

Model Machine Learning untuk Rekomendasi Hotel Berdasarkan Profil dan Preferensi Wisatawan

Candra Agustina^{1*}, Angga Ardiansyah², Raja Sabarudin³, Lisnawanty⁴, Eka Rahmawati⁵

^{1,2,3}Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika
Jakarta, Indonesia

⁴Informatika, Universitas Bina Sarana Informatika
Jakarta, Indonesia

⁵Sistem Informasi, Universitas Bina Sarana Informatika
Jakarta, Indonesia

Correspondence e-mail: candra.caa@bsi.ac.id

ABSTRAK

Perkembangan platform pariwisata digital mendorong kebutuhan terhadap sistem rekomendasi hotel yang lebih personal dan berbasis data. Namun, sistem rekomendasi yang bergantung pada popularitas, rating, atau riwayat interaksi pengguna kurang optimal ketika data historis belum tersedia. Penelitian ini mengusulkan pendekatan *classification-based recommendation* untuk memprediksi kategori pilihan hotel wisatawan berdasarkan profil demografis, pola perjalanan, faktor ekonomi, dan preferensi fasilitas. Dataset diperoleh dari 290 responden dengan variabel meliputi usia, pendapatan, status pernikahan, pekerjaan, rekan perjalanan, lama liburan, besaran biaya liburan, jarak, fasilitas penginapan, dan layanan internet. Tiga algoritma klasifikasi digunakan, yaitu *Decision Tree*, *Gaussian Naive Bayes*, dan *Random Forest*. Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Hasil penelitian menunjukkan bahwa *Random Forest* memperoleh performa terbaik dengan *accuracy* sebesar 68,97% dan *F1-score* sebesar 67,87%. Analisis *feature importance* menunjukkan bahwa lama liburan, pekerjaan, besaran biaya liburan, jarak, dan pendapatan menjadi faktor dominan dalam pemilihan hotel. Temuan ini menunjukkan bahwa pendekatan klasifikasi dapat digunakan sebagai dasar awal rekomendasi hotel yang lebih personal, terutama pada kondisi ketika data rating atau riwayat interaksi pengguna belum tersedia.

Kata Kunci: klasifikasi, *machine learning*, preferensi wisatawan, *random forest*, rekomendasi hotel

ABSTRACT

The development of digital tourism platforms has increased the need for more personalized and data-driven hotel recommendation systems. However, recommendation systems that rely on popularity, ratings, or user interaction history are less optimal when historical data are not yet available. This study proposes a *classification-based recommendation* approach to predict tourists' hotel choice categories based on demographic profiles, travel patterns, economic factors, and facility preferences. The dataset was obtained from 290 respondents and included variables such as age, income, marital status, occupation, travel companions, length of vacation, vacation budget, distance, accommodation facilities, and internet services. Three classification algorithms were applied: *Decision Tree*, *Gaussian Naive Bayes*, and *Random Forest*. The models were evaluated using *accuracy*, *precision*, *recall*, *F1-score*, and *confusion matrix*. The results show that *Random Forest* achieved the best performance, with an *accuracy* of 68.97% and an *F1-score* of 67.87%. *Feature importance* analysis indicates that length of vacation, occupation, vacation budget, distance, and income are the dominant factors influencing hotel selection. These findings suggest that the *classification-based* approach can serve as an initial basis for more personalized hotel recommendations, particularly when rating data or user interaction history are not yet available.

Keywords: *classification*, *hotel recommendation*, *machine learning*, *random forest*, *tourist preferences*



1. PENDAHULUAN

Transformasi digital dalam industri pariwisata telah mengubah pola perilaku wisatawan dalam merencanakan perjalanan, termasuk dalam proses pemilihan akomodasi. Platform pariwisata digital kini menyediakan berbagai informasi mengenai hotel, harga, fasilitas, lokasi, serta ulasan pengguna yang dapat diakses secara real-time (Topsakal et al., 2021; Ernawati et al., 2023). Namun, sistem pencarian dan rekomendasi yang tersedia pada banyak platform masih cenderung bersifat umum, misalnya berdasarkan popularitas, harga, lokasi, atau ulasan pengguna secara agregat. Kondisi tersebut menyebabkan rekomendasi yang diberikan belum sepenuhnya mempertimbangkan karakteristik individual wisatawan, sehingga hasil rekomendasi dapat kurang sesuai dengan kebutuhan spesifik pengguna.

Pemilihan hotel merupakan salah satu keputusan strategis dalam perjalanan wisata karena berkaitan dengan kenyamanan, efisiensi biaya, aksesibilitas, serta kesesuaian dengan tujuan perjalanan. Wisatawan dengan profil demografis berbeda, seperti usia, tingkat pendapatan, status pernikahan, maupun pekerjaan, cenderung memiliki preferensi akomodasi yang berbeda pula. Selain itu, pola perjalanan seperti lama liburan, rekan perjalanan (keluarga, teman, atau individu), serta besaran biaya perjalanan juga memengaruhi keputusan pemilihan hotel. Faktor preferensi fasilitas seperti ketersediaan internet, akses lokasi, serta layanan pendukung turut menjadi variabel penting dalam menentukan pilihan akomodasi (Dewiyanti et al., 2021; Perdana et al., 2025).

Beberapa penelitian sebelumnya telah membahas pemanfaatan *machine learning* dalam bidang pariwisata dan perhotelan, seperti prediksi permintaan hotel, analisis ulasan pengguna, dan sistem rekomendasi berbasis rating. Namun, sebagian besar pendekatan rekomendasi masih bergantung pada data rating, ulasan, atau riwayat interaksi pengguna. Pendekatan tersebut efektif apabila data historis tersedia dalam jumlah memadai, tetapi menjadi kurang optimal pada kondisi awal pengembangan sistem ketika data interaksi pengguna masih terbatas (Dowlut & Gobin-Rahimbux, 2023; N. Li et al., 2025). Dengan demikian, masih terdapat celah penelitian dalam pengembangan model rekomendasi hotel yang memanfaatkan data profil demografis, pola perjalanan, faktor ekonomi, dan preferensi fasilitas sebagai dasar prediksi, terutama pada kondisi ketika data rating atau riwayat interaksi pengguna belum tersedia. Oleh karena itu, *machine learning* menjadi pendekatan yang relevan karena mampu mempelajari pola keterkaitan antarvariabel dan menghasilkan prediksi yang dapat digunakan sebagai dasar rekomendasi hotel yang lebih personal (Solano-Barliza et al., 2024).

Dalam konteks *supervised learning*, metode klasifikasi dapat digunakan untuk memprediksi kategori pilihan hotel wisatawan berdasarkan atribut profil demografis, pola perjalanan, faktor ekonomi, dan preferensi fasilitas. Model dilatih menggunakan dataset yang memiliki label kategori pilihan hotel wisatawan, sehingga dapat mengidentifikasi pola hubungan antara variabel prediktor dan kategori hotel yang dipilih (Pamungkas et al., 2019). Pendekatan ini tidak hanya menghasilkan prediksi, tetapi juga memungkinkan analisis faktor-faktor dominan yang memengaruhi keputusan wisatawan melalui evaluasi *feature importance*.

Berdasarkan celah tersebut, penelitian ini mengusulkan pendekatan *classification-based recommendation* untuk membangun model rekomendasi hotel berdasarkan profil demografis, pola perjalanan, faktor ekonomi, dan preferensi fasilitas wisatawan. Pendekatan klasifikasi digunakan untuk memprediksi kategori pilihan hotel wisatawan, kemudian hasil prediksi dari model terbaik digunakan sebagai dasar dalam penyusunan rekomendasi hotel yang lebih personal. Dalam penelitian ini, tiga algoritma klasifikasi digunakan, yaitu *Decision Tree*, *Gaussian Naive Bayes*, dan *Random Forest*. Ketiga algoritma tersebut dipilih karena mewakili pendekatan yang berbeda, yaitu pohon keputusan, probabilistik, dan *ensemble learning*, sehingga dapat digunakan untuk membandingkan kemampuan masing-masing model dalam menangkap pola hubungan antar variabel. Hasil prediksi dari model terbaik selanjutnya dapat dimanfaatkan sebagai dasar dalam memberikan rekomendasi hotel yang lebih personal dan relevan bagi wisatawan (Ayu Firnanda et al., 2025; Hadi et al., 2025; Novita Sari et al., 2025; Suryani, 2023).

Pendekatan ini berbeda dengan metode *collaborative filtering* yang memerlukan data interaksi pengguna dan rating historis dalam menghasilkan rekomendasi (Y. Li, 2025; Tejashri et al., 2025). Penelitian ini mengadopsi pendekatan *classification-based recommendation* yang lebih sesuai dengan karakteristik dataset berbasis profil dan preferensi individu (Alrassi et al., 2025; Gadkari, 2023). Dalam pendekatan ini, model klasifikasi mempelajari hubungan antara karakteristik wisatawan, pola perjalanan, faktor ekonomi, dan preferensi fasilitas terhadap kategori pilihan hotel wisatawan. Hasil prediksi dari

model terbaik selanjutnya digunakan sebagai dasar dalam memberikan rekomendasi hotel yang lebih personal dan relevan, terutama pada kondisi ketika data rating atau riwayat interaksi pengguna belum tersedia dalam jumlah besar.

Kontribusi utama penelitian ini adalah pengembangan dan evaluasi model rekomendasi hotel berbasis klasifikasi dengan memanfaatkan kombinasi profil demografis, pola perjalanan, faktor ekonomi, dan preferensi fasilitas sebagai variabel prediktor untuk menentukan kategori pilihan hotel wisatawan. Secara ilmiah, penelitian ini memperluas penerapan *classification-based recommendation* dalam konteks pemilihan hotel, khususnya ketika data rating atau riwayat interaksi pengguna belum tersedia. Secara praktis, model yang dihasilkan dapat menjadi dasar awal bagi pengembangan sistem rekomendasi hotel yang lebih personal dan relevan dengan karakteristik wisatawan. Dengan demikian, tujuan penelitian ini adalah membangun dan mengevaluasi model *machine learning* untuk memprediksi kategori pilihan hotel wisatawan berdasarkan profil dan preferensi wisatawan.

2. METODE PENELITIAN

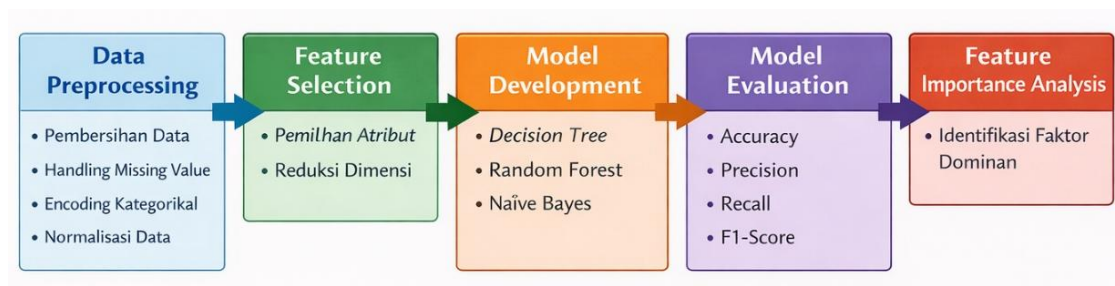
Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen komputasional menggunakan metode *machine learning* berbasis *supervised learning*. Pendekatan ini digunakan karena dataset memiliki variabel target berupa kategori pilihan hotel wisatawan. Eksperimen dilakukan dengan membandingkan beberapa algoritma klasifikasi untuk mengetahui model yang memiliki performa terbaik dalam memprediksi kategori pilihan hotel berdasarkan profil demografis, pola perjalanan, faktor ekonomi, dan preferensi fasilitas wisatawan.

Dataset yang digunakan dalam penelitian ini diperoleh melalui kuesioner kepada 290 responden wisatawan. Data tersebut merepresentasikan profil demografis, pola perjalanan, faktor ekonomi, serta preferensi fasilitas wisatawan dalam memilih hotel. Variabel prediktor yang digunakan meliputi usia, pendapatan, status pernikahan, pekerjaan, rekan perjalanan, lama liburan, besaran biaya liburan, jarak, fasilitas penginapan, dan layanan internet.

Variabel yang digunakan dalam penelitian ini dikelompokkan ke dalam tiga kategori utama, yaitu profil demografis, pola perjalanan, serta preferensi fasilitas dan layanan. Profil demografis mencakup usia, pendapatan, status pernikahan, dan pekerjaan. Pola perjalanan mencakup rekan perjalanan, lama liburan, besaran biaya liburan, dan jarak. Sementara itu, preferensi fasilitas dan layanan mencakup fasilitas penginapan dan layanan internet.

Variabel target dalam penelitian ini adalah kategori pilihan hotel wisatawan yang digunakan sebagai label kelas dalam proses klasifikasi. Kategori tersebut merepresentasikan kelompok pilihan hotel yang diprediksi oleh model berdasarkan karakteristik dan preferensi wisatawan. Dengan demikian, model yang dibangun bertujuan untuk mempelajari pola hubungan antara variabel prediktor dan kategori pilihan hotel wisatawan.

Proses penelitian dalam pengembangan model rekomendasi dilakukan melalui beberapa tahapan sistematis sebagaimana ditunjukkan pada Gambar 1. Gambar 1 menunjukkan alur penelitian yang dimulai dari tahap *data preprocessing* untuk membersihkan dan menyiapkan data, dilanjutkan dengan *feature selection* untuk memilih atribut yang relevan. Selanjutnya dilakukan tahap *model development* menggunakan beberapa algoritma klasifikasi, kemudian *model evaluation* untuk mengukur kinerja model, dan diakhiri dengan analisis *feature importance* guna mengidentifikasi faktor dominan dalam keputusan pemilihan hotel.



Gambar 1. Tahapan Penelitian Model Rekomendasi Hotel Berbasis Machine Learning

Pada tahap *preprocessing*, data diperiksa untuk mengidentifikasi nilai kosong, data duplikat, kesalahan penulisan, dan ketidakkonsistenan kategori. Data yang tidak lengkap atau tidak sesuai dengan kebutuhan analisis dibersihkan agar tidak memengaruhi proses pemodelan. Variabel kategorikal seperti status pernikahan, pekerjaan, rekan perjalanan, fasilitas penginapan, dan layanan internet dikonversi ke dalam bentuk numerik menggunakan teknik encoding. Sementara itu, variabel numerik seperti usia, pendapatan, lama liburan, besaran biaya liburan, dan jarak dipertahankan dalam format numerik agar dapat diproses oleh algoritma *machine learning*. Setelah proses pembersihan dan encoding selesai, dataset dibagi menjadi data latih dan data uji dengan proporsi 80:20.

Pada tahap selanjutnya *feature selection*. *Feature selection* dilakukan berdasarkan relevansi variabel terhadap tujuan penelitian dan dukungan literatur mengenai faktor-faktor yang memengaruhi pemilihan hotel. Variabel yang digunakan mencakup profil demografis, pola perjalanan, faktor ekonomi, dan preferensi fasilitas. Seluruh variabel tersebut dipertahankan sebagai prediktor karena dinilai relevan dalam menjelaskan kategori pilihan hotel wisatawan. Tahap ini bertujuan untuk memastikan bahwa fitur yang digunakan sesuai dengan tujuan penelitian dan dapat membantu interpretasi faktor-faktor yang memengaruhi kategori pilihan hotel wisatawan.

Pada tahap model *development*, data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengevaluasi kemampuan model dalam memprediksi data yang belum pernah digunakan pada proses pelatihan. Tiga algoritma yang digunakan adalah *Decision Tree*, *Gaussian Naive Bayes*, dan *Random Forest*. *Decision Tree* dipilih karena mampu menghasilkan aturan klasifikasi yang mudah diinterpretasikan, *Gaussian Naive Bayes* digunakan sebagai model probabilistik yang sederhana dan efisien, sedangkan *Random Forest* digunakan sebagai model *ensemble* yang mampu meningkatkan stabilitas prediksi melalui penggabungan beberapa pohon keputusan. Implementasi model dilakukan menggunakan bahasa pemrograman *Python* dengan library *Scikit-learn*. *Decision Tree* dan *Gaussian Naive Bayes* menggunakan konfigurasi dasar, sedangkan *Random Forest* dijalankan dengan jumlah estimator sebanyak 100 pohon keputusan. Nilai *random_state* ditetapkan untuk menjaga konsistensi dan replikasi hasil eksperimen.

Tahap evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. *Accuracy* digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data uji. *Precision* digunakan untuk mengukur ketepatan model dalam memprediksi suatu kelas, sedangkan *recall* digunakan untuk mengukur kemampuan model dalam mengenali data pada kelas yang sebenarnya. *F1-score* digunakan sebagai ukuran keseimbangan antara *precision* dan *recall*. *Confusion matrix* digunakan untuk melihat distribusi prediksi benar dan salah pada setiap kelas target. Penelitian ini menggunakan skema evaluasi *train-test split* dengan proporsi 80:20. *Cross-validation* belum diterapkan dalam penelitian ini dan menjadi salah satu arah pengembangan pada penelitian selanjutnya untuk memperoleh estimasi performa model yang lebih stabil.

Analisis *feature importance* dilakukan pada model *Random Forest* karena model tersebut memperoleh performa terbaik. *Feature importance* digunakan untuk mengidentifikasi kontribusi relatif setiap variabel prediktor terhadap proses klasifikasi. Nilai *feature importance* yang lebih tinggi menunjukkan bahwa variabel tersebut memiliki kontribusi lebih besar dalam membedakan kategori pilihan hotel wisatawan.

Model dikembangkan menggunakan bahasa pemrograman *Python* dengan bantuan library *Scikit-learn*. Proses implementasi meliputi pembagian data latih dan data uji, pelatihan model, pengujian model, evaluasi performa, serta analisis *feature importance*.

Algoritma *Decision Tree* diimplementasikan menggunakan pendekatan pohon keputusan untuk membentuk aturan klasifikasi berdasarkan variabel prediktor. *Gaussian Naive Bayes* diimplementasikan sebagai model probabilistik dengan asumsi distribusi *Gaussian* pada fitur numerik. *Random Forest* diimplementasikan sebagai metode *ensemble* yang membangun sejumlah pohon keputusan dan menggabungkan hasil prediksinya untuk memperoleh keputusan akhir. Parameter dasar yang digunakan dalam proses pemodelan disesuaikan dengan konfigurasi standar *Scikit-learn*, sedangkan *Random Forest* digunakan untuk memperoleh nilai *feature importance* dari setiap variabel prediktor.

Seluruh model dievaluasi menggunakan data uji yang sama agar perbandingan performa antar algoritma dapat dilakukan secara adil. Model dengan nilai *accuracy* dan *F1-score* terbaik dipilih sebagai model utama dalam sistem rekomendasi hotel berbasis klasifikasi.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Evaluasi Model Klasifikasi

Penelitian ini menguji tiga algoritma klasifikasi, yaitu *Decision Tree*, *Gaussian Naive Bayes*, dan *Random Forest*, untuk memprediksi kategori pilihan hotel wisatawan berdasarkan profil demografis, pola perjalanan, faktor ekonomi, dan preferensi fasilitas. Pengujian dilakukan menggunakan data uji yang telah dipisahkan dari data latih, sehingga performa model dapat diukur berdasarkan kemampuannya dalam memprediksi data yang belum digunakan pada proses pelatihan. Berdasarkan hasil pengujian, diperoleh ringkasan performa model sebagaimana ditunjukkan pada Tabel 1.

Tabel 1. Perbandingan Kinerja Model Klasifikasi

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.6897	0.6687	0.6897	0.6787
Gaussian NB	0.5345	0.5822	0.5345	0.5518
Decision Tree	0.4310	0.4970	0.4310	0.4443

Hasil pada Tabel 1 menunjukkan bahwa algoritma *Random Forest* memperoleh performa terbaik dibandingkan *Gaussian Naive Bayes* dan *Decision Tree*. *Random Forest* menghasilkan *accuracy* sebesar 68,97%, *precision* sebesar 66,87%, *recall* sebesar 68,97%, dan *F1-score* sebesar 67,87%. Nilai tersebut menunjukkan bahwa *Random Forest* memiliki kemampuan lebih baik dalam mempelajari pola hubungan antara variabel prediktor dan kategori pilihan hotel wisatawan.

Performa *Decision Tree* yang relatif rendah menunjukkan bahwa model pohon tunggal kurang mampu menangkap kompleksitas hubungan antar variabel dalam dataset. Hal ini dapat terjadi karena *Decision Tree* cenderung sensitif terhadap variasi data dan menghasilkan aturan klasifikasi yang terlalu sederhana. *Gaussian Naive Bayes* memperoleh performa lebih baik dibandingkan *Decision Tree*, tetapi masih berada di bawah *Random Forest*. Kondisi ini dapat disebabkan oleh asumsi independensi antarfitur pada *Naive Bayes*, sedangkan dalam konteks pemilihan hotel, variabel seperti pendapatan, lama liburan, biaya liburan, dan jarak kemungkinan saling berkaitan.

Random Forest memberikan performa terbaik karena menggunakan pendekatan *ensemble learning* yang menggabungkan beberapa pohon keputusan. Dengan mekanisme tersebut, *Random Forest* dapat menghasilkan prediksi yang lebih stabil dibandingkan *Decision Tree* tunggal. Meskipun demikian, nilai *accuracy* sebesar 68,97% menunjukkan bahwa performa model masih berada pada tingkat moderat. Oleh karena itu, model ini lebih tepat diposisikan sebagai dasar awal dalam pengembangan sistem rekomendasi hotel, bukan sebagai model keputusan akhir yang sepenuhnya menggantikan pertimbangan pengguna.

Untuk melihat distribusi hasil prediksi secara lebih rinci, *confusion matrix* model *Random Forest* ditampilkan pada Tabel 2. Berdasarkan Tabel 2, model *Random Forest* berhasil mengklasifikasikan 26 data pada kelas 1 dengan benar dari total 33 data aktual. Pada kelas 2, model mampu mengklasifikasikan 14 data dengan benar dari total 21 data aktual. Namun, pada kelas 3, model belum mampu menghasilkan prediksi yang benar karena seluruh data aktual kelas 3 diprediksi sebagai kelas 2.

Tabel 2. Confusion Matrix Model Random Forest

Aktual / Prediksi	Prediksi Kelas 1	Prediksi Kelas 2	Prediksi Kelas 3	Total Aktual
Aktual Kelas 1	26	5	2	33
Aktual Kelas 2	7	14	0	21
Aktual Kelas 3	0	4	0	4
Total Prediksi	33	23	2	58

Secara keseluruhan, terdapat 40 data yang berhasil diklasifikasikan dengan benar dari total 58 data uji, sehingga menghasilkan *accuracy* sebesar 68,97%. Hasil ini menunjukkan bahwa *Random Forest* memiliki kemampuan prediksi yang cukup baik pada kelas 1 dan kelas 2, tetapi masih lemah dalam mengenali kelas 3. Kelemahan tersebut kemungkinan disebabkan oleh jumlah data kelas 3 yang relatif sedikit, yaitu hanya 4 data pada data uji. Kondisi ini menunjukkan adanya indikasi ketidakseimbangan distribusi kelas yang dapat memengaruhi kemampuan model dalam mengenali kelas minoritas.

Hasil *classification report Random Forest* juga menunjukkan perbedaan performa antar kelas. Kelas 1 memperoleh *precision*, *recall*, dan *F1-score* masing-masing sebesar 0,79. Kelas 2 memperoleh *precision* sebesar 0,61, *recall* sebesar 0,67, dan *F1-score* sebesar 0,64. Sementara itu, kelas 3 memperoleh *precision*, *recall*, dan *F1-score* sebesar 0,00 karena tidak ada data kelas 3 yang berhasil diprediksi dengan benar.

Temuan ini menunjukkan adanya indikasi ketidakseimbangan distribusi kelas, terutama pada Kelas 3 yang jumlah datanya lebih sedikit dibandingkan kelas lainnya. Ketidakseimbangan kelas dapat menyebabkan model lebih banyak mempelajari pola pada kelas mayoritas, sehingga kemampuan model dalam mengenali kelas minoritas menjadi rendah. Dalam penelitian ini, penanganan *class imbalance* belum dilakukan secara khusus melalui teknik resampling atau *class weighting* sehingga kondisi tersebut menjadi salah satu keterbatasan penelitian. Pada penelitian selanjutnya, penanganan *class imbalance* dapat dilakukan melalui penambahan data pada kelas minoritas, penerapan teknik *oversampling*, *undersampling*, SMOTE, atau penggunaan parameter *class weight* pada algoritma klasifikasi.

3.2. Analisis Faktor yang Memengaruhi Pilihan Hotel

Analisis *feature importance* dilakukan menggunakan model *Random Forest* karena model ini memperoleh performa terbaik dibandingkan algoritma lainnya. *Feature importance* digunakan untuk mengidentifikasi kontribusi relatif setiap variabel prediktor terhadap proses klasifikasi. Semakin tinggi nilai *feature importance*, semakin besar kontribusi variabel tersebut dalam membedakan kategori pilihan hotel wisatawan. Berdasarkan hasil analisis, variabel yang memiliki kontribusi terbesar adalah lama liburan, pekerjaan, besaran biaya liburan, jarak, dan pendapatan.

Lama liburan memperoleh nilai *feature importance* tertinggi sebesar 0,1714. Hal ini menunjukkan bahwa durasi perjalanan menjadi faktor dominan dalam pemilihan hotel. Wisatawan dengan durasi liburan yang lebih panjang cenderung memiliki kebutuhan akomodasi yang berbeda dibandingkan wisatawan dengan perjalanan singkat, baik dari sisi kenyamanan, biaya, maupun fasilitas pendukung.

Variabel pekerjaan memperoleh nilai *feature importance* sebesar 0,1265. Temuan ini menunjukkan bahwa latar belakang pekerjaan dapat berkaitan dengan pola perjalanan, kemampuan finansial, serta preferensi terhadap fasilitas hotel. Besaran biaya liburan dengan nilai 0,1145 dan pendapatan dengan nilai 0,1083 menunjukkan bahwa faktor ekonomi berperan penting dalam pemilihan hotel. Wisatawan dengan anggaran dan pendapatan yang berbeda cenderung memiliki preferensi yang berbeda terhadap kategori hotel. Selain itu, variabel jarak dengan nilai 0,1138 menunjukkan bahwa aksesibilitas lokasi hotel tetap menjadi pertimbangan penting dalam pemilihan hotel.

3.3. Implikasi Terhadap Sistem Rekomendasi

Dalam konteks sistem rekomendasi, hasil klasifikasi dari model terbaik dapat digunakan sebagai dasar untuk menyaring dan menampilkan hotel yang sesuai dengan kategori pilihan wisatawan. Ketika pengguna memasukkan data seperti usia, pendapatan, pekerjaan, status pernikahan, rekan perjalanan, lama liburan, besaran biaya liburan, jarak, fasilitas penginapan, dan layanan internet, sistem akan memprediksi kategori pilihan hotel yang paling sesuai. Hasil prediksi tersebut kemudian dapat digunakan untuk memfilter daftar hotel yang berada pada kategori relevan, sehingga rekomendasi yang diberikan menjadi lebih personal dibandingkan pencarian manual berdasarkan popularitas, harga, atau lokasi saja.

Pendekatan ini berbeda dengan *collaborative filtering* yang membutuhkan data rating atau riwayat interaksi pengguna. *Classification-based recommendation* lebih sesuai digunakan pada tahap awal pengembangan sistem ketika data historis pengguna belum tersedia dalam jumlah memadai, tetapi sistem telah memiliki data profil dan preferensi wisatawan. Dengan demikian, model yang dikembangkan dalam penelitian ini dapat menjadi dasar awal bagi sistem rekomendasi hotel berbasis profil pengguna.

Namun, penelitian ini masih berfokus pada evaluasi model klasifikasi, sehingga evaluasi khusus untuk sistem rekomendasi seperti *precision@k*, *recall@k*, atau *top-k recommendation* belum dilakukan. Oleh karena itu, hasil rekomendasi dalam penelitian ini sebaiknya diposisikan sebagai rekomendasi pendukung. Evaluasi *top-k* perlu ditambahkan pada penelitian selanjutnya agar kualitas daftar rekomendasi hotel dapat dinilai secara lebih spesifik.

3.4. Perbandingan dengan Penelitian Terdahulu

Hasil penelitian ini sejalan dengan penelitian terdahulu yang menunjukkan bahwa *machine learning* dapat digunakan untuk mendukung analisis dan pengambilan keputusan dalam bidang pariwisata dan perhotelan. Namun, penelitian ini memiliki perbedaan pada pendekatan yang digunakan, yaitu

classification-based recommendation berbasis profil dan preferensi wisatawan. Beberapa penelitian terdahulu banyak berfokus pada prediksi permintaan hotel, analisis ulasan pengguna, sistem rekomendasi berbasis rating, dan *collaborative filtering*. Pendekatan tersebut umumnya membutuhkan data historis berupa rating, ulasan, atau riwayat interaksi pengguna. Berbeda dari pendekatan tersebut, penelitian ini menggunakan data profil demografis, pola perjalanan, faktor ekonomi, dan preferensi fasilitas untuk memprediksi kategori pilihan hotel wisatawan. Dengan demikian, pendekatan yang diusulkan lebih relevan untuk kondisi awal pengembangan sistem rekomendasi ketika data historis pengguna belum tersedia dalam jumlah besar.

Dibandingkan dengan pendekatan konvensional yang hanya mengandalkan popularitas hotel, harga, atau pencarian manual, pendekatan *machine learning* dalam penelitian ini mampu mempertimbangkan hubungan antar variabel secara lebih komprehensif. Selain itu, hasil *feature importance* memberikan nilai tambah karena dapat menjelaskan faktor-faktor yang paling berpengaruh dalam pemilihan hotel. Dengan demikian, penelitian ini tidak hanya menghasilkan model prediksi, tetapi juga memberikan interpretasi terhadap variabel yang berperan dalam proses rekomendasi hotel. Meskipun nilai *accuracy Random Forest* sebesar 68,97% belum dapat dikategorikan sangat tinggi, hasil tersebut menunjukkan bahwa penggunaan profil dan preferensi wisatawan memiliki potensi untuk dikembangkan sebagai dasar sistem rekomendasi hotel. Performa yang moderat dapat dipengaruhi oleh jumlah data yang masih terbatas, variasi preferensi wisatawan, serta ketidakseimbangan distribusi kelas target. Oleh karena itu, penelitian ini memberikan dasar awal bagi pengembangan model rekomendasi hotel yang lebih adaptif melalui perluasan dataset, optimasi parameter, penyeimbangan kelas, dan pengujian metode ensemble atau hybrid recommendation pada penelitian selanjutnya.

4. KESIMPULAN

Penelitian ini mengembangkan model *classification-based recommendation* untuk memprediksi kategori pilihan hotel wisatawan berdasarkan profil demografis, pola perjalanan, faktor ekonomi, dan preferensi fasilitas. Hasil pengujian menunjukkan bahwa *Random Forest* memperoleh performa terbaik dibandingkan *Decision Tree* dan *Gaussian Naive Bayes*, dengan *accuracy* sebesar 68,97%, *precision* sebesar 66,87%, *recall* sebesar 68,97%, dan *F1-score* sebesar 67,87%. Hasil ini menunjukkan bahwa *Random Forest* sebagai metode *ensemble* memiliki kemampuan prediksi paling baik, meskipun performanya masih berada pada tingkat moderat. Analisis *feature importance* menunjukkan bahwa lama liburan menjadi faktor paling berpengaruh dalam pemilihan hotel, diikuti oleh pekerjaan, besaran biaya liburan, jarak, dan pendapatan. Temuan ini memperlihatkan bahwa faktor perjalanan dan ekonomi lebih dominan dibandingkan karakteristik demografis dasar dalam menentukan pilihan hotel wisatawan. Kontribusi penelitian ini terletak pada penerapan *classification-based recommendation* sebagai alternatif awal sistem rekomendasi hotel ketika data rating atau riwayat interaksi pengguna belum tersedia. Keterbatasan penelitian terletak pada jumlah data yang terbatas dan distribusi kelas yang belum seimbang, terutama karena model belum mampu mengenali kelas minoritas dengan baik. Penelitian selanjutnya disarankan untuk menambah jumlah responden, menerapkan teknik penyeimbangan data, melakukan *hyperparameter tuning*, serta menguji evaluasi rekomendasi seperti *precision*, *recall* atau *top-k recommendation*.

5. REFERENSI

- Alrassi, K., Al Moaiad, Y., & Alrasi, K. (2025). Review Paper on Recommendation Systems: Different Methods and Techniques. *International Journal of Contemporary Computer Research (IJCCR)*, 1(1).
- Ayu Firnanda, P., Shofwatillah, L., Rahma, F., Fauzi, F., Studi Statistika, P., Muhammadiyah Semarang, U., Kedungmundu No, J., Tembalang, K., & Tengah, J. (2025). Analisis Perbandingan Decision Tree dan Random Forest dalam Klasifikasi Penjualan Produk pada Supermarket. *Emerging Statistics and Data Science Journal*, 3(1).
- Dewiyanti, D., Tantarto, T., & Penulis, K. (2021). Preferensi Wisatawan dalam Memilih Hotel pada Lokasi yang Tidak Strategis Tourist Preferences in Choosing Hotels Located in Non-Strategic Locations. *Jurnal Lingkungan Binaan Indonesia*, 10(3). <https://doi.org/10.32315/jlbi.v10i3.68>
- Dowlut, N., & Gobin-Rahimbux, B. (2023). Forecasting resort hotel tourism demand using deep learning techniques – A systematic literature review. *Heliyon*, 9(7). <https://doi.org/10.1016/j.heliyon.2023.e18385>

- Ernawati, H., Hananto, K., & Tinggi Pariwisata Ambarrukmo Yogyakarta, S. (2023). PARIWISATA DIGITAL: PERSPEKTIF DAN AGENDA RISET MASA DEPAN. *Kepariwisata : Jurnal Ilmiah*, 17(2).
- Gadkari, C. T. (2023). Recommendation System And It's Types. *International Research Journal of Modernization in Engineering Technology and Science*. <https://doi.org/10.56726/irjmets33481>
- Hadi, A., Qamal, M., & Afrillia, Y. (2025). Comparison of Random Forest Algorithm Classifier and Naive Bayes Algorithm in Whatsapp Message Type Classification. *Journal of Renewable Energy, Electrical, and Computer Engineering*, 5(1), 9–17. <https://doi.org/10.29103/jreece.v5i1.21227>
- Li, N., Liu, X., & Meng, F. (2025). Machine Learning Techniques in Tourism and Hospitality Research: A Critical Assessment. *Journal of Travel Research*. <https://doi.org/10.1177/00472875251401912>
- Li, Y. (2025). A Review of Collaborative Filtering Recommendation Systems. In *Computing & Informatics* (Vol. 1, Number 1). <https://pinnaclepubs.com/index.php/EJACI>
- Novita Sari, I., Fakhri Lidimilah, L., & Lutfi, A. (2025). Proceeding National Conference Of Research And Community Service Sisi Indonesia 2025 Analisis Perbandingan Algoritma Decision Tree, Random Forest, Dan Naive Bayes Dalam Klasifikasi Gangguan Tidur. Retrieved <https://conference.sinesia.id/ncrcs-sinesia>
- Pamungkas, F. S., Prasetya, B. D., & Kharisudin, I. (2019). Perbandingan Metode Klasifikasi Supervised Learning pada Data Bank Customers Menggunakan Python. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 689–694. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Perdana, D. A., Manajemen, J., Tinggi, S., & Galileo -Batam, I. E. (2025). YUME : Journal of Management Analisis Faktor-Faktor Yang Mempengaruhi Konsumen Dalam Memilih Tempat Menginap Di Grup Hotel Ibis. In *YUME : Journal of Management* (Vol. 8, Number 1).
- Solano-Barliza, A., Arregocés-Julio, I., Aarón-Gonzalvez, M., Zamora-Musa, R., De-La-Hoz-Franco, E., Escorcia-Gutierrez, J., & Acosta-Coll, M. (2024). Recommender systems applied to the tourism industry: a literature review. In *Cogent Business and Management* (Vol. 11, Number 1). Cogent OA. <https://doi.org/10.1080/23311975.2024.2367088>
- Suryani, I. (2023). Accredited rank 4 (SINTA 4), excerpts from the decision of the Comparison Of Decision Tree, Naive Bayes And Random Forest Algorithm To Get The Best Performance Of Algorithm For Customer Credit Classification Accredited rank 4 (SINTA 4), excerpts from the decision of the. *JURNAL RISET INFORMATIKA*, 6(1). <https://doi.org/10.34288/jri.v6i1>
- Tejashri, M., Phalle, S., & Bhushan, S. (2025). Content Based Filtering And Collaborative Filtering: A Comparative Study. *Journal of Advanced Zoology*.
- Topsakal, Y., Icoz, O., & Icoz, O. (2021). Digital Transformation and Tourist Experiences (pp. 19–41). <https://doi.org/10.4018/978-1-7998-8528-3.ch002>