JIIFKOM (Jurnal Ilmiah Informatika & Komputer) STTR Cepu
E-Mail: jiifkom@sttrcepu.ac.id

Klasifikasi Popularitas Menu Kafe Menggunakan Random Forest dengan Labeling Berbasis Quantile pada Data Transaksi Penjualan

Aryanti Aryanti[a*], Dea Putri Amanda[b], Naila Nadra[c], Riska Riska[d], Muhammad Zaid Akbar[e], Fikran Tamma Adistra[f]
e-mail: aryanti@polsri.ac.id

Politeknik Negeri Sriwijaya[abcdef]

Intisari
Perkembangan industri kafe menuntut pemahaman yang lebih akurat terhadap preferensi pelanggan, khususnya dalam menentukan popularitas menu. Penelitian ini bertujuan mengembangkan model klasifikasi tingkat popularitas menu kafe menggunakan algoritma Random Forest berbasis data transaksi penjualan. Dataset yang digunakan mencakup atribut harga menu (Price Per Unit) dan waktu transaksi (Transaction Date), yang diproses melalui tahap preprocessing serta ekstraksi fitur temporal meliputi Month, DayOfWeek, dan IsWeekend. Label popularitas menu ditentukan berdasarkan frekuensi kemunculan setiap item dalam data transaksi menggunakan nilai ambang persentil ke-60. Hasil pengujian menunjukkan bahwa model yang dibangun memiliki performa yang baik dengan nilai akurasi 94,79%, precision 94,10%, recall 93,74%, dan F1-score 93,92%. Model ini mampu mengidentifikasi pola popularitas menu secara efektif dan dapat dimanfaatkan sebagai referensi dalam memahami preferensi pelanggan berbasis data.

Kata kunci:
Random Forest, Klasifikasi, Popularitas Menu, Data Transaksi, Machine Learning

Tentang naskah:
-diterima, 10 Juni 2026
-diterbitkan, 1 Juli 2026

Abstract
The rapid growth of the cafe industry requires a more accurate understanding of customer preferences, particularly in determining menu popularity. This study aims to develop a classification model for menu popularity levels using the Random Forest algorithm based on cafe sales transaction data. The dataset includes menu price attributes (Price Per Unit) and transaction time (Transaction Date), which are processed through a preprocessing stage and temporal feature extraction including Month, DayOfWeek, and IsWeekend. Popularity labels are determined based on the frequency of each item's occurrence in transaction data using the 60th percentile as the threshold value. The results show that the model achieves strong performance with 94.79% accuracy, 94.10% precision, 93.74% recall, and 93.92% F1-score. The model effectively identifies menu popularity patterns and can serve as a reference for data-driven customer preference analysis.

Keywords:
Random Forest, Classification, Menu Popularity, Transaction Data, Machine Learning

1. Pendahuluan

Perkembangan industri kuliner, khususnya kafe, menunjukkan perubahan fungsi yang tidak lagi hanya sebagai tempat konsumsi makanan dan minuman, tetapi juga dimanfaatkan sebagai ruang aktivitas lain seperti belajar dan bekerja. Kafe kini berperan sebagai *informal learning space* yang mendukung berbagai aktivitas produktif, terutama di kalangan mahasiswa (Adityawirawan & Kusuma, 2021) seiring dengan perubahan tersebut, pemahaman terhadap pola konsumsi pelanggan menjadi penting untuk mendukung pengelolaan usaha kafe. Salah satu sumber informasi yang dapat dimanfaatkan adalah data transaksi penjualan yang merekam aktivitas pembelian pelanggan secara historis dan dapat digunakan untuk mengidentifikasi pola preferensi terhadap menu tertentu.

Pada tahun 2001, Breiman memperkenalkan algoritma *Random Forest* dan menunjukkan bahwa metode ini memiliki tingkat kesalahan yang lebih rendah serta mampu menghasilkan klasifikasi yang akurat, terutama dalam mengolah

data berukuran besar (Breiman, 2001). Pemanfaatan data transaksi melalui pendekatan *machine learning*, khususnya metode klasifikasi, memungkinkan pengelompokan menu berdasarkan tingkat popularitasnya secara lebih objektif dibandingkan pendekatan konvensional (Rindiyani et al., 2022). Selain itu, penerapan algoritma *Random Forest* dalam analisis data penjualan terbukti efektif dalam menangani data yang kompleks serta meningkatkan akurasi prediksi dalam mengidentifikasi pola permintaan pelanggan (Amelia & R, 2025).

Seiring dengan perkembangan teknologi, pemanfaatan *machine learning* dalam sektor kuliner dan *hospitality* menunjukkan peningkatan yang signifikan, khususnya dalam mendukung analisis berbasis data. Penerapan model *machine learning* pada data operasional, seperti data permintaan dan transaksi, terbukti mampu meningkatkan akurasi prediksi dibandingkan pendekatan tradisional sehingga dapat membantu dalam memahami pola konsumsi dan mengoptimalkan pengelolaan sumber daya (Rodrigues et al., 2024). Kajian literatur terbaru dalam bidang *hospitality* menunjukkan adanya tren peningkatan penggunaan kecerdasan buatan untuk berbagai kebutuhan analitis, seperti analisis perilaku pelanggan, prediksi permintaan, serta pengambilan keputusan berbasis data (To & Yu, 2025).

Salah satu pendekatan yang banyak digunakan dalam analisis data adalah algoritma *Random Forest* yang dikenal memiliki kemampuan dalam menangani data kompleks serta memberikan performa yang baik pada tugas klasifikasi dan prediksi. Algoritma *Random Forest* dikenal mampu menangani data dengan tingkat kompleksitas yang tinggi serta mengekstraksi pola dari berbagai atribut yang tersedia (Arifuddin et al., 2025). Kemampuan tersebut menjadikan algoritma ini relevan untuk digunakan dalam proses klasifikasi, termasuk dalam mengidentifikasi tingkat popularitas menu berdasarkan data transaksi dan preferensi pelanggan.

Selain itu, *Random Forest* juga telah dimanfaatkan dalam sistem rekomendasi makanan untuk menganalisis preferensi pengguna berdasarkan berbagai atribut sehingga dapat menghasilkan rekomendasi yang lebih sesuai (Akbar & Ramadhan, 2026). Berbagai faktor seperti harga, jenis produk, serta data transaksi dan volume penjualan dapat digunakan sebagai dasar dalam analisis untuk memahami pola perilaku konsumen. Pendekatan *machine learning* dalam sistem berbasis data juga telah dimanfaatkan untuk mengolah preferensi pengguna secara lebih sistematis sehingga berkontribusi dalam peningkatan kualitas layanan dan pengambilan keputusan (Solano-Barliza et al., 2024).

Berdasarkan kajian pada industri kuliner, algoritma *Random Forest* mampu menganalisis tren *penjualan* menu dengan tingkat akurasi mencapai 96,25%, yang menunjukkan performa tinggi dalam menangani data penjualan yang kompleks (Ramadhan et al., 2025). Meskipun berbagai penelitian telah membahas penerapan *machine learning* dalam industri makanan, sebagian besar masih berfokus pada prediksi permintaan, sistem rekomendasi, atau analisis sentimen. Penelitian yang secara khusus mengkaji klasifikasi tingkat popularitas menu pada kafe masih relatif terbatas, terutama yang menggunakan algoritma *Random Forest* sebagai pendekatan utama. Hal ini menunjukkan adanya peluang untuk mengembangkan model klasifikasi yang lebih spesifik dalam menganalisis popularitas menu.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengklasifikasikan tingkat popularitas menu kafe menggunakan algoritma *Random Forest* berdasarkan atribut harga menu dan waktu transaksi. Penelitian ini menguji *kemampuan Random Forest* dalam membedakan menu dengan tingkat popularitas tinggi dan rendah melalui pendekatan *quantile-based labeling* pada data transaksi penjualan. Hasil penelitian diharapkan dapat memberikan informasi yang lebih objektif mengenai pola preferensi pelanggan serta mendukung pengambilan keputusan berbasis data dalam pengelolaan menu kafe.

2. Kerangka Teori

2.1. Random Forest

Menurut (Breiman, 2001), Random Forest adalah algoritma pembelajaran ensemble yang *membangun* beberapa pohon keputusan secara otonom dengan data pelatihan dan karakteristik yang dipilih secara acak. Kemudian, untuk klasifikasi, algoritma ini menggabungkan prediksi dari semua pohon menggunakan mekanisme pemungutan suara mayoritas. Dengan membuat model lebih tahan terhadap data dengan pola yang rumit dan non-linear, strategi *bagging* mengurangi varians tanpa secara signifikan meningkatkan bias. Menurut (Ibrahim, 2022), *Random Forest* adalah iterasi dari pohon keputusan berbasis *bagging* yang bertujuan untuk meningkatkan generalisasi dan stabilitas dalam klasifikasi dengan mengoptimalkan *trade-off* antara varians dan bias.

2.2. Confusion Matrix

Kinerja model klasifikasi dapat diukur menggunakan *Confusion Matrix*, yang membandingkan prediksi model dengan kondisi

aktual. Untuk lebih memahami kinerja model, pendekatan ini memberikan matriks dengan data tentang jumlah prediksi yang benar dan salah. *Confusion Matrix* berguna untuk mengukur keberhasilan klasifikasi dan untuk menentukan berbagai jenis kesalahan prediksi (Sathyanarayanan & Tantri, 2024).

Keempat elemen ini membentuk *True Negative* (TN), *False Positive* (FP), *True Positive* (TP), dan *False Negative* (FN) yang membentuk *Confusion Matrix* dalam klasifikasi biner. Sejumlah metrik penilaian, termasuk *F1-score, recall, specificity, accuracy,* dan *precision,* didasarkan pada data ini. Peneliti dapat mempelajari lebih lanjut tentang kemampuan klasifikasi suatu model dan model mana yang paling sesuai untuk tugas tertentu dengan menggunakan ukuran-ukuran ini untuk mengevaluasi kinerja model (Sathyanarayanan & Tantri, 2024).

2.3. Quantile-Based Labeling

Klasifikasi data berdasarkan distribusi frekuensi transaksi memungkinkan algoritma berbasis kuantil menentukan label popularitas tanpa menggunakan nilai absolut. Strategi ini memiliki dasar statistik yang kuat karena dapat menghilangkan bias dominasi kelas, yang khususnya bermasalah pada data non-normal. (Sch et al., 2022) menunjukkan bahwa algoritma pengelompokan kuantil mengelompokkan data dengan lebih tepat, menghasilkan kualitas klasifikasi yang lebih baik dan kemampuan model untuk mendeteksi pola yang lebih konsisten.

3. Metodologi

3.1. Pengumpulan Dataset

Penelitian ini menggunakan data transaksi penjualan kafe yang bersumber dari platform *Kaggle*, khususnya *Cafe Sales Dataset* dalam format .csv. Kaggle adalah platform komunitas untuk praktisi dan akademisi ilmu data, yang menawarkan berbagai *dataset* terbuka di berbagai bidang (Hermawan, 2024). Dataset ini memiliki properti termasuk nama item (menu), harga satuan (Harga Per Unit), dan waktu transaksi. Penggunaan *dataset* publik meningkatkan kemampuan penggunaan kembali dan transparansi studi, memfasilitasi replikasi dan perbandingan dengan studi lain, yang pada gilirannya meningkatkan kredibilitas temuan penelitian (Mohammed et al., 2025).

Pendekatan pengumpulan data melibatkan pengambilan dataset dari *Kaggle* dan selanjutnya mengidentifikasi atribut yang relevan. Pemilihan atribut mempertimbangkan kualitas data, konsistensi nilai, dan kemungkinan dampak setiap atribut pada kinerja model. (Pudjihartono et al., 2022) menyatakan bahwa metode pemilihan fitur meningkatkan akurasi model dan mengurangi kutukan dimensi dengan mempertahankan fitur yang paling informatif. Setelah seleksi, data tersebut menjalani fase validasi pendahuluan untuk memastikan tidak adanya anomali substansial, termasuk nilai ekstrem yang tidak masuk akal atau format temporal yang tidak konsisten.

3.2. Preprocessing Dataset

Fase persiapan dilakukan untuk menjamin kualitas, konsistensi, dan struktur data yang sesuai dengan persyaratan analitis. Prosedur ini mencakup pembersihan data, modifikasi fitur, dan pembuatan label kategorisasi. (Strasser, 2024) menyatakan bahwa keterbukaan dan keterlacakan dalam pipeline pra-pemrosesan sangat penting untuk melacak dan memvalidasi modifikasi data, sehingga meningkatkan keandalan dan reproduksibilitas sistem pembelajaran mesin.

Dari 9.741 set data awal, 9.121 set data bersih diperoleh setelah proses pembersihan. Pengurangan ini menandakan keberhasilan penghapusan data duplikat, nilai yang tidak akurat, dan outlier. (Gong et al., 2023) menyatakan bahwa kualitas dataset yang diperoleh melalui pembersihan dan pra-pemrosesan data sangat penting untuk meningkatkan akurasi, efisiensi, dan generalisasi model pembelajaran mesin.

Fitur model melibatkan pengambilan informasi dari *Price Per Unit*, *Month*, *DayOfWeek*, dan *IsWeekend*. Atribut *Month*, *DayOfWeek*, dan *IsWeekend* diperoleh melalui proses ekstraksi fitur dari atribut *Transaction Date* yang terdapat pada dataset. Pada tahap *preprocessing*, *Transaction Date* yang masih berupa *timestamp* dikonversi ke format *datetime* menggunakan *library Pandas*. Selanjutnya, fitur *Month* dibentuk untuk merepresentasikan bulan terjadinya transaksi, *DayOfWeek* digunakan untuk menunjukkan hari transaksi dalam satu minggu (0–6), sedangkan *IsWeekend* merupakan variabel biner yang bernilai 1 apabila transaksi terjadi pada akhir pekan (Sabtu atau Minggu) dan bernilai 0 apabila transaksi terjadi pada hari kerja. Ketiga fitur temporal tersebut digunakan untuk menangkap pola pembelian pelanggan berdasarkan dimensi waktu transaksi. Label popularitas ditetapkan berdasarkan nilai ambang batas 0,60 yang mewakili persentil ke-60 dari distribusi jumlah transaksi untuk setiap item menu.

Menu dengan jumlah transaksi sama dengan atau di atas nilai ambang batas diklasifikasikan sebagai Popularitas Tinggi, dan yang di bawah ambang batas diklasifikasikan sebagai Popularitas Rendah. (Berrettini et al., 2024) menunjukkan bahwa strategi klasifikasi berbasis kuantil meningkatkan akurasi dan stabilitas

model karena kemampuannya beradaptasi dengan distribusi data yang kompleks.

Tabel 1. Jumlah Data Sebelum dan Setelah Preprocessing

Tahap Data	Jumlah Data
Data awal	9741
Data setelah di cleaning	9121

```
menu_count = train_data.groupby('Item').size().reset_index(name='Total')

threshold = menu_count['Total'].quantile(0.6)

menu_count['Popularity'] = menu_count['Total'].apply(
    lambda x: 'Tinggi' if x >= threshold else 'Rendah'
)
```

Gambar 3.1 Implementasi Quantile-Based Labeling untuk Penentuan Tingkat Popularitas Menu

Tabel 2. Variabel Penelitian

No.	Variabel	Keterangan
1	X1 – Harga Menu (*Variabel Input*)	Menunjukkan harga satuan setiap menu yang digunakan sebagai *variabel input* dalam model klasifikasi.
2	X2 – Waktu Transaksi *(Variabel Input*)	Waktu saat transaksi terjadi. Meliputi *Month*, *DayOfWeek*, dan *IsWeekend*. Menggambarkan pola pembelian pelanggan berdasarkan dimensi waktu.
3	Y – Label Popularitas (*Variabel Output*/Target)	Label biner hasil klasifikasi frekuensi transaksi menu menggunakan ambang *quantile* 0,60 (persentil ke-60), terdiri atas kategori Tinggi dan Rendah.

3.3. Training Model

Fase pelatihan model menggunakan metode pengklasifikasi Random Forest sebagai model utama, karena kemampuannya dalam mengelola data tabular yang terdiri dari variabel numerik dan kategorikal (Aryanti et al., 2026). Data dibagi dengan rasio 80:20, mengalokasikan 80% untuk set pelatihan dan 20% untuk set pengujian. Posch et al. (2022) membuktikan bahwa Random Forest secara efektif menangkap pola permintaan yang rumit dan non-linear dalam prediksi penjualan makanan dan minuman dengan akurasi yang cukup tinggi. Model ini dibangun dengan menggunakan atribut *Price Per Unit*, *Month*, *DayOfWeek*, dan *IsWeekend*, sebagai *variabel input*, dengan Popularitas sebagai variabel tujuan.

4. Hasil dan Pembahasan

4.1 Hasil Klasifikasi Popularitas Menu

Penelitian ini menggunakan algoritma Random Forest untuk mengkategorikan popularitas menu kafe melalui data transaksi penjualan. Tahapan penelitian meliputi pembersihan data, ekstraksi fitur temporal dari atribut Tanggal Transaksi, penetapan label popularitas berdasarkan frekuensi transaksi menu, dan pelatihan serta evaluasi model klasifikasi. Dari 9.741 set data awal, 9.121 dianggap layak untuk digunakan setelah pra-pemrosesan. Label popularitas dikategorikan menjadi Tinggi dan Rendah berdasarkan ambang batas persentil ke-60 transaksi untuk setiap item menu.

Hasil klasifikasi menunjukkan bahwa *Juice*, *Coffee*, dan *Cake* diklasifikasikan sebagai populer Tinggi, karena volume transaksinya yang lebih tinggi dibandingkan item menu lainnya. Sebaliknya, *Sandwich*, *Cookie*, *Smoothie*, *Tea*, dan *Salad* diklasifikasikan sebagai populer Rendah. Temuan ini mengungkapkan bahwa pola pembelian pelanggan tersebar tidak merata di seluruh menu, dengan konsentrasi pada produk-produk tertentu yang menunjukkan peningkatan permintaan. Informasi ini dapat memandu keputusan tentang manajemen inventaris, perumusan rencana promosi, dan pengembangan menu yang selaras dengan preferensi pelanggan.

Tabel 3. Klasifikasi Tingkat Popularitas Menu

No	Item	Total Transaksi	Popularitas
1	Juice	1.187	Tinggi
2	Coffee	928	Tinggi
3	Cake	905	Tinggi
4	Sand wich	903	Rendah

5	Cookie	887	Rendah
6	Smoothie	886	Rendah
7	Tea	872	Rendah
8	Salad	728	Rendah

4.2 Evaluasi dan Kinerja Model

Kinerja model dinilai menggunakan *Confusion Matrix* bersama dengan parameter *accuracy*, *precision*, *recall*, dan F1 *Score*. Hasil pengujian menunjukkan *accuracy* 94,79%, *precision* 94,10%, *recall* 93,74%, dan F1 *Score* 93,92%. Nilai-nilai tersebut menunjukkan bahwa model dapat mengkategorikan tingkat popularitas menu dengan akurasi yang signifikan dan sedikit kesalahan klasifikasi. Angka presisi dan recall yang tinggi menunjukkan bahwa model tersebut akurat dalam prediksinya dan secara teratur mengidentifikasi item menu yang populer.

Temuan penelitian ini sesuai dengan temuan (Ramadhan et al., 2025) dan (Posch et al., 2022), yang menunjukkan bahwa algoritma Random Forest menunjukkan efikasi yang kuat dalam mengevaluasi tren penjualan dan mencapai kinerja klasifikasi yang tinggi pada data transaksi. Penelitian ini menggunakan metode kategorisasi popularitas menu yang mengintegrasikan parameter harga dan karakteristik transaksi temporal yang berasal dari data penjualan. Metode ini memberikan wawasan yang lebih tepat tentang popularitas menu, sehingga memfasilitasi pengambilan keputusan berbasis data dalam manajemen kafe.

Tabel 4. Hasil Evaluasi Performa Model

No	Metrik Evaluasi	Nilai	Persentase
1	*Accuracy*	0,9479	94,79%
2	*Precision*	0,9410	94,10%
3	*Recall*	0,9374	93,74%
4	*F1-Score*	0,9392	93,92%

5. Simpulan

Studi ini menyimpulkan bahwa algoritma Random Forest efektif untuk mengklasifikasikan tingkat popularitas menu kafe berdasarkan data transaksi penjualan, dengan memanfaatkan atribut harga dan variabel waktu transaksi. Hasil kategorisasi secara efektif membedakan antara menu dengan popularitas tinggi dan rendah, sehingga menjelaskan pola preferensi pelanggan terhadap produk yang tersedia. Lebih lanjut, hasil pengujian menunjukkan bahwa model memiliki kemampuan klasifikasi yang kuat dalam membedakan antara kedua kategori popularitas tersebut, mencapai tingkat akurasi yang tinggi dan kesalahan prediksi yang relatif rendah. Temuan ini menunjukkan bahwa pendekatan pembelajaran mesin Random Forest sangat baik dalam memfasilitasi pengambilan keputusan berbasis data dalam manajemen menu kafe, khususnya untuk strategi promosi dan penilaian produk yang diminati pelanggan. Penelitian selanjutnya dapat ditingkatkan dengan menggabungkan variasi data, menambah karakteristik yang digunakan, dan mengevaluasi efektivitas Random Forest terhadap algoritma klasifikasi alternatif untuk mencapai model yang lebih ideal dalam mencerminkan pola perilaku pelanggan.

Daftar Pustaka

Artikel jurnal:

Adityawirawan, S. S. K., & Kusuma, H. E. (2021). *Cafe As Student ' S Informal Learning Space : A Case Study In Bandung , Indonesia*. *48*(2), 109–119. https://doi.org/10.9744/dimensi.48.2.109-120

Akbar, M. S., & Ramadhan, I. (2026). *Pemanfaatan Random Forest Classifier Untuk Sistem Rekomendasi Makanan Berbasis Preferensi Pengguna*. *15*.

Amelia, D., & R, R. K. (2025). *Penerapan Algoritma Random Forest Untuk Prediksi Pejualan Dan Persedian Produk Pada Toko Frozen Food Anisa*. 843–848.

Aryanti, Putri, A. T., Khairunisya, A., Pratama, I. Y., & Azizah, P. N. (2026). *Klasifikasi Tempat Wisata Di Yogyakarta Berdasarkan Harga Dan Rating Menggunakan Algoritma Random Forest*. *18*(2), 192–204.

Berrettini, M., Hennig, C., & Viroli, C. (2024). *The quantile-based classifier with variable-wise parameters The quantile classifier*.

Gong, Y., Liu, G., Xue, Y., Li, R., & Meng, L. (2023). A survey on dataset quality in machine learning. *Information and Software Technology*, *162*(September 2022), 107268. https://doi.org/10.1016/j.infsof.2023.107268

Hermawan, G. (2024). *Memahami Peran Dataset dalam Penelitian Kecerdasan Buatan : Kualitas , Aksesibilitas , dan Tantangan*. 1–9.

Ibrahim, M. (2022). *Evolution of Random Forest from Decision Tree and Bagging : A Bias-Variance Perspective*. *7*(1), 66–71.

Mohammed, S., Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Naumann, F., & Harmouch, H. (2025). The effects of data quality on machine learning performance on tabular data. *Information Systems*, *132*(June 2024), 102549. https://doi.org/10.1016/j.is.2025.102549

Posch, K., Truden, C., Hungerländer, P., & Pilz, J. (2022). A Bayesian approach for predicting food and beverage sales in staff canteens and restaurants ☆. *International Journal of Forecasting*, *38*(1), 321–338. https://doi.org/10.1016/j.ijforecast.2021.06.001

Pudjihartono, N., Fadason, T., & Kempa-liehr, A. W. (2022). *A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction*. *2*(June), 1–17. https://doi.org/10.3389/fbinf.2022.927312

Ramadhan, G., Faqih, A., Permana, F. E., Informatika, T., Informatika, M., & Cirebon, K. (2025). *Analisis Tren Penjualan Menu Seafood Dengan Algoritma Random Forest Untuk Meningkatkan Strategi Pemasaran*. *9*(4),

5942–5949.
Rindiyani, Primadewi, A., Maimunah, & Purwantini, A. H. (2022). *Klasifikasi Penjualan berdasarkan Platform pada UMKM Omah Branded Menggunakan Random Forest*. *9*(5), 1520–1529. https://doi.org/10.30865/jurikom.v9i5.4949
Rodrigues, M., Miguéis, V., Freitas, S., & Machado, T. (2024). Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste. *Journal of Cleaner Production*, *435*(November 2023), 140265. https://doi.org/10.1016/j.jclepro.2023.140265
Sathyanarayanan, S., & Tantri, B. R. (2024). *Confusion Matrix-Based Performance Evaluation Metrics*. *27*(4).
Sch, L., Swift, A. J., Lu, H., & Member, S. (2022). *Uncertainty Estimation for Heatmap-based Landmark Localization*. *XX*(Xx), 1–14.
Solano-Barliza, A., Valls, A., Acosta-Coll, M., Moreno, A., Escorcia-Gutierrez, J., De-La-Hoz-Franco, E., & Arregoces-Julio, I. (2024). Enhancing Fair Tourism Opportunities in Emerging Destinations by Means of Multi-criteria Recommender Systems: The Case of Restaurants in Riohacha, Colombia. *International Journal of Computational Intelligence Systems*, *17*(1). https://doi.org/10.1007/s44196-024-00700-8
Strasser, S. (2024). *Transparent Data Preprocessing for Machine Learning*. https://doi.org/10.1145/3665939.3665960
To, W. M., & Yu, B. T. W. (2025). Artificial Intelligence Research in Tourism and Hospitality Journals: Trends, Emerging Themes, and the Rise of Generative AI. *Tourism and Hospitality*, *6*(2). https://doi.org/10.3390/tourhosp6020063

Buku :

Breiman, L. (2001). *Random Forest*. 1–33.
Arifuddin, N. A., Dahlan, A., Insani, C. N., Nur, N., Arifin, N., Setiawan, H., Sutanto, A., & Kristanto, T. (2025). *Machine Learning*.