Analisis Perbandingan Kinerja Algoritma Random Forest dan XGBoost dalam Klasifikasi Preferensi Wisatawan Kabupaten Kendal

Performance Comparison Analysis of Random Forest and XGBoost Algorithms in Classifying Tourist Preferences in Kendal Regency

Yuni Handayani *[1] , Taufik Hidayat[2], Muhammad Khozin[3], Tri Muji Waluyo[4]
[1,2,3,4]*Jurusan Teknik Informatika, Universitas Selamat Sri*
E-mail : yuni0406handayani@gmail.com[1], taufikhidayat.jc@gmail.com[2], khozin.dsn@gmail.com[3], trimujiwaluyo056@gmail.com[4]*
**Corresponding author*
Received 16 May 2026; Revised 26 June 2026; Accepted 1 July 2026

Abstrak - Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma *Random Forest* dan *XGBoost* dalam mengklasifikasikan preferensi pengunjung destinasi wisata di Kabupaten Kendal. Data diperoleh melalui kuesioner yang mencakup atribut usia, jenis kelamin, asal, transportasi, harga tiket, fasilitas, dan akses. Tahapan penelitian meliputi pembagian data menggunakan *train-test split* dengan proporsi 80:20 serta penanganan ketidakseimbangan data menggunakan teknik *SMOTE* pada data pelatihan. Proses pemodelan dilakukan dengan optimasi parameter menggunakan *GridSearchCV* untuk memperoleh model terbaik dari masing-masing algoritma. Evaluasi model dilakukan menggunakan metrik akurasi, *presisi*, *recall*, dan *F1-score* pada data uji, serta diperkuat dengan *Stratified K-Fold Cross Validation (k=5)* untuk memastikan stabilitas dan kemampuan generalisasi model. Hasil penelitian menunjukkan bahwa *Random Forest* memperoleh akurasi sebesar 89,55%, sedangkan *XGBoos*t menghasilkan akurasi yang lebih tinggi yaitu 90,30%. Hasil *cross-validation* juga menunjukkan bahwa *XGBoost* memiliki performa yang lebih stabil dibandingkan Random Forest. Analisis *feature importance* menunjukkan bahwa atribut usia menjadi faktor paling berpengaruh, diikuti oleh akses, asal, dan harga tiket. Hasil penelitian ini diharapkan dapat menjadi dasar dalam perumusan strategi pengembangan dan promosi pariwisata berbasis data.

Kata kunci – *Random Forest*, *XGBoost*, Klasifikasi, Preferensi Wisata

Abstract *– This study aims to analyze and compare the performance of the Random Forest and XGBoost algorithms in classifying visitor preferences for tourist destinations in Kendal Regency. Data were collected through questionnaires covering attributes such as age, gender, origin, transportation, ticket price, facilities, and accessibility. The research stages included splitting the dataset using an 80:20 train-test ratio and handling data imbalance with the SMOTE technique on the training set. Model development was conducted with parameter optimization using GridSearchCV to obtain the best model for each algorithm. Model evaluation employed accuracy, precision, recall, and F1-score metrics on the test data, reinforced by Stratified K-Fold Cross Validation (k=5) to ensure stability and generalization capability. The results show that Random Forest achieved an accuracy of 89.55%, while XGBoost yielded a higher accuracy of 90.30%. Cross-validation results also indicate that XGBoost demonstrates more stable performance compared to Random Forest. Feature importance analysis revealed that age was the most influential factor, followed by accessibility, origin, and ticket price. This study is expected to serve as a foundation for formulating data-driven strategies in tourism development and promotion.*

Keywords *– Random Forest, XGBoost, Classification, Tourist Preferences*

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214

1. PENDAHULUAN

Kabupaten Kendal merupakan salah satu wilayah di Provinsi Jawa Tengah yang memiliki potensi pariwisata yang cukup besar dan beragam [1]. Keanekaragaman tersebut mencakup berbagai jenis destinasi wisata, seperti wisata alam, budaya, religi, hingga wisata edukasi. Beberapa objek wisata unggulan yang dimiliki antara lain Curug Sewu, Pantai Ngebum, Pantai Cahaya, serta kawasan pegunungan dan perbukitan yang menawarkan panorama alam yang menarik. Selain itu, kekayaan budaya lokal daerah turut menjadi daya tarik bagi wisatawan. Dengan potensi tersebut, sektor pariwisata memiliki peran strategis dalam mendorong pertumbuhan ekonomi daerah serta meningkatkan kesejahteraan masyarakat [2].

Beberapa penelitian terdahulu telah membuktikan efektivitas kedua algoritma tersebut dalam berbagai bidang. Penelitian yang dilakukan [3] menunjukkan bahwa *Random Forest* memiliki performa yang lebih baik dalam klasifikasi mutu air dibandingkan metode lainnya. Penelitian [4] juga membuktikan bahwa *Random Forest* menghasilkan akurasi yang lebih tinggi dibandingkan SVM dalam memprediksi capaian studi. Selain itu, [5] menunjukkan bahwa *Random Forest* efektif dalam klasifikasi data dengan ketidakseimbangan kelas menggunakan teknik *oversampling*. Di sisi lain, algoritma *XGBoost* berbasis *gradient boosting* juga menunjukkan performa yang kompetitif. Penelitian [6] membuktikan bahwa *XGBoost* dan *Random Forest* menghasilkan performa yang lebih baik dibandingkan algoritma lain dalam klasifikasi kualitas udara dengan *SMOTE*. Penelitian mengintegrasikan SMOTE dengan *XGBoost* dan membuktikan keunggulannya dalam menangani ketidakseimbangan kelas pada berbagai domain dataset [7]. Sementara itu, [8] membandingkan *XGBoost*, *Random Forest*, dan *CatBoost* dalam memprediksi kunjungan wisatawan mancanegara dan menemukan bahwa *XGBoost* menghasilkan nilai *RMSE* terkecil. Penelitian menerapkan *Random Forest* dengan *SMOTE* pada data imbalanced dan membuktikan peningkatan akurasi sebesar 3,45% dibandingkan tanpa SMOTE.

Berdasarkan penelitian terdahulu [9][10] tersebut, dapat diketahui bahwa performa *Random Forest* dan *XGBoost* sangat bergantung pada karakteristik data dan konteks permasalahan. Namun, penelitian yang secara khusus [11] mengkaji klasifikasi preferensi wisatawan dengan mempertimbangkan faktor demografis dan faktor pendukung lainnya, serta membandingkan kedua algoritma tersebut dalam konteks pariwisata daerah, masih terbatas.
Oleh karena itu, penelitian ini dilakukan dengan mengumpulkan data secara daring melalui penyebaran kuesioner menggunakan *Google Forms* kepada 670 responden dengan latar belakang yang beragam. Data yang diperoleh mencakup atribut seperti jenis kelamin, usia, asal daerah, transportasi, harga tiket, fasilitas, serta akses menuju lokasi wisata. Selanjutnya, data tersebut melalui tahapan *preprocessing* yang komprehensif, meliputi penanganan *missing value*, penghapusan data duplikat, transformasi data kategorikal menggunakan label *encoding*, serta penerapan teknik *oversampling* untuk mengatasi ketidakseimbangan kelas.

Data yang telah diproses kemudian digunakan dalam pemodelan klasifikasi menggunakan algoritma *Random Forest* dan *XGBoost*. Model yang dihasilkan dievaluasi menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengetahui performa masing-masing algoritma. Hasil penelitian menunjukkan bahwa kedua algoritma mampu mengklasifikasikan preferensi wisata pengunjung dengan baik, di mana *XGBoost* menghasilkan tingkat akurasi yang lebih tinggi dibandingkan *Random Forest*. Selain itu, penelitian ini juga berhasil mengidentifikasi faktor-faktor penting yang mempengaruhi preferensi wisata, seperti harga tiket, akses, fasilitas, usia, dan asal pengunjung.

Adapun kebaruan *(novelty)* dalam penelitian ini terletak pada beberapa aspek, yaitu: (1) penggunaan pendekatan komparatif antara algoritma *Random Forest* dan *XGBoost* dalam klasifikasi preferensi wisatawan; (2) pemanfaatan data yang lebih komprehensif dengan melibatkan tidak hanya faktor demografis tetapi juga faktor pendukung seperti harga, akses, dan fasilitas; (3) penerapan proses *preprocessing* yang lengkap termasuk penanganan ketidakseimbangan data; serta (4) penggunaan dataset primer dengan jumlah responden yang

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214

cukup besar (670 responden) dalam konteks pariwisata daerah, khususnya Kabupaten Kendal. (5) penerapan evaluasi model yang lebih *robust* menggunakan *Stratified K-Fold Cross Validation*, sehingga menghasilkan performa yang lebih stabil dan representatif. (6) Selain itu, penelitian ini mengintegrasikan atribut demografis dan faktor kontekstual pariwisata seperti harga, akses, dan fasilitas dalam satu model klasifikasi. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi tidak hanya secara akademis dalam pengembangan metode klasifikasi berbasis *machine learning*, tetapi juga secara praktis sebagai dasar pengambilan keputusan bagi pemerintah daerah dan pengelola destinasi wisata dalam merancang strategi pengembangan dan promosi pariwisata yang lebih efektif, terarah, dan berbasis data.

2. METODE PENELITIAN

Penelitian dilakukan melalui beberapa tahapan untuk menganalisis preferensi pengunjung destinasi wisata di Kabupaten Kendal menggunakan algoritma *Random Forest* dan *XGBoost*. Kedua algoritma tersebut merupakan metode *ensemble learning* yang memanfaatkan kombinasi beberapa model pohon keputusan *(decision tree)* untuk meningkatkan akurasi klasifikasi serta mengurangi kesalahan prediksi.

1. Pemodelan *Random Forest*

Random Forest merupakan salah satu algoritma *machine learning* yang termasuk dalam metode ensemble learning dan banyak digunakan dalam proses klasifikasi. Algoritma ini bekerja dengan membangun sejumlah *Decision Tree* secara paralel dari data pelatihan. Setiap pohon keputusan menghasilkan prediksi masing-masing, kemudian seluruh hasil prediksi tersebut digabungkan untuk memperoleh keputusan akhir yang lebih akurat [18]. Rumus *Random Forest* dapat dilihat pada persamaan berikut. Rumus 1 menjelaskan bahwa nilai prediksi akhir diperoleh dengan menjumlahkan seluruh hasil prediksi dari masing-masing pohon keputusan $T_i(x)$ terhadap data masukan x kemudian dibagi dengan jumlah total pohon N. Dengan demikian, setiap pohon memberikan kontribusi berupa nilai prediksi yang selanjutnya dirata-ratakan sehingga menghasilkan output yang lebih stabil dan memiliki tingkat akurasi yang lebih baik. Selanjutnya, Rumus 2 menunjukkan bahwa prediksi akhir $\widehat{(y)}$ dalam proses klasifikasi ditentukan berdasarkan kelas yang paling sering muncul dari seluruh pohon keputusan $T_i(x)$ yang terdapat dalam forest. Setiap pohon memberikan satu "suara" terhadap suatu kelas, dan kelas dengan jumlah suara terbanyak akan dipilih sebagai hasil prediksi. Melalui mekanisme ini, Random Forest mampu mengurangi bias yang mungkin muncul dari satu pohon keputusan tunggal serta meningkatkan keandalan model klasifikasi karena keputusan dihasilkan dari kombinasi banyak model yang bekerja secara bersama-sama.

Regresi: $\hat{y} = \frac{1}{N}\sum_{i=1}^{N} T_i(x)$ ..... (1)

Klasifikasi: $\hat{y} = mode\ \{T_1(x).T_2(x).,,,,,.T_N(x)\}$.....(2)

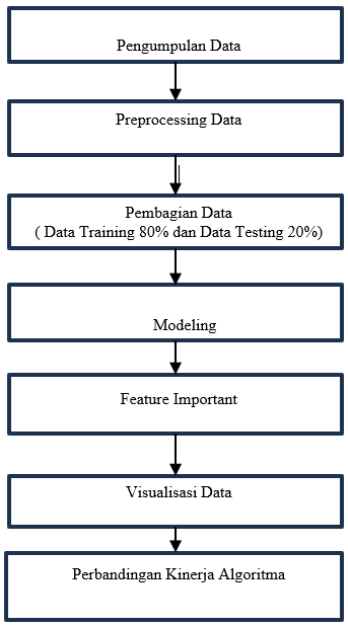
2. Pemodelan *XGBoost*

Extreme Gradient Boosting (XGBoost) merupakan salah satu algoritma *machine learning* berbasis *ensemble learning* [12][13] yang digunakan dalam penelitian ini untuk melakukan klasifikasi preferensi pengunjung terhadap destinasi wisata di Kabupaten Kendal. Algoritma ini memanfaatkan data latih yang terdiri dari berbagai variabel yang memengaruhi preferensi wisatawan, seperti usia, jenis kelamin, asal daerah, jenis destinasi wisata, fasilitas, serta faktor lainnya yang berkaitan dengan pilihan wisata. Melalui proses pembelajaran tersebut, *XGBoost* [14] bertujuan menemukan parameter optimal (θ) yang mampu merepresentasikan hubungan antara atribut-atribut tersebut dengan kelas preferensi pengunjung. Dalam proses pelatihannya, algoritma *XGBoost* membangun sejumlah pohon keputusan secara bertahap. Setiap pohon yang terbentuk berfungsi untuk memperbaiki kesalahan prediksi dari model sebelumnya, sehingga

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214

performa model semakin meningkat [12]. Dengan pendekatan tersebut, model mampu menghasilkan hasil klasifikasi yang lebih akurat terhadap variabel target, yaitu preferensi pengunjung terhadap destinasi wisata. Rumus dasar *XGBoost* dapat dilihat pada Persamaan 3. Fungsi objektif pada *XGBoost* terdiri dari dua komponen utama. Komponen pertama adalah fungsi *loss* $l(y_i.\hat{y_i}))$ yang digunakan untuk mengukur perbedaan antara nilai aktual dengan hasil prediksi model. Komponen kedua adalah fungsi regularisasi $\Omega(f_x)$ yang berfungsi untuk mengontrol kompleksitas model agar tidak mengalami *overfitting*. Dengan kombinasi kedua komponen tersebut, *XGBoost* mampu menghasilkan model klasifikasi yang tidak hanya akurat tetapi juga memiliki kemampuan generalisasi yang baik ketika diterapkan pada data baru.

$Obj(0) = \sum_i^n l(y_{i.}\, \hat{y} + \sum_{i=1}^K \Omega(fx))$.....(3)



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Data dalam penelitian ini diperoleh melalui metode survei *online* yang disebarkan kepada responden yang memiliki minat terhadap kegiatan pariwisata di Kabupaten Kendal. Penyebaran kuesioner dilakukan melalui berbagai *platform digital* seperti media sosial dan aplikasi pesan instan untuk menjangkau responden secara lebih luas dan beragam. Instrumen survei dirancang dalam bentuk kuesioner terstruktur yang memuat beberapa pertanyaan terkait karakteristik demografis responden serta preferensi mereka terhadap jenis destinasi wisata.

Atribut demografis yang dikumpulkan dalam penelitian ini meliputi jenis kelamin, usia, tingkat pendidikan, dan jenis pekerjaan. Selain itu, responden juga diminta untuk memilih jenis wisata yang paling diminati, seperti wisata alam, wisata budaya, wisata religi serta wisata edukasi. Informasi tersebut digunakan sebagai variabel dalam proses analisis data untuk mengidentifikasi pola hubungan antara karakteristik demografis wisatawan dengan preferensi jenis wisata yang dipilih.

Data yang terkumpul dari hasil survei kemudian digunakan sebagai dataset penelitian yang selanjutnya akan melalui tahap *preprocessing* sebelum diterapkan pada model klasifikasi menggunakan algoritma *Random Forest* dan *XGBoost*.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214

2.2 *Preprocessing* Data

Setelah data diperoleh dari hasil survei *online*, tahap selanjutnya adalah melakukan *preprocessing* data untuk memastikan bahwa dataset yang digunakan memiliki kualitas yang baik dan siap digunakan dalam proses pemodelan machine learning. Tahap ini bertujuan untuk meningkatkan konsistensi, kebersihan, serta keterbacaan data sehingga model klasifikasi dapat bekerja secara optimal.

Langkah awal *preprocessing* dilakukan dengan pembersihan dan penyesuaian struktur data. Pada tahap ini, nama kolom pada dataset diperiksa dan disesuaikan agar memiliki format yang konsisten.

Selanjutnya dilakukan proses transformasi data kategorikal menjadi data numerik menggunakan teknik *Label Encoding*. Proses ini dilakukan karena algoritma machine learning seperti *Random Forest* dan *XGBoost* hanya dapat memproses data dalam bentuk numerik. Setiap nilai kategorikal pada atribut seperti jenis kelamin, usia, pendidikan, pekerjaan, dan preferensi jenis wisata diubah menjadi representasi angka menggunakan pustaka *LabelEncoder* dari *scikit-learn*.

Setelah proses *encoding* selesai, dataset kemudian dipisahkan menjadi dua bagian utama, yaitu variabel fitur dan variabel target. Variabel fitur berisi atribut demografis responden seperti jenis kelamin, usia, tingkat pendidikan, dan pekerjaan, sedangkan variabel target berisi kategori jenis wisata yang menjadi preferensi responden.

Tahap berikutnya adalah pembagian dataset menjadi data pelatihan (*training* data) dan data pengujian (*testing* data) menggunakan metode *train-test split*. Pada penelitian ini, dataset dibagi dengan proporsi 80% data *training* dan 20% data *testing*. Data *training* digunakan untuk melatih model klasifikasi, sedangkan data *testing* digunakan untuk mengevaluasi performa model yang dihasilkan.

Melalui proses preprocessing ini, dataset yang dihasilkan menjadi lebih terstruktur, bersih, dan siap digunakan pada tahap pemodelan menggunakan algoritma *Random Forest* dan *XGBoost* untuk melakukan klasifikasi preferensi pengunjung terhadap destinasi wisata di Kabupaten Kendal.

2.3 Pembagian Data

Setelah melalui tahap *preprocessing*, dataset kemudian dibagi menjadi dua bagian, yaitu data *training* dan data *testing*. Pembagian ini bertujuan untuk melatih model serta mengevaluasi kinerja model dalam melakukan prediksi pada data yang belum pernah digunakan sebelumnya.

Pada penelitian ini, dataset dibagi dengan proporsi 80% data *training* dan 20% data *testing* menggunakan metode *train-test split* dari pustaka *scikit-learn*. Data training digunakan untuk membangun model klasifikasi menggunakan algoritma *Random Forest* dan *XGBoost*, sedangkan data *testing* digunakan untuk menguji performa model dalam mengklasifikasikan preferensi wisatawan. Pembagian data dilakukan dengan parameter *random_state* = 42 agar proses pembagian data bersifat konsisten dan dapat direproduksi.

2.4 Modeling Penerapan Algoritma *Random Forest*

Pada tahap ini dilakukan proses pembangunan model klasifikasi menggunakan algoritma *Random Forest*. Algoritma ini merupakan metode *ensemble learning* yang bekerja dengan membangun sejumlah pohon keputusan *(decision tree)* secara acak dari subset data pelatihan [12][15]. Setiap pohon keputusan akan menghasilkan prediksi, kemudian hasil prediksi tersebut digabungkan menggunakan metode *majority voting* untuk menentukan kelas akhir. Pendekatan ini bertujuan untuk meningkatkan akurasi prediksi serta mengurangi risiko *overfitting* yang sering terjadi pada model pohon keputusan tunggal.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214

Dalam penelitian ini, model *Random Forest* dilatih menggunakan data training dengan parameter *n_estimators* = 200 yang menunjukkan jumlah pohon keputusan yang dibangun dalam model, serta *max_depth* = 10 yang berfungsi untuk membatasi kedalaman maksimum pohon guna menjaga kompleksitas model. Setelah proses pelatihan selesai, model digunakan untuk melakukan prediksi terhadap data testing sehingga dapat diketahui kemampuan model dalam mengklasifikasikan preferensi jenis wisata berdasarkan atribut demografis responden.

2.5 Modeling Algoritma *XGBoost*

Selain menggunakan *Random Forest*, penelitian ini juga menerapkan algoritma *XGBoost (Extreme Gradient Boosting)* sebagai metode pembanding. *XGBoost* merupakan algoritma *ensemble learning* [13] berbasis teknik *gradient boosting* yang bekerja dengan membangun pohon keputusan secara bertahap, di mana setiap model baru berfungsi untuk memperbaiki kesalahan prediksi dari model sebelumnya [16][17]. Pendekatan ini memungkinkan model menghasilkan performa prediksi yang lebih optimal serta mampu menangani dataset dengan kompleksitas yang tinggi.

Pada penelitian ini, model *XGBoost* dilatih menggunakan data training dengan beberapa parameter utama, yaitu *n_estimators* = 200 yang menunjukkan jumlah iterasi pembentukan pohon keputusan, *max_depth* = 6 untuk mengatur kedalaman maksimum pohon, serta *learning_rate* = 0,1 yang berfungsi untuk mengontrol kontribusi setiap pohon terhadap model secara keseluruhan. Setelah model dilatih, proses prediksi dilakukan pada data testing untuk mengevaluasi kemampuan algoritma dalam mengklasifikasikan preferensi wisata responden. Hasil dari model *XGBoost* kemudian dibandingkan dengan model *Random Forest* untuk mengetahui algoritma yang memiliki kinerja terbaik dalam penelitian ini.

2.6 Evaluasi Model

Setelah model klasifikasi menggunakan algoritma *Random Forest* dan *XGBoost* dibangun, tahap selanjutnya adalah mengevaluasi performa kedua model tersebut. Evaluasi model bertujuan untuk mengetahui sejauh mana kemampuan algoritma dalam mengklasifikasikan preferensi jenis wisata berdasarkan atribut demografis responden. Proses evaluasi dilakukan dengan menggunakan data testing yang sebelumnya tidak digunakan pada tahap pelatihan model sehingga hasil pengujian dapat menggambarkan kemampuan generalisasi model terhadap data baru.

Dalam penelitian ini, performa model dievaluasi menggunakan beberapa metrik evaluasi yang umum digunakan dalam klasifikasi, yaitu *Accuracy, Precision, Recall, F1-Score*, serta *Confusion Matrix* [12].

Accuracy digunakan untuk mengukur tingkat ketepatan model dalam melakukan klasifikasi [12]. Nilai *accuracy* menunjukkan proporsi jumlah prediksi yang benar dibandingkan dengan keseluruhan data yang diuji. Semakin tinggi nilai *accuracy*, maka semakin baik kemampuan model dalam memprediksi kelas yang benar.

Precision mengukur tingkat ketepatan model dalam memprediksi suatu kelas tertentu [12]. *Precision* menunjukkan perbandingan antara jumlah prediksi yang benar pada suatu kelas dengan seluruh data yang diprediksi sebagai kelas tersebut [12]. Nilai precision yang tinggi menunjukkan bahwa model memiliki tingkat kesalahan prediksi yang rendah.

Recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh data yang benar-benar termasuk dalam suatu kelas tertentu [12]. *Recall* menunjukkan perbandingan antara jumlah data yang berhasil diprediksi dengan benar dengan seluruh data yang sebenarnya termasuk dalam kelas tersebut.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214

F1-Score merupakan nilai harmonisasi antara p*recision* dan *recall*. Metrik ini digunakan untuk memberikan gambaran yang lebih seimbang mengenai performa model, terutama ketika distribusi data antar kelas tidak seimbang [12].

Selain itu, evaluasi juga dilakukan menggunakan *Confusion Matrix*, yaitu sebuah matriks yang menggambarkan perbandingan antara nilai prediksi model dengan nilai sebenarnya [18], [19]. *Confusion Matrix* membantu dalam memahami pola kesalahan klasifikasi yang dilakukan oleh model, sehingga dapat diketahui kelas mana yang sering salah diprediksi.

Melalui serangkaian metrik evaluasi tersebut, performa model *Random Forest* dan XGBoost dapat dianalisis secara komprehensif. Hasil evaluasi ini kemudian digunakan untuk menentukan algoritma yang memiliki kinerja terbaik dalam mengklasifikasikan preferensi pengunjung destinasi wisata di Kabupaten Kendal.

2.7 Perbandingan Kinerja Algoritma

Tahap terakhir dalam penelitian ini adalah melakukan perbandingan kinerja algoritma *Random Forest* dan *XGBoost* untuk menentukan model yang memiliki performa terbaik dalam mengklasifikasikan preferensi jenis wisata pengunjung di Kabupaten Kendal. Perbandingan ini dilakukan berdasarkan hasil evaluasi model yang telah diperoleh pada tahap sebelumnya menggunakan beberapa metrik evaluasi seperti *accuracy, precision, recall, F1-score*, serta analisis *confusion matrix*.

Melalui perbandingan tersebut, dapat diketahui sejauh mana masing-masing algoritma mampu mempelajari pola hubungan antara atribut demografis responden—seperti jenis kelamin, usia, tingkat pendidikan, dan pekerjaan—dengan preferensi jenis wisata yang dipilih. Selain itu, analisis ini juga bertujuan untuk melihat perbedaan performa kedua algoritma dalam menangani data klasifikasi yang memiliki beberapa kategori kelas. Hasil dari perbandingan ini kemudian digunakan untuk mengidentifikasi algoritma yang paling efektif dan memiliki tingkat akurasi yang lebih baik dalam memprediksi preferensi wisatawan. Algoritma dengan performa terbaik diharapkan mampu memberikan model klasifikasi yang lebih akurat dan dapat digunakan sebagai dasar dalam memahami pola minat wisatawan, sehingga dapat menjadi referensi dalam pengambilan keputusan terkait pengembangan dan strategi promosi destinasi wisata di Kabupaten Kendal.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Penelitian ini melaksanakan proses pengumpulan data dengan pendekatan yang modern, efektif, serta adaptif terhadap perkembangan teknologi digital melalui penyebaran kuesioner secara daring. Pendekatan ini dipilih karena dinilai sebagai strategi yang efisien untuk mengumpulkan data dalam jumlah besar dalam waktu yang relatif singkat, sekaligus meminimalkan keterbatasan yang sering ditemui pada metode pengumpulan data konvensional. Selain itu, penggunaan teknologi berbasis internet juga mencerminkan upaya peneliti untuk menyesuaikan metode penelitian dengan tren digitalisasi masyarakat modern, di mana sebagian besar individu telah terbiasa menggunakan internet dan perangkat mobile dalam aktivitas sehari-hari.

Tautan tersebut kemudian disebarluaskan secara strategis melalui berbagai kanal komunikasi digital seperti media sosial dan aplikasi pesan instan. Penyebaran dilakukan melalui platform seperti *Facebook, Instagram*, dan *WhatsApp*, serta dibagikan pada berbagai grup komunitas lokal, forum daring, dan jaringan masyarakat yang memiliki keterkaitan dengan sektor pariwisata di Kabupaten Kendal. Strategi ini memungkinkan peneliti menjangkau responden secara lebih luas, baik masyarakat lokal maupun calon wisatawan dari luar daerah yang memiliki minat terhadap destinasi wisata di wilayah tersebut.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214

Data yang diperoleh dari kuesioner kemudian digunakan sebagai dataset penelitian untuk menganalisis preferensi pengunjung terhadap berbagai jenis destinasi wisata di Kabupaten Kendal. Selanjutnya, data tersebut diolah menggunakan pendekatan data mining dengan membandingkan performa algoritma klasifikasi, yaitu *Random Forest* dan *XGBoost*. Kedua algoritma tersebut digunakan untuk membangun model klasifikasi yang mampu mengidentifikasi pola preferensi pengunjung berdasarkan beberapa atribut seperti usia, asal pengunjung, jenis transportasi, harga tiket, fasilitas, dan tingkat aksesibilitas destinasi wisata.

Penggunaan metode pengumpulan data berbasis digital juga memberikan keuntungan dalam proses pengelolaan data, karena data yang terkumpul dapat tersimpan secara otomatis dalam sistem, sehingga mempermudah proses pengolahan, analisis, dan visualisasi data. Selain itu, pendekatan ini juga dapat meminimalkan potensi kesalahan manusia *(human error)* yang sering terjadi pada proses pencatatan data secara manual.

Tabel 1. Data Kuesioner Penelitian

Usia	Jenis Kelamin	Asal	Transportasi	Harga Tiket	Fasilitas	Akses	Jenis Wisata
<20	Perempuan	LuarKota	Bus	Sedang	Cukup	Sedang	Alam
20-30	Perempuan	Demak	Mobil	Sedang	Lengkap	Sulit	Alam
>40	Laki-laki	Kendal	Bus	Mahal	Lengkap	Sulit	Religi
>40	Perempuan	Kendal	Mobil	Murah	Kurang	Sedang	Religi
<20	Perempuan	Semarang	Motor	Mahal	Cukup	Sedang	Edukasi
>40	Perempuan	Kendal	Motor	Sedang	Lengkap	Sulit	Pantai
<20	Laki-laki	LuarKota	Bus	Murah	Kurang	Sulit	Pantai
>40	Perempuan	LuarKota	Mobil	Murah	Kurang	Sedang	Religi
30-40	Laki-laki	Semarang	Motor	Sedang	Cukup	Sedang	Religi
>40	Laki-laki	Batang	Motor	Murah	Cukup	Sulit	Religi
20-30	Perempuan	Semarang	Mobil	Sedang	Lengkap	Sedang	Religi

3.2 Preprocessing Data

Dataset yang diperoleh dari penyebaran kuesioner secara online berjumlah 670 responden, mencakup informasi demografis responden, preferensi terhadap destinasi wisata, serta faktor-faktor pendukung seperti harga tiket, aksesibilitas, dan fasilitas. Meskipun data telah tersimpan dalam bentuk terstruktur, tahapan preprocessing tetap diperlukan untuk memastikan kualitas, konsistensi, dan kelengkapan dataset sebelum digunakan dalam proses pemodelan.

Tahap pertama adalah pemeriksaan dan penanganan nilai yang hilang (missing value) pada setiap atribut dataset. *Missing value* dapat terjadi akibat responden yang tidak mengisi seluruh pertanyaan pada kuesioner. Apabila tidak ditangani, kondisi ini berpotensi menurunkan akurasi model dan menimbulkan bias pada hasil prediksi. Hasil pemeriksaan menunjukkan terdapat beberapa nilai kosong pada atribut tertentu, yang seluruhnya ditangani menggunakan metode imputasi modus.

Metode ini dipilih karena sebagian besar atribut dalam dataset bersifat kategorikal, seperti jenis kelamin, tingkat pendidikan, dan jenis pekerjaan, sehingga penggunaan modus dinilai mampu merepresentasikan karakteristik umum data tanpa mengubah distribusi secara signifikan. Selain itu, imputasi modus lebih diutamakan dibandingkan penghapusan baris data tidak lengkap, mengingat penghapusan dapat mengurangi jumlah sampel secara signifikan dan berpotensi menghilangkan informasi yang relevan untuk proses analisis.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214

3.3 Melakukan identifikasi dan penanganan terhadap data yang terduplikasi.

Tahap selanjutnya dalam preprocessing adalah pemeriksaan data duplikat, yaitu entri yang tercatat lebih dari satu kali dengan nilai atribut yang identik pada seluruh kolom. Duplikasi dapat terjadi akibat responden mengisi kuesioner lebih dari satu kali, kesalahan sistem penyimpanan, maupun redundansi saat penggabungan data.

Keberadaan data duplikat berpotensi menimbulkan bias pada hasil analisis karena informasi yang sama dihitung lebih dari satu kali, serta dapat meningkatkan beban komputasi secara tidak perlu. Data yang teridentifikasi duplikat kemudian dihapus, sehingga setiap baris dataset merepresentasikan satu responden yang unik. Langkah ini menghasilkan dataset yang lebih bersih dan representatif untuk digunakan pada tahap pemodelan selanjutnya.

3.4 Melakukan penanganan ketidakseimbangan data menggunakan metode *oversampling*

Ketidakseimbangan distribusi kelas *(class imbalance)* merupakan kondisi ketika jumlah data pada satu kelas jauh lebih besar dibandingkan kelas lainnya. Kondisi ini dapat menyebabkan model cenderung memihak kelas mayoritas, sehingga kemampuan mengenali kelas minoritas menjadi kurang optimal meskipun akurasi keseluruhan tampak tinggi.

Untuk mengatasi hal tersebut, diterapkan teknik *Synthetic Minority Oversampling Technique (SMOTE)* yang menghasilkan sampel sintetis baru pada kelas minoritas melalui interpolasi antar sampel terdekat, bukan sekadar menduplikasi data yang ada. *SMOTE* hanya diterapkan pada data pelatihan untuk menghindari data leakage, sehingga evaluasi model tetap dilakukan pada data asli yang tidak dimodifikasi. Hasil penerapan *SMOTE* menunjukkan distribusi kelas yang lebih seimbang sebagaimana ditunjukkan pada Gambar 7, sehingga model dapat mempelajari pola setiap kategori preferensi wisata secara lebih proporsional.

Jenis_Wisata	Sebelum_SMOTE	Sesudah_SMOTE
0	186	214
1	267	214
2	63	214
3	153	214

Gambar 2. Perbandingan Distribusi Kelas Sebelum dan Sesudah Oversampling SMOTE

3.5 Tahap akhir *preprocessing* data

Tahap akhir dalam proses data preprocessing merupakan langkah penting yang menandai selesainya seluruh proses persiapan data sebelum memasuki tahap pemodelan. Setelah berbagai tahapan sebelumnya dilakukan, seperti pembersihan data, penanganan *missing value*, penghapusan data duplikat, serta penanganan ketidakseimbangan kelas menggunakan teknik *oversampling*, seluruh hasil pengolahan tersebut kemudian digabungkan menjadi satu kesatuan dataset yang telah terstruktur dengan baik.

Dataset hasil pengolahan tersebut selanjutnya disimpan dalam sebuah tabel baru yang diberi nama *new_master_training*. Tabel ini merupakan versi akhir dari data pelatihan yang telah melalui proses pembersihan sehingga memiliki format yang sesuai untuk kebutuhan analisis dan pemodelan. Penyimpanan data dalam tabel baru juga bertujuan untuk menjaga keaslian data mentah *(raw data)* agar tetap tersimpan tanpa mengalami perubahan.

Dengan selesainya tahap ini, peneliti telah memperoleh dataset yang siap digunakan pada proses pemodelan. Data yang tersedia telah memenuhi aspek kualitas,

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214

konsistensi, dan kelengkapan sehingga dapat mendukung proses analisis secara optimal. Dataset tersebut kemudian digunakan pada tahap berikutnya, yaitu penerapan dan perbandingan kinerja algoritma *Random Forest* dan *XGBoost* dalam melakukan klasifikasi preferensi pengunjung terhadap destinasi wisata di Kabupaten Kendal.

Tabel 2. Hasil *Preprocessing Encoding* Numerik

Usia	Jenis Kelamin	Asal	Transportasi	Harga Tiket	Fasilitas	Akses	Jenis Wisata
0	0	0	2	0	0	0	1
3	0	2	3	1	0	2	3
0	0	4	2	0	1	1	1
1	1	0	2	0	1	0	1
0	1	2	2	1	0	1	1
0	1	2	2	0	0	1	0
0	0	0	2	1	1	0	1
3	1	1	2	1	1	2	3
2	0	4	2	0	0	1	0
3	1	4	3	1	1	2	3
1	0	0	2	1	0	0	1
1	1	3	2	0	1	0	0
1	1	4	3	0	1	1	1
1	1	2	2	1	0	0	1

3.6 Penerapan *Algoritma Random Forest* dan *XGBoost*

Dataset yang telah melalui tahap preprocessing dibagi menjadi data latih (80%) dan data uji (20%) menggunakan metode *train-test split* dengan parameter *stratify=y* untuk menjaga proporsi kelas tetap seimbang pada kedua subset. Teknik *SMOTE* kemudian diterapkan hanya pada data latih guna menghindari data *leakage*.

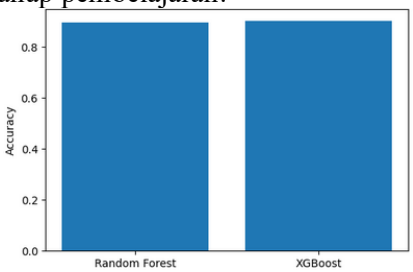
Dua algoritma *machine learning* diterapkan dalam penelitian ini. *Random Forest* membangun sejumlah pohon keputusan secara paralel dari berbagai subset data latih, kemudian menggabungkan hasilnya melalui *majority voting* untuk menghasilkan prediksi yang lebih stabil dan tahan terhadap *overfitting*. Sementara itu, *XGBoost* bekerja secara bertahap *(boosting)*, di mana setiap model baru berfungsi memperbaiki kesalahan prediksi model sebelumnya secara *iteratif*, sehingga mampu menangani kompleksitas pola data dengan lebih baik. Kedua model dilatih menggunakan parameter optimal hasil *GridSearchCV*, kemudian dievaluasi pada data uji untuk membandingkan kinerjanya dalam mengklasifikasikan preferensi wisata pengunjung di Kabupaten Kendal.

3.7 Evaluasi Model

Proses evaluasi dilakukan dengan menggunakan metrik akurasi serta classification report yang meliputi nilai presisi, *recall*, dan *f1-score*. Meskipun demikian, peneliti menyadari bahwa evaluasi yang dilakukan pada data yang sama dengan data pelatihan berpotensi menghasilkan nilai yang terlalu optimistis, karena model telah mengenali pola-pola yang terdapat dalam data tersebut. Oleh karena itu, meskipun metrik evaluasi tersebut dapat memberikan gambaran awal mengenai kinerja model, diperlukan pengujian lebih lanjut menggunakan data yang belum pernah digunakan sebelumnya agar diperoleh estimasi performa model yang lebih objektif dan representatif. Hasil pengujian model menunjukkan bahwa algoritma yang diterapkan memiliki tingkat akurasi yang tinggi.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214

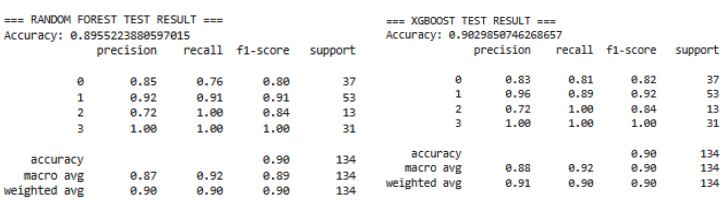
Model *Random Forest* memperoleh akurasi sebesar 89,55%, sedangkan model *XGBoost* menunjukkan performa yang lebih baik dengan akurasi sebesar 90%. Hal ini menunjukkan bahwa kedua model mampu melakukan prediksi dengan tingkat ketepatan yang baik, sehingga dapat memberikan rekomendasi yang cukup andal dalam konteks preferensi wisata. Kinerja tersebut mengindikasikan bahwa model mampu menangkap pola-pola yang kompleks dalam data. Pada *Random Forest*, kemampuan ini diperoleh melalui penerapan teknik ensemble learning berbasis bagging, yang menggabungkan beberapa model untuk meningkatkan stabilitas dan mengurangi risiko *overfitting*. Sementara itu, *XGBoost* mampu meningkatkan performa melalui pendekatan boosting, yang secara iteratif memperbaiki kesalahan prediksi pada setiap tahap pembelajaran.



Gambar 3. Perbandingan Akurasi Algoritma

Hasil evaluasi model menunjukkan bahwa kedua algoritma yang digunakan memiliki performa yang baik dalam mengklasifikasikan preferensi wisata pengunjung. Model *Random Forest* memperoleh tingkat akurasi sebesar 89,55%, sedangkan model *XGBoost* menunjukkan performa yang lebih tinggi dengan akurasi sebesar 90%. Hal ini mengindikasikan bahwa algoritma *XGBoost* memiliki kemampuan yang lebih optimal dalam melakukan klasifikasi dibandingkan *Random Forest* pada dataset yang digunakan.

Berdasarkan *classification report*, model *Random Forest* menunjukkan nilai *presisi* dan *recall* yang cukup tinggi pada sebagian besar kelas, khususnya pada kelas 1 dengan nilai *presisi* sebesar 0,95 dan *recall* sebesar 0,89, serta kelas 3 yang mencapai nilai sempurna (1,00) untuk *presisi* dan *recall*. Namun, pada kelas 2 masih terdapat penurunan performa dengan nilai presisi sebesar 0,69 meskipun *recall* mencapai 0,85, yang menunjukkan bahwa model masih mengalami kesulitan dalam mengklasifikasikan kelas tersebut secara konsisten.Sementara itu, model *XGBoost* menunjukkan peningkatan performa yang lebih baik dibandingkan *Random Forest*. Hal ini terlihat dari nilai presisi dan *recall* yang lebih tinggi pada sebagian besar kelas, khususnya pada kelas 1 dengan presisi sebesar 0,96 dan *recall* sebesar 0,93. Selain itu, kelas 3 tetap menunjukkan performa sempurna dengan nilai presisi dan recall sebesar 1,00. Meskipun demikian, performa pada kelas 2 masih relatif sama dengan *Random Forest*, yang menunjukkan bahwa kelas tersebut memiliki tingkat kompleksitas yang lebih tinggi dibandingkan kelas lainnya.Secara keseluruhan, kedua model mampu memberikan hasil klasifikasi yang cukup baik, namun algoritma *XGBoost* lebih unggul dalam hal akurasi dan konsistensi prediksi. Oleh karena itu, *XGBoost* dapat direkomendasikan sebagai model yang lebih efektif dalam analisis preferensi wisata pengunjung pada penelitian ini.



=== RANDOM FOREST TEST RESULT ===
Accuracy: 0.8955223880597015

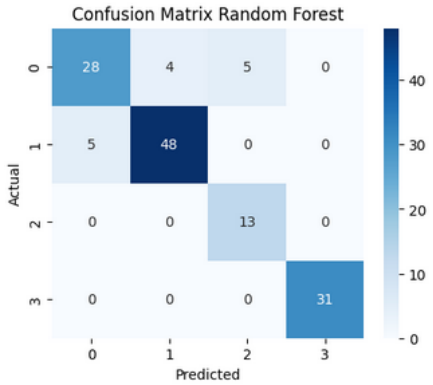
	precision	recall	f1-score	support
0	0.85	0.76	0.80	37
1	0.92	0.91	0.91	53
2	0.72	1.00	0.84	13
3	1.00	1.00	1.00	31
accuracy			0.90	134
macro avg	0.87	0.92	0.89	134
weighted avg	0.90	0.90	0.90	134

=== XGBOOST TEST RESULT ===
Accuracy: 0.9029850746268657

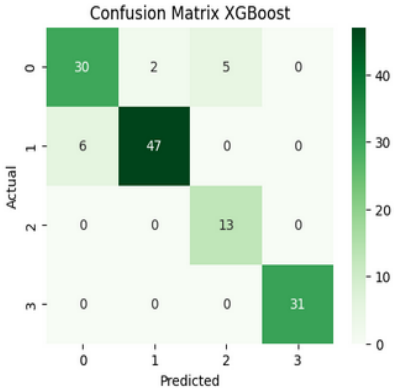
	precision	recall	f1-score	support
0	0.83	0.81	0.82	37
1	0.96	0.89	0.92	53
2	0.72	1.00	0.84	13
3	1.00	1.00	1.00	31
accuracy			0.90	134
macro avg	0.88	0.92	0.90	134
weighted avg	0.91	0.90	0.90	134

Gambar 4. Hasil Nilai Akurasi Algoritma

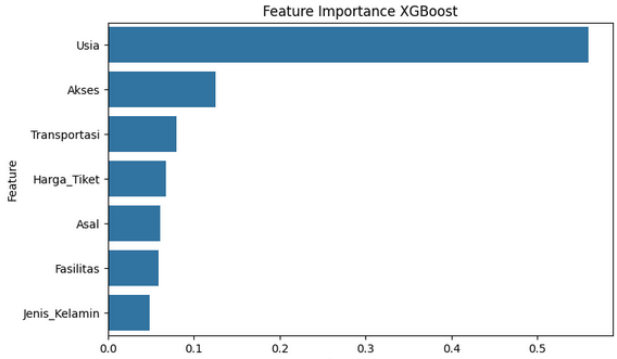
e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214



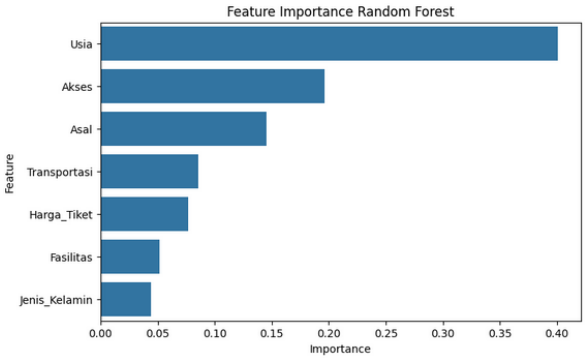
Gambar 5. *Confusion Matrix Random Forest*



Gambar 6. *Confusion Matrix XGBoost*



Gambar 7. *Feature Important XGBoost*



Gambar 8. *Feature Important Random Forest*

Gambar menunjukkan hasil analisis *feature importance* pada model *XGBoost* dan *Random Forest* dalam menentukan faktor-faktor yang memengaruhi preferensi wisata pengunjung. Berdasarkan kedua grafik tersebut, dapat diketahui bahwa atribut Usia merupakan variabel yang paling dominan dalam kedua model, dengan nilai kepentingan yang jauh lebih tinggi dibandingkan atribut lainnya. Hal ini menunjukkan bahwa faktor usia memiliki pengaruh yang sangat signifikan dalam menentukan pilihan jenis wisata.

Pada model *XGBoost*, terlihat bahwa kontribusi atribut selain usia relatif kecil dan cenderung merata, seperti Harga Tiket, Transportasi, Akses, Asal, Fasilitas, dan Jenis Kelamin. Hal ini mengindikasikan bahwa *XGBoost* lebih berfokus pada atribut utama (usia) dalam proses pengambilan keputusan, sementara atribut lainnya berperan sebagai faktor pendukung dengan pengaruh yang tidak terlalu dominan. Sementara itu, pada model *Random Forest*, distribusi kepentingan fitur terlihat lebih bervariasi.

Selain Usia sebagai faktor utama, atribut lain seperti Akses dan Asal juga memiliki kontribusi yang cukup besar dalam memengaruhi hasil prediksi. Hal ini menunjukkan bahwa *Random Forest* mampu menangkap hubungan yang lebih kompleks antar variabel dan memanfaatkan lebih banyak fitur dalam proses klasifikasi dibandingkan *XGBoost*. Secara keseluruhan, kedua model menunjukkan bahwa usia merupakan faktor paling berpengaruh dalam menentukan preferensi wisata, diikuti oleh atribut seperti Akses, Asal, dan Harga Tiket. Temuan ini mengindikasikan bahwa segmentasi berdasarkan usia menjadi aspek penting dalam perencanaan strategi pengembangan dan promosi destinasi wisata, sementara faktor aksesibilitas dan karakteristik asal pengunjung juga perlu diperhatikan untuk meningkatkan daya tarik wisata.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214

```
=== CROSS VALIDATION RESULT ===
RF Mean: 0.9073616878015937 | Std: 0.021819525682760216
XGB Mean: 0.9118617439120189 | Std: 0.032408507354526495
```

Gambar 9. Nilai *Cross Validation*

Berdasarkan hasil pengujian menggunakan metode *Stratified K-Fold Cross Validation* (k=5), diperoleh nilai rata-rata akurasi untuk algoritma *Random Forest* sebesar 0,9073 atau 90,73%, dengan nilai standar deviasi sebesar 0,0218. Nilai rata-rata akurasi tersebut menunjukkan bahwa model *Random Forest* memiliki kemampuan yang sangat baik dalam mengklasifikasikan preferensi wisata pengunjung, dengan tingkat ketepatan yang tinggi secara konsisten pada berbagai pembagian data. Sementara itu, nilai standar deviasi yang relatif kecil (±2,18%) mengindikasikan bahwa performa model stabil dan tidak mengalami fluktuasi yang signifikan antar fold. Hal ini menunjukkan bahwa model memiliki kemampuan generalisasi yang baik dan tidak bergantung pada satu subset data tertentu. Dengan demikian, dapat disimpulkan bahwa algoritma *Random Forest* tidak hanya memberikan akurasi yang tinggi, tetapi juga menunjukkan konsistensi performa yang baik, sehingga layak digunakan dalam analisis klasifikasi preferensi wisata pada dataset yang digunakan dalam penelitian ini.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, diperoleh beberapa temuan penting yang menggambarkan keberhasilan sekaligus kontribusi penelitian ini dalam menganalisis preferensi pengunjung terhadap destinasi wisata di Kabupaten Kendal.

4.1. Pengumpulan data efektif

Penelitian ini berhasil menerapkan metode pengumpulan data yang efektif melalui penyebaran kuesioner secara daring menggunakan *platform Google Forms*. Pendekatan tersebut memungkinkan peneliti menjangkau responden dalam jumlah yang lebih luas tanpa adanya keterbatasan wilayah, serta mempermudah proses pengumpulan data secara cepat, sistematis, dan terstruktur. Dalam penelitian ini, sebanyak 670 responden turut berpartisipasi yang terdiri dari masyarakat dengan latar belakang demografis yang beragam. Data yang dikumpulkan mencakup beberapa atribut penting, antara lain jenis kelamin, usia, asal daerah, transportasi yang digunakan, harga tiket, fasilitas, serta akses menuju lokasi wisata. Informasi yang diperoleh dari responden tersebut dinilai sangat relevan dalam mengidentifikasi pola serta preferensi pengunjung terhadap berbagai jenis destinasi wisata yang terdapat di Kabupaten Kendal, seperti wisata alam, wisata religi, wisata pantai, hingga wisata edukasi.

4.2 *Preprocessing* data yang komprehensif

Tahapan preprocessing dilakukan secara sistematis untuk memastikan kualitas data yang digunakan dalam proses pemodelan berada dalam kondisi yang optimal. Proses ini meliputi pemeriksaan serta penanganan nilai yang hilang *(missing value)* menggunakan metode modus, penghapusan data duplikat untuk menghindari potensi bias dalam analisis, serta validasi kategori pada atribut jenis kelamin guna menjaga konsistensi data. Selanjutnya, data kategorikal diubah ke dalam bentuk numerik menggunakan teknik *Label Encoding* agar dapat diproses oleh algoritma *machine learning*. Selain itu, diterapkan pula teknik *oversampling* untuk mengatasi ketidakseimbangan distribusi kelas, sehingga model dapat mempelajari pola data secara lebih seimbang dan menghasilkan prediksi yang lebih akurat terhadap setiap kategori preferensi wisata.

4.3 Implementasi Algoritma *Random Forest* dan *XGBoost*

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma *Random Forest* dan *XGBoost* mampu digunakan secara efektif dalam mengklasifikasikan preferensi wisata pengunjung. Algoritma *Random Forest* yang menggunakan teknik *ensemble learning* berbasis bagging mampu menghasilkan model yang stabil melalui penggabungan beberapa pohon keputusan, sedangkan algoritma *XGBoost* yang berbasis *boosting* mampu meningkatkan performa model dengan memperbaiki kesalahan prediksi secara bertahap pada setiap iterasi.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214

Hasil pengujian menunjukkan bahwa algoritma Random Forest memperoleh tingkat akurasi sebesar 89.55%, sedangkan algoritma *XGBoost* menghasilkan akurasi yang lebih tinggi yaitu sebesar 90.30%. Hal ini menunjukkan bahwa algoritma *XGBoost* memiliki kinerja yang lebih optimal dalam mengklasifikasikan preferensi wisata pengunjung dibandingkan *Random Forest* pada dataset yang digunakan.

Selain itu, kedua algoritma mampu mengidentifikasi atribut-atribut yang memiliki pengaruh signifikan terhadap pemilihan jenis destinasi wisata, seperti harga tiket, akses, fasilitas, usia, dan asal pengunjung. Atribut-atribut tersebut menjadi faktor penting dalam menentukan preferensi wisata, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan.

Dengan demikian, hasil penelitian ini diharapkan dapat memberikan kontribusi baik secara akademik maupun praktis, khususnya bagi pemerintah daerah dan pengelola destinasi wisata, dalam merumuskan strategi pengembangan dan promosi pariwisata yang lebih efektif, terarah, dan berbasis data. Selain itu, penelitian ini juga dapat menjadi referensi bagi penelitian selanjutnya dalam pengembangan model klasifikasi yang lebih akurat dengan memanfaatkan data yang lebih besar dan variatif.

DAFTAR PUSTAKA

[1] U. S. Sri, A. A. Syaifudin, dan S. Wahyuni, "STRATEGI PROMOSI PARIWISATA : STUDI ATAS MARKETING," *J. Kepariwisataan Indones. J. Penelit. dan Pengemb. Kepariwisataan Indones.*, vol. 14, no. 2, hal. 45–58, 2022.

[2] M. Rizaludin *et al.*, "Analisis Preferensi Pengunjung Wisata Kabupaten Pekalongan Menggunakan Algoritma Decision Tree," *Techno.Com J. Teknol. Inf.*, vol. 25, no. 1, hal. 30–46, 2026.

[3] G. Joseph, L. Jason, C. Putra, P. Agusia, dan M. Y. Utama, "PERBANDINGAN NAIVE BAYES , RANDOM FOREST , XGBOOST UNTUK KLASIFIKASI MUTU AIR".

[4] S. Mahasiswa, "Perbandingan metode klasifikasi random forest dan support vector machine dalam memprediksi capaian studi mahasiswa," vol. 9, no. 4, hal. 1821–1833, 2024.

[5] S. Amaliah dan M. Nusrang, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi Di Kedai Kopi Konijiwa Bantaeng," vol. 4, no. 2, hal. 121–127, 2022, doi: 10.35580/variansiunm31.

[6] F. P. Arifianti dan A. Salam, "XGBoost and Random Forest Optimization using SMOTE to Classify Air Quality," *Adv. Sustain. Sci. Eng. Technol.*, vol. 6, no. 1, hal. 1–8, 2024, doi: https://doi.org/10.26877/asset.v6i1.17951 XGBoost.

[7] N. A. Siagian, S. P. Sipayung, A. Rikki, dan N. Marbun, "Integrating SMOTE with XGBoost for Robust Classification on Imbalanced Datasets : A Dual-Domain Evaluation," *Sink. J. dan Penelit. Tek. Inform.*, vol. 9, no. 3, hal. 1094–1107, 2025, doi: https://doi.org/10.33395/sinkron.v9i3.15029.

[8] A. Prayuda dan I. Pratama, "PREDIKSI JUMLAH KEDATANGAN WISATAWAN MANCANEGARA DI INDONESIA BERDASARKAN PINTU MASUK KEDATANGAN UDARA," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 9, no. 2, hal. 232–241, 2024.

[9] Y. Desnelita, N. Nasution, L. Suryati, dan F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang Impact of SMOTE on Random Forest Classifier Performance based on Imbalanced Data," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, hal. 515–526, 2022, doi: 10.30812/matrik.v21i3.1726.

[10] D. Sangaji dan T. Sutabri, "E ISSN : 2809-4069 Analisis XGBoost dan Random Forest untuk Prediksi Curah Hujan dalam Mendukung Mitigasi Karhutla," *E-JURNAL JUSITI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 5, no. 1, hal. 13–18, 2025, doi: https://doi.org/10.55382/jurnalpustakaai.v5i1.905.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214

[11] K. J. Waciko, L. A. Susanti, dan R. Nur, "Forecasting Tourist Arrivals in Bali : A Grid Search-Tuned Comparative Study of Random Forest , XGBoost , and a Hybrid RF-XGBoost Model," *Inferensi J. Stat.*, vol. 8, no. 3, hal. 251–261, 2025.

[12] E. Jurnal, "Klasifikasi Preferensi Destinasi Wisata Gunung dan Pantai Menggunakan Metode Deep Learning," *ELKOM J. Elektron. dan Komput.*, vol. 18, no. 2, hal. 285–291, 2025.

[13] S. F. A. Khansa, N. Ulinnuha, dan W. D. Utami, "Grid Search and Random Search Hyperparameter Tuning Optimization in Xgboost Algorithm for Parkinson'S Disease Classification," *Barekeng*, vol. 19, no. 3, hal. 1609–1624, 2025, doi: 10.30598/barekengvol19iss3pp1609-1624.

[14] Y. Handayani, D. Setiawan, T. Hidayat, dan T. M. Waluyo, "An XGBoost-Driven Intelligent Classification Model for Textile Product Quality Eligibility : A Case Study at PT ABC Textile," *Techno.Com*, vol. 25, no. 1, hal. 211–222, 2026.

[15] F. Di dan K. Gayo, "Prediksi potensi wisata menggunakan algoritma random forest di kabupaten gayo lues 1) 1,2,3)," vol. 10, no. 2, hal. 741–751, 2025.

[16] W. Lokal dan D. I. Lombok, "Penerapan algoritma xgboost dalam penentuan potensi sektor wisata lokal di lombok timur," *J. Inform. dan Rekayasa Elektron.*, vol. 03, hal. 155–160, 2025.

[17] A. A. Syahputra dan R. E. Saputro, "Application of the XGBoost Model with Hyperparameter Tuning for Industry Classification for Job Applicants," *Sinkron*, vol. 8, no. 3, hal. 1920–1931, 2024, doi: 10.33395/sinkron.v8i3.13840.

[18] Y. Handayani, "Development of optimization formula for Neural Network-Based Automatic Control System in manufacturing industry," *J. Appl. Eng. Technol. Sci.*, vol. 7, no. 4, hal. 302–308, 2024.

[19] Y. Handayani dan A. R. Hakim, "Classification of Naive Bayes Algorithm on Dengue Hemorrhagic Fever and Typhoid Fever Based on Hematology Results," *J. Appl. Intell. Syst.*, vol. 8, no. 1, hal. 94–99, 2023.

e-ISSN : 2356-2579 | p-ISSN : 1412-2693, DOI : 10.62411/tc.v25i3.16214