data historis yang dipadukan dengan variabel eksternal seperti kondisi cuaca, hari libur, promosi daring, dan event khusus. Hasil penelitian tersebut menunjukkan bahwa algoritma Random Forest mampu memberikan performa prediksi yang cukup baik berdasarkan nilai evaluasi MAE, RMSE, dan R^2 . Sementara itu, penelitian lain oleh Haifan Ghani Raihan Santos dkk. menerapkan algoritma Random Forest untuk memprediksi jumlah pengunjung perpustakaan menggunakan data historis kunjungan dan menghasilkan model yang mampu menggambarkan pola kunjungan secara cukup baik. Meskipun demikian, penelitian yang secara khusus menerapkan algoritma Random Forest untuk memprediksi jumlah pengunjung Wisata Guci menggunakan data historis kunjungan dan data pendapatan retribusi yang diperoleh langsung dari Dinas Pemuda, Olahraga, dan Pariwisata Kabupaten Tegal masih sangat terbatas. Meskipun demikian, penelitian yang secara khusus menerapkan algoritma Random Forest untuk memprediksi jumlah pengunjung Wisata Guci menggunakan data historis yang diperoleh langsung dari Dinas Pemuda, Olahraga, dan Pariwisata Kabupaten Tegal masih sangat terbatas. Oleh karena itu, penelitian ini dilakukan untuk membangun model prediksi yang mampu memperkirakan jumlah pengunjung Wisata Guci berdasarkan pola historis sehingga dapat menjadi alternatif dalam mendukung pengambilan keputusan pada sektor pariwisata.

Berdasarkan permasalahan dan penelitian terdahulu tersebut, penelitian ini bertujuan untuk membangun model prediksi jumlah pengunjung Wisata Guci menggunakan algoritma Random Forest berdasarkan variabel tahun dan bulan. Model yang dibangun kemudian dievaluasi menggunakan metrik Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan koefisien determinasi (R^2) untuk mengetahui tingkat akurasi prediksi yang dihasilkan. Penelitian ini diharapkan dapat memberikan kontribusi dalam menyediakan model prediksi jumlah pengunjung Wisata Guci yang akurat serta menjadi referensi dalam penerapan algoritma Random Forest pada bidang pariwisata.

2. Tinjauan Literatur

2.1. Data Mining

Data mining merupakan proses menggali pola, hubungan, atau pengetahuan dari sekumpulan data untuk menghasilkan informasi yang bernilai dalam mendukung proses pengambilan keputusan [19]. Proses ini memanfaatkan berbagai teknik analisis untuk menemukan pola yang tidak mudah dikenali secara manual sehingga data yang tersedia dapat diolah menjadi informasi yang bermanfaat. Dalam perkembangannya, data mining telah banyak diterapkan pada berbagai bidang, seperti bisnis, kesehatan, pendidikan, keuangan, dan pariwisata [7]. Pada bidang pariwisata, data mining dapat dimanfaatkan untuk membangun model prediksi berdasarkan data historis sehingga mampu memberikan informasi yang mendukung perencanaan dan pengambilan keputusan secara lebih efektif. Oleh karena itu, pada penelitian ini data mining digunakan sebagai pendekatan untuk membangun model prediksi jumlah pengunjung Wisata Guci menggunakan algoritma Random Forest [8].

2.2 CRISP-DM

Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan metodologi standar yang banyak digunakan dalam pengembangan proyek data mining karena menyediakan alur kerja yang sistematis dan terstruktur [16]. Metodologi ini terdiri atas enam tahapan, yaitu business understanding, data understanding, data preparation, modeling,

evaluation, dan deployment [9]. Setiap tahapan saling berkaitan sehingga proses pengembangan model dapat dilakukan secara bertahap, mulai dari identifikasi permasalahan hingga evaluasi model yang dihasilkan. Dalam penelitian ini, CRISP-DM digunakan sebagai kerangka kerja untuk memastikan setiap tahapan pembangunan model prediksi jumlah pengunjung Wisata Guci dilakukan secara sistematis, terstruktur, dan sesuai dengan tujuan penelitian [10].

2.3 Random Forest

Random Forest merupakan algoritma machine learning berbasis ensemble learning yang bekerja dengan membangun sejumlah decision tree dari data pelatihan, kemudian menggabungkan hasil prediksi dari setiap pohon untuk memperoleh hasil akhir yang lebih stabil dan akurat [11]. Proses pembentukan setiap decision tree dilakukan menggunakan teknik bootstrap sampling dan pemilihan atribut secara acak (random feature selection), sehingga setiap pohon memiliki karakteristik yang berbeda. Pendekatan tersebut mampu meningkatkan kemampuan generalisasi model serta mengurangi risiko overfitting yang sering terjadi pada penggunaan satu decision tree [12].

Random Forest banyak digunakan pada berbagai permasalahan prediksi karena memiliki kemampuan dalam menangani hubungan nonlinier antarvariabel, menghasilkan model yang stabil, serta memberikan performa prediksi yang baik. Pada penelitian ini, algoritma Random Forest dipilih karena dinilai mampu memodelkan hubungan antara variabel tahun dan bulan terhadap jumlah pengunjung berdasarkan data historis sehingga dapat menghasilkan prediksi jumlah pengunjung Wisata Guci secara lebih akurat.

2.4 Evaluasi Model

Evaluasi model bertujuan untuk mengukur performa model prediksi yang dibangun menggunakan algoritma Random Forest. Pada penelitian ini digunakan tiga metrik evaluasi, yaitu Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan koefisien determinasi (R^2). MAE digunakan untuk mengukur rata-rata selisih absolut antara nilai aktual dan nilai prediksi, sedangkan RMSE mengukur besarnya kesalahan prediksi dengan memberikan penalti yang lebih besar terhadap kesalahan yang memiliki nilai tinggi. Sementara itu, koefisien determinasi (R^2) digunakan untuk mengukur kemampuan model dalam menjelaskan variasi data aktual. Semakin kecil nilai MAE dan RMSE serta semakin mendekati 1 nilai R^2 , maka semakin baik performa model prediksi yang dihasilkan [13].

3. Metode

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Dinas Pemuda, Olahraga, dan Pariwisata Kabupaten Tegal. Dataset terdiri atas 60 data historis bulanan Wisata Guci selama periode Januari 2021 hingga Desember 2025 yang memuat informasi tahun, bulan, pendapatan per bulan, dan jumlah pengunjung. Data historis tersebut digunakan sebagai sumber data dalam proses pembangunan model prediksi jumlah pengunjung menggunakan algoritma Random Forest [14]. Sebelum proses pemodelan dilakukan, dataset terlebih dahulu diperiksa untuk memastikan kelengkapan data, konsistensi format, serta kesesuaian tipe data sehingga siap digunakan pada tahap analisis selanjutnya.

3.2. Variabel Penelitian

Penelitian ini menggunakan tiga variabel penelitian yang terdiri atas dua variabel masukan (feature), yaitu tahun dan bulan, serta satu variabel keluaran (target), yaitu jumlah pengunjung. Variabel masukan digunakan dalam proses pelatihan model Random Forest untuk mempelajari hubungan antara periode waktu dan jumlah pengunjung, sedangkan variabel jumlah pengunjung digunakan sebagai variabel keluaran yang diprediksi oleh model. Rincian variabel yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1 Variabel Penelitian

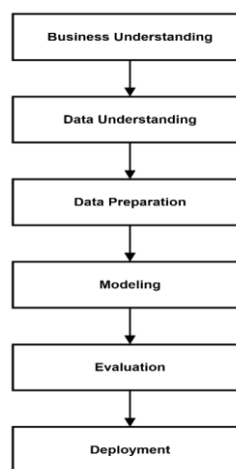
Variabel	Peran	Keterangan
Tahun	Feature	Tahun pengamatan data
Bulan	Feature	Bulan pengamatan (1–12)

Jumlah Pengunjung	Target	Jumlah pengunjung Wisata Guci setiap bulan yang menjadi target prediksi
-------------------	--------	---

Berdasarkan Tabel 1, variabel tahun dan bulan digunakan sebagai variabel masukan (feature) dalam proses pelatihan model Random Forest. Kedua variabel tersebut digunakan untuk mempelajari pola hubungan antara waktu pengamatan dan jumlah pengunjung sehingga model dapat menghasilkan prediksi jumlah pengunjung pada periode yang diinginkan. Sementara itu, variabel jumlah pengunjung berperan sebagai variabel keluaran (target) yang digunakan sebagai acuan dalam proses pelatihan dan evaluasi model.

3.3. Tahapan Penelitian

Penelitian ini menggunakan metodologi Cross Industry Standard Process for Data Mining (CRISP-DM) yang terdiri atas enam tahapan, yaitu business understanding, data understanding, data preparation, modeling, evaluation, dan deployment [15]. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1 Tahapan Metode CRISP-DM

Alur penelitian pada Gambar 1 menggambarkan tahapan penerapan metodologi CRISP-DM dalam membangun model prediksi jumlah pengunjung Wisata Guci menggunakan algoritma Random Forest. Setiap tahapan memiliki tujuan yang berbeda, mulai dari memahami permasalahan, menyiapkan data, membangun model, mengevaluasi performa model, hingga menghasilkan model prediksi yang siap digunakan [20]. Penjelasan setiap tahapan dijelaskan sebagai berikut.

1) Business Understanding

Menentukan tujuan penelitian, yaitu membangun model prediksi jumlah pengunjung Wisata Guci untuk mendukung proses perencanaan dan pengambilan keputusan.

2) Data Understanding

Mengumpulkan dan memahami karakteristik data historis bulanan Wisata Guci periode 2021–2025 yang diperoleh dari Dinas Pemuda, Olahraga, dan Pariwisata Kabupaten Tegal.

3) Data Preparation

Melakukan pemeriksaan terhadap dataset, memastikan tidak terdapat missing value, serta memastikan tipe data telah sesuai sehingga data siap digunakan pada proses pelatihan model.

4) Modeling

Membangun model prediksi menggunakan algoritma Random Forest dengan variabel masukan berupa tahun dan bulan serta jumlah pengunjung sebagai variabel keluaran. Proses pelatihan dilakukan menggunakan parameter yang telah ditentukan untuk menghasilkan model prediksi yang optimal.

5) Evaluation

Mengevaluasi performa model menggunakan metrik Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan koefisien determinasi (R^2).

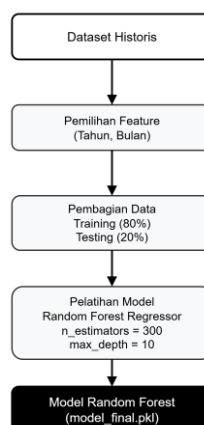
6) Deployment

Menyimpan model hasil pelatihan sehingga dapat digunakan kembali pada proses prediksi atau diintegrasikan ke dalam sistem lain sesuai kebutuhan.

3.4. Pemodelan Random Forest

Tahap pemodelan dilakukan menggunakan algoritma Random Forest Regressor untuk membangun model prediksi jumlah pengunjung Wisata Guci. Variabel masukan (feature) yang digunakan terdiri atas tahun dan bulan, sedangkan variabel keluaran (target) adalah jumlah pengunjung. Sebelum proses pelatihan dilakukan, dataset dibagi menjadi data pelatihan (training data) sebesar 80% dan data pengujian (testing data) sebesar 20% menggunakan metode train-test split dengan nilai random state sebesar 42. Pembagian data tersebut bertujuan untuk melatih model menggunakan data pelatihan serta mengevaluasi kemampuan model dalam melakukan prediksi terhadap data yang belum pernah digunakan selama proses pelatihan [17].

Model prediksi dibangun menggunakan algoritma Random Forest Regressor dengan parameter 300 decision tree ($n_estimators = 300$), kedalaman maksimum pohon (max_depth) sebesar 10, dan nilai random state sebesar 42. Parameter tersebut digunakan untuk menghasilkan model yang stabil, meningkatkan kemampuan generalisasi, serta mengurangi risiko overfitting [18]. Setelah proses pelatihan selesai, model disimpan dalam format `model_final.pkl` sehingga dapat digunakan kembali untuk menghasilkan prediksi pada data baru tanpa perlu melakukan proses pelatihan ulang.



Gambar 2 Alur Pemodelan Menggunakan Algoritma Random Forest

Gambar 2 menunjukkan alur proses pemodelan menggunakan algoritma Random Forest. Proses dimulai dari penyusunan dataset historis, dilanjutkan dengan pemilihan variabel masukan berupa tahun dan bulan. Dataset kemudian dibagi menjadi data pelatihan sebesar 80% dan data pengujian sebesar 20%. Selanjutnya dilakukan proses pelatihan menggunakan algoritma Random Forest Regressor dengan parameter $n_estimators$ sebanyak 300 dan max_depth sebesar 10. Model yang telah berhasil dilatih kemudian disimpan dalam format `model_final.pkl` untuk digunakan kembali pada proses prediksi.

3.5. Evaluasi Model

Tahap evaluasi dilakukan untuk mengukur kinerja model Random Forest dalam memprediksi jumlah pengunjung Wisata Guci. Evaluasi dilakukan menggunakan data pengujian (testing data) yang tidak digunakan selama proses pelatihan sehingga hasil evaluasi dapat menggambarkan kemampuan model dalam melakukan prediksi terhadap data baru. Pada penelitian ini, performa model diukur menggunakan tiga metrik evaluasi, yaitu Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan koefisien determinasi (R^2).

Mean Absolute Error (MAE) digunakan untuk mengukur rata-rata selisih absolut antara nilai aktual dan nilai hasil prediksi. Semakin kecil nilai MAE yang diperoleh, semakin baik tingkat akurasi model dalam melakukan prediksi. Root Mean Square Error (RMSE) digunakan untuk mengukur besarnya kesalahan prediksi dengan memberikan penalti yang lebih besar

terhadap kesalahan yang memiliki nilai tinggi sehingga mampu menggambarkan tingkat kesalahan model secara lebih sensitif. Sementara itu, koefisien determinasi (R^2) digunakan untuk mengukur kemampuan model dalam menjelaskan variasi data aktual. Semakin kecil nilai MAE dan RMSE serta semakin mendekati 1 nilai R^2 , maka semakin baik performa model prediksi yang dihasilkan. Perhitungan masing-masing metrik evaluasi mengacu pada Persamaan (1), Persamaan (2), dan Persamaan (3).

Persamaan (1) Mean Absolute Error (MAE)

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

Keterangan:

n = Jumlah data

y_i = Nilai aktual

$\hat{y}_i$ = Nilai hasil prediksi

Persamaan (2) Root Mean Square Error (RMSE)

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

Keterangan:

n = Jumlah data

y_i = Nilai aktual

$\hat{y}_i$ = Nilai hasil prediksi

Persamaan (3) Koefisien Determinasi (R^2)

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

Keterangan:

n = Jumlah data

y_i = Nilai aktual

$\hat{y}_i$ = Nilai hasil prediksi

$\bar{y}$ = Rata-Rata nilai aktual

4. Hasil dan Pembahasan

4.1 . Persiapan Data

Tahap persiapan data dilakukan untuk memastikan bahwa dataset yang digunakan telah memenuhi kebutuhan proses pemodelan. Dataset yang digunakan merupakan data historis bulanan Wisata Guci periode tahun 2021–2025 yang diperoleh dari Dinas Pemuda, Olahraga, dan Pariwisata Kabupaten Tegal. Dataset terdiri atas empat atribut, yaitu tahun, bulan, pendapatan per bulan, dan jumlah pengunjung. Namun, berdasarkan proses pemilihan variabel pada tahap data preparation, hanya atribut tahun dan bulan yang digunakan sebagai variabel masukan (feature), sedangkan jumlah pengunjung digunakan sebagai variabel keluaran (target) dalam proses pelatihan model Random Forest. Cuplikan lima data pertama dari dataset penelitian disajikan pada Tabel 2.

Tabel 2 merupakan cuplikan lima data pertama dari dataset penelitian.

Tahun	Bulan	Pendapatan Per Bulan	Jumlah Pengunjung
2021	1	42.281.500	59.110
2021	2	4.809.000	10.255
2021	3	12.369.500	15.926
2021	4	6.006.000	14.984
2021	5	76.380.000	58.250

Berdasarkan Tabel 2, dataset memuat informasi periode pengamatan yang direpresentasikan oleh variabel tahun dan bulan, serta jumlah pengunjung pada setiap periode. Meskipun dataset juga memuat atribut pendapatan per bulan, atribut tersebut tidak digunakan dalam proses pemodelan karena penelitian ini difokuskan pada pembangunan model prediksi

jumlah pengunjung berdasarkan variabel waktu. Oleh karena itu, proses pelatihan model hanya menggunakan variabel tahun dan bulan sebagai feature dengan jumlah pengunjung sebagai target prediksi.

Selanjutnya dilakukan pemeriksaan terhadap struktur dataset untuk mengetahui jumlah data, tipe data, serta kelengkapan setiap atribut yang digunakan dalam penelitian. Hasil pemeriksaan menunjukkan bahwa dataset terdiri atas 60 data historis bulanan yang berasal dari periode Januari 2021 hingga Desember 2025 dengan empat atribut, yaitu tahun, bulan, pendapatan per bulan, dan jumlah pengunjung. Ringkasan struktur dataset disajikan pada Tabel 3.

Tabel 3 Struktur Dataset

Atribut	Tipe Data	Jumlah Data Valid
Tahun	Integer	60
Bulan	Integer	60
Pendapatan Per Bulan	Integer	60
Jumlah Pengunjung	Integer	60

Berdasarkan Tabel 3, seluruh atribut memiliki tipe data integer (int64) dan masing-masing terdiri atas 60 data valid (non-null). Hasil tersebut menunjukkan bahwa tidak terdapat data yang hilang (missing value) pada dataset sehingga data telah memenuhi kebutuhan proses pelatihan dan pengujian model menggunakan algoritma Random Forest.

4.2 . Hasil Pemodelan Random Forest

Setelah tahap persiapan data selesai dilakukan, proses pemodelan dilaksanakan menggunakan algoritma Random Forest Regressor. Dataset dibagi menjadi data pelatihan (training data) sebesar 80% dan data pengujian (testing data) sebesar 20% menggunakan metode train-test split. Model dibangun menggunakan variabel tahun dan bulan sebagai feature, sedangkan jumlah pengunjung digunakan sebagai target. Parameter yang digunakan meliputi `n_estimators` sebanyak 300, `max_depth` sebesar 10, dan `random_state` sebesar 42.

Proses pelatihan menunjukkan bahwa model Random Forest berhasil dibangun sesuai dengan parameter yang telah ditentukan. Penggunaan 300 decision tree bertujuan untuk meningkatkan stabilitas hasil prediksi melalui mekanisme ensemble learning, sedangkan pembatasan kedalaman pohon (`max_depth`) sebesar 10 dilakukan untuk mengurangi risiko overfitting. Selain itu, penggunaan nilai `random_state` sebesar 42 bertujuan untuk memastikan proses pembagian data dan pelatihan model menghasilkan keluaran yang konsisten setiap kali proses dijalankan. Dengan konfigurasi tersebut, model siap digunakan pada tahap evaluasi untuk mengukur tingkat akurasi prediksi terhadap data pengujian.

4.3 . Evaluasi Model

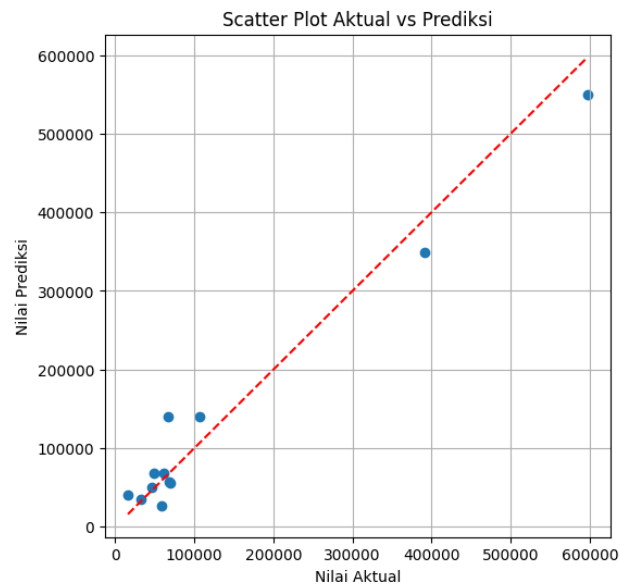
Tahap evaluasi dilakukan untuk mengukur kemampuan model Random Forest dalam memprediksi jumlah pengunjung Wisata Gucci menggunakan data pengujian (testing data). Evaluasi dilakukan menggunakan tiga metrik, yaitu Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan koefisien determinasi (R^2). Hasil evaluasi model disajikan pada Tabel 4.

Tabel 4 Hasil Evaluasi Model Random Forest

Metrik Evaluasi	Nilai
MAE	25.950,41
RMSE	32.835,06
R^2	0,9624

Berdasarkan Tabel 4, model Random Forest menghasilkan nilai MAE sebesar 25.950,41, yang menunjukkan bahwa rata-rata kesalahan prediksi model terhadap data aktual sebesar 25.950 pengunjung. Nilai RMSE sebesar 32.835,06 menunjukkan bahwa model masih memiliki tingkat kesalahan prediksi yang relatif rendah meskipun memberikan penalti yang lebih besar terhadap kesalahan dengan nilai tinggi. Sementara itu, nilai koefisien determinasi (R^2) sebesar 0,9624 menunjukkan bahwa model mampu menjelaskan sekitar 96,24% variasi jumlah pengunjung. Hasil tersebut menunjukkan bahwa algoritma Random Forest mampu

membangun model prediksi dengan tingkat akurasi yang baik berdasarkan variabel tahun dan bulan.



Gambar 3 Scatter Plot Nilai Aktual dan Hasil Prediksi Random Forest

Gambar 3 menunjukkan hubungan antara nilai aktual dan nilai hasil prediksi yang dihasilkan oleh model Random Forest pada data pengujian. Garis merah putus-putus merupakan garis ideal ($y = x$), yang menunjukkan kondisi ketika nilai prediksi sama dengan nilai aktual. Berdasarkan grafik tersebut, sebagian besar titik berada di sekitar garis ideal, sehingga menunjukkan bahwa hasil prediksi model memiliki kedekatan dengan nilai aktual. Temuan ini sejalan dengan nilai koefisien determinasi (R^2) sebesar 0,9624, yang mengindikasikan bahwa model memiliki kemampuan yang baik dalam memprediksi jumlah pengunjung Wisata Guci.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model Random Forest mampu menghasilkan prediksi yang mendekati nilai aktual. Nilai R^2 yang tinggi menunjukkan bahwa sebagian besar variasi jumlah pengunjung dapat dijelaskan oleh model, sedangkan nilai MAE dan RMSE yang relatif rendah mengindikasikan tingkat kesalahan prediksi yang kecil. Dengan demikian, model Random Forest dinilai memiliki performa prediksi yang baik dan dapat digunakan sebagai model prediksi jumlah pengunjung Wisata Guci.

5. Kesimpulan

Penelitian ini berhasil membangun model prediksi jumlah pengunjung Wisata Guci menggunakan algoritma Random Forest berdasarkan variabel tahun dan bulan. Proses pembangunan model dilakukan menggunakan metodologi CRISP-DM, yang meliputi tahapan business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Berdasarkan hasil evaluasi menggunakan data pengujian, model memperoleh nilai Mean Absolute Error (MAE) sebesar 25.950,41, Root Mean Square Error (RMSE) sebesar 32.835,06, dan koefisien determinasi (R^2) sebesar 0,9624. Hasil tersebut menunjukkan bahwa model memiliki tingkat kesalahan prediksi yang relatif rendah serta mampu menjelaskan sekitar 96,24% variasi data jumlah pengunjung, sehingga mampu menghasilkan prediksi yang mendekati nilai aktual.

Berdasarkan hasil tersebut, algoritma Random Forest terbukti memiliki performa yang baik dalam memprediksi jumlah pengunjung Wisata Guci berdasarkan data historis. Penelitian selanjutnya disarankan untuk menambahkan variabel lain yang berpotensi memengaruhi jumlah kunjungan, seperti kondisi cuaca, hari libur nasional, maupun penyelenggaraan event wisata, sehingga performa model prediksi dapat ditingkatkan.

Kontribusi Penulis: Penelitian ini memberikan kontribusi pada penerapan teknik data mining dalam bidang pariwisata, khususnya pada prediksi jumlah pengunjung Wisata Guci menggunakan algoritma Random Forest. Penelitian ini menghasilkan model prediksi

berdasarkan data historis bulanan yang mampu memberikan performa prediksi yang baik berdasarkan nilai evaluasi MAE, RMSE, dan koefisien determinasi (R^2). Hasil penelitian ini diharapkan dapat menjadi referensi dalam pengembangan model prediksi jumlah pengunjung pada sektor pariwisata serta mendukung proses pengambilan keputusan yang berbasis data.

Pendanaan: Penelitian ini tidak menerima pendanaan dari lembaga atau pihak eksternal mana pun. Seluruh proses penelitian, mulai dari pengumpulan data, pengolahan dan analisis data, hingga penyusunan artikel ilmiah, dilakukan secara mandiri oleh penulis. Dengan demikian, tidak terdapat kepentingan finansial yang berpotensi memengaruhi hasil maupun interpretasi penelitian.

Pernyataan Ketersediaan Data: Data yang digunakan dalam penelitian ini diperoleh dari Dinas Pemuda, Olahraga, dan Pariwisata Kabupaten Tegal. Data yang mendukung hasil penelitian tersedia dari penulis korespondensi dan dapat diberikan berdasarkan permintaan yang wajar dengan tetap memperhatikan ketentuan serta kebijakan dari instansi penyedia data.

Ucapan Terima Kasih: Penulis mengucapkan terima kasih kepada Dinas Pemuda, Olahraga, dan Pariwisata Kabupaten Tegal atas dukungan berupa penyediaan data penelitian. Penulis juga menyampaikan apresiasi kepada dosen pembimbing yang telah memberikan arahan, masukan, dan bimbingan selama proses penelitian hingga penyusunan artikel ini. Selain itu, penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan sehingga penelitian ini dapat diselesaikan dengan baik.

Konflik Kepentingan: Penulis menyatakan tidak ada konflik kepentingan. Penelitian ini tidak menerima pendanaan dari pihak mana pun sehingga tidak terdapat pihak yang memiliki peran dalam desain penelitian, pengumpulan dan analisis data, interpretasi hasil, penulisan naskah, maupun keputusan untuk mempublikasikan hasil penelitian.

Referensi

- [1] N. J. Nafiisah, "Pengaruh Sektor Pariwisata terhadap Pertumbuhan Ekonomi di Provinsi Jawa Tengah", *COSTING: Journal of Economic, Business and Accounting*, vol. 8, no. 3, pp. 2574–2586, 2025."
- [2] Marjo and E. Sumantri, "Implementasi Data Mining dalam Prediksi Target Produksi pada Proses Kerja Mesin Molding Menggunakan Algoritma Linear Regression (Studi Kasus: PT. AIM Karawang)," *Jurnal Indonesia: Manajemen Informatika dan Komunikasi (JIMIK)*, vol. 4, no. 3, pp. 1694–1703, Sep. 2023, doi: 10.35870/jimik.v4i3.397.
- [3] B. Susilo, N. A. Ramdhan, and O. S. Bachri, "Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Penjualan Produk Digital", *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1466–1476, Oct. 2024, doi:10.57152/malcom.v4i4.1517.
- [4] Marjo and E. Sumantri, "Implementasi Data Mining Dalam Prediksi Target Produksi Pada Proses Kerja Mesin Molding Menggunakan Algoritma Linear Regression (Studi Kasus: Pt. Aim Karawang)," *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, vol. 4, no. 3, pp. 1694–1703, Sep. 2023, doi: 10.35870/jimik.v4i3.397.
- [5] A. Riyandi, I. Nur Ardiansyah, and R. Dany, "Analisis Data Mining Untuk Prediksi Harga Saham: Perbandingan Metode Regresi Linier Dan Pola Historis Data Mining Analysis for Stock Price Prediction: A Comparison of Linear Regression Method and Historical Patterns," 2023.
- [6] H. Andrianof, A. P. Gusman, and O. A. Putra, "Implementasi Algoritma Random Forest untuk Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik: Studi Kasus di Perguruan Tinggi Indonesia," *Jurnal Sains Informatika Terapan (JSIT)*, vol. 4, no. 1, pp. 24–28, 2025.
- [7] A. Ernawati, Khairul, Z. Sitorus, M. Iqbal, and D. Nasution, "Penerapan Data Mining Untuk Klasifikasi Penduduk Miskin di Kabupaten Labuhanbatu Menggunakan Random Forest dan K-Nearest Neighbors," *Bulletin of Information Technology (BIT)*, vol. 6, no. 2, pp. 23–35, Jun. 2025.
- [8] Feri SLN, *Basic Data Mining from A to Z: Dasar Membangun Tindakan Bisnis*. 2024.
- [9] A. Rianti, N. W. A. Majid, and A. Fauzi, "CRISP-DM: Metodologi Proyek Data Science," *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*, pp. 107–114, 2023.
- [10] A. Muzakir, K. Adi, and R. Kusumaningrum, *Penerapan Konsep Machine Learning & Deep Learning*. Semarang: UNDIP Press, 2024.
- [11] J. Budiasto, T. M. Tallulembang, S. Pare, and M. Hasbi, *Machine Learning untuk Pemula (Konsep dan Implementasi)*. Jakarta: Penerbit Buku Indonesia, 2025.
- [12] O. K. Sari and A. A. Sari, "Prediksi Kunjungan Wisatawan dengan Random Forest Berbasis Data Historis dan Eksternal," *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*, pp. 328–335, 2025.
- [13] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Computer Science*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [14] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques*, 4th ed. Cambridge, MA, USA: Morgan Kaufmann, 2022.
- [15] L. Rahmadi, Hadiyanto, R. Sanjaya, and A. Prambayun, "Crop Prediction Using Machine Learning with CRISP-DM Approach," in *Proceedings of Data Analytics and Management*. Singapore: Springer, 2023, pp. 399–421.
- [16] P. M. Agata, "Integration of CRISP-DM and Machine Learning in Prediction Analysis Using Random Forest," *InfoSains*, vol. 5, no. 4, 2024.

-
- [17] Y. Yennimar, W. Leonardi, H. Weide, D. Cantona, and G. M. Hutagalung, "Comparison of Data Mining Algorithms (Random Forest, C4.5, CatBoost) Based on Adaptive Boosting in Predicting Diabetes Mellitus," *Jurnal Teknik Informatika C.I.T Medicom*, vol. 16, no. 1, pp. 1–12, 2024, doi: 10.35335/cit.Vol16.2024.730.pp1-12.
 - [18] S. Fadli, "Implementation of Data Mining on Tourist Visits Patterns Using Decision Tree Algorithm," *JISA (Jurnal Informatika dan Sains)*, 2022.
 - [19] E. Saputri, "Teknik dan Aplikasi Data Mining di Indonesia: Tinjauan Literatur Satu Dekade (2015–2024)," *IT-Explore*, vol. 4, no. 2, pp. 138–149, 2025, doi: 10.24246/itexplore.v4i2.2025.pp138-149.
 - [20] W. Silfianti et al., "A Deep Learning Approach for Tourism Destination Recommendation," *ILKOM Jurnal Ilmiah*, 2025.