

Klasifikasi Konsumen Berdasarkan Loyalitas Belanja Online Menggunakan Algoritma Random Forest

Syifa Aurellia Rahma^{1*}, Tiara Putri²,

^{1,2}Program Studi Teknik Informatika, Universitas Pelita Bangsa, Indonesia

Email: ¹syifaurellia.31210009@mhs.pelitabangsa.ac.id, ²tiaraputri312210064@mhs.pelitabangsa.ac.id

INFORMASI ARTIKEL

Histori artikel:

Naskah masuk, 2 Juli 2025

Direvisi, 30 Juli 2025

Diiterima, 31 Juli 2025

Kata Kunci:

Loyalitas Konsumen
Random Forest
Klasifikasi
Machine Learning

ABSTRAK

Abstract- The rapid growth of online shopping has driven e-commerce companies to understand consumer behavior, particularly in measuring customer loyalty. This study aims to classify consumer loyalty based on sales transaction data using the Random Forest algorithm. The dataset was obtained from the Kaggle platform, containing e-commerce transaction records from the United Kingdom. Data analysis involved several stages, including data cleaning, categorical attribute encoding, numerical data normalization, and feature engineering to generate attributes such as shopping frequency, total quantity, spending, and transaction duration. Loyalty classes were categorized into four groups: Loyal, Not Loyal, New, and Potential Loyal, with an imbalanced distribution dominated by the Loyal class. Random Forest was chosen for its ability to handle large, imbalanced data and provide useful feature importance. The model achieved an accuracy of 91%, although its performance on minority classes remained low due to data imbalance. The *TotalPrice* feature was the most influential factor in determining loyalty. The results show that a data-driven classification approach can effectively support customer retention strategies in the e-commerce industry.

Abstrak- Pertumbuhan pesat belanja online mendorong perusahaan e-commerce untuk memahami perilaku konsumen, terutama dalam mengukur loyalitas pelanggan. Penelitian ini bertujuan mengklasifikasikan loyalitas konsumen berdasarkan data transaksi penjualan menggunakan algoritma *Random Forest*. Dataset diambil dari platform Kaggle yang memuat transaksi e-commerce di Inggris. Analisis data dilakukan melalui tahapan pembersihan, *encoding* atribut kategorikal, normalisasi data numerik, dan *feature engineering* untuk menghasilkan atribut seperti frekuensi belanja, total kuantitas, pembelanjaan, dan durasi transaksi. Kelas loyalitas dibentuk menjadi empat kategori: Loyal, Tidak Loyal, Baru, dan Potensial Loyal, dengan distribusi tidak seimbang, didominasi kelas Loyal. *Random Forest* dipilih karena kemampuannya menangani data besar dan tidak seimbang, serta memberikan *feature importance* yang bermanfaat. Model mencapai akurasi 91%, namun performa pada kelas minoritas masih rendah karena ketimpangan data. Fitur *TotalPrice* menjadi faktor paling berpengaruh dalam menentukan loyalitas. Hasil penelitian menunjukkan bahwa klasifikasi berbasis data dapat digunakan secara efektif untuk mendukung strategi retensi pelanggan pada industri e-commerce.

Copyright © 2025 LPPM - STMIK IKMI Cirebon
This is an open access article under the CC-BY license

Penulis Korespondensi:

Syifa Aurellia Rahma

Program Studi Teknik Informatika,

Universitas Pelita Bangsa

Indonesia

Email: syifaurellia.31210009@mhs.pelitabangsa.ac.id

1. Pendahuluan

Di era digital, belanja online telah menjadi bagian dari gaya hidup masyarakat. Namun, tantangan utama bagi perusahaan e-commerce adalah mempertahankan loyalitas pelanggan, yang penting untuk keberlangsungan bisnis. Banyak perusahaan belum memiliki sistem otomatis berbasis data untuk mengidentifikasi tingkat loyalitas konsumen, sehingga pengambilan keputusan masih bersifat manual dan kurang efektif.

Penelitian ini mengusulkan penggunaan algoritma *Random Forest*, salah satu metode *machine learning* yang andal untuk klasifikasi data berskala besar. Tidak seperti studi sebelumnya yang berfokus pada prediksi pembelian, penelitian ini menargetkan klasifikasi loyalitas pelanggan berdasarkan data transaksi. Hasilnya diharapkan dapat membantu perusahaan menyusun strategi retensi pelanggan yang lebih tepat sasaran dan berbasis data.

Perkembangan teknologi digital telah mengubah pola konsumsi masyarakat, khususnya dalam sektor *e-commerce*. Untuk memahami perilaku belanja konsumen secara lebih akurat, berbagai pendekatan *machine learning* telah banyak diterapkan. Beberapa algoritma seperti *Random Forest*, *Support Vector Machine (SVM)*, dan *Neural Network* digunakan dalam mengklasifikasikan pelanggan berdasarkan pola pembelajarannya [1]. Dalam konteks loyalitas konsumen, algoritma *Random Forest* menjadi salah satu pilihan utama karena mampu memberikan gambaran yang lebih jelas tentang preferensi konsumen berdasarkan data historis dan umpan balik yang spesifik [2]. *Random Forest* merupakan teknik *ensemble learning* yang menggabungkan banyak pohon keputusan guna menghasilkan prediksi yang lebih stabil dan akurat. Kelebihan ini membuat *Random Forest* cocok untuk klasifikasi data yang kompleks dan beragam [3].

Random Forest bekerja dengan membentuk sejumlah pohon keputusan yang dilatih secara individual, lalu menggabungkan hasilnya untuk klasifikasi (melalui voting mayoritas) atau regresi (melalui rata-rata). Pendekatan ini dapat mengurangi risiko *overfitting* yang sering terjadi pada pohon tunggal dan cocok digunakan dalam berbagai masalah prediksi [4]. Salah satu keunggulan *Random Forest* adalah kemampuannya dalam melakukan seleksi fitur, yang sangat bermanfaat saat bekerja dengan dataset besar dan kompleks. Ini membuat algoritma ini tetap mampu memberikan prediksi yang akurat meskipun berhadapan dengan variabel input yang banyak dan tidak seimbang [5]. Proses kerjanya juga melibatkan perhitungan *entropy* dan *information*

gain untuk menilai ketidakhadiran atribut, yang menjadi dasar dalam pemilihan cabang pada pohon keputusan [6]. Dengan cara ini, *Random Forest* tidak hanya menghasilkan klasifikasi yang kuat, tetapi juga membantu peneliti memahami variabel-variabel yang paling berpengaruh.

Secara konseptual, *Random Forest* termasuk dalam kategori *supervised learning* yang diperkenalkan oleh Leo Breiman. Metode ini dikenal karena akurasi yang tinggi dalam memprediksi dan kemampuannya menangani data dalam skala besar tanpa menyebabkan *overfitting*. Selain itu, ia mampu mengurangi korelasi antar pohon keputusan berkat teknik *ensemble* yang digunakan [7]. *Random Forest* dibangun atas dasar algoritma *Decision Tree* dan memperluasnya dengan pendekatan *ensemble*, yaitu dengan membentuk banyak pohon dan mengambil keputusan berdasarkan *majority voting*. Proses pembentukan *Decision Tree* dalam *Random Forest* mirip dengan metode CART, hanya saja tidak melibatkan proses pemangkasan (*pruning*), yang justru meningkatkan ketahanan terhadap noise [8]. Kombinasi teknik *Bagging* dan *Random Subspaces* menjadikan *Random Forest* mampu melakukan klasifikasi dan regresi secara lebih efisien [9].

Meski telah terbukti efektif, algoritma ini tetap memiliki tantangan, seperti dalam penentuan parameter optimal, pengelolaan data yang tidak seimbang, dan pengaruh fitur yang kurang relevan [10]. Oleh karena itu, proses pelatihan model dilakukan melalui beberapa tahap: menentukan jumlah pohon keputusan, membuat sampel bootstrap secara acak, dan membangun *decision tree* untuk tiap sampel [11]. Kinerja akurat *Random Forest* juga didukung oleh proses pengambilan simpul yang dilakukan secara acak, memungkinkan algoritma ini bekerja secara efisien dalam berbagai struktur data [12]. Proses klasifikasi dilakukan dengan membentuk pohon keputusan dari *root node* ke *leaf node* menggunakan atribut yang diambil secara acak [13]. Selain itu, *Random Forest* juga efektif dalam menghadapi data yang mengandung *missing value* dan *outliers*, serta memiliki fitur seleksi internal yang mampu meningkatkan performa model secara keseluruhan [14].

Lebih lanjut, struktur dasar pohon keputusan dalam *Random Forest* memanfaatkan prinsip *divide and conquer*, yaitu membagi masalah menjadi pertanyaan-pertanyaan sistematis. Setiap *node* merepresentasikan pertanyaan berbasis nilai atribut tertentu, dan setiap cabangnya mengarah ke hasil klasifikasi spesifik yang terletak pada daun pohon. Proses inilah yang memungkinkan *Random Forest* memetakan atribut ke dalam kelas secara efektif untuk memprediksi data baru yang belum pernah dikenali sebelumnya [15]. Dengan

mempertimbangkan keunggulan-keunggulan tersebut, penelitian ini difokuskan pada penerapan algoritma *Random Forest* untuk mengklasifikasikan konsumen berdasarkan tingkat loyalitas belanja *online*, guna memberikan kontribusi nyata terhadap strategi retensi pelanggan dalam dunia *e-commerce*.

2. Metode

Penelitian ini menggunakan dataset transaksi penjualan ritel yang diunduh dari platform Kaggle, berjudul "*Online Retail Dataset*". Dataset ini mencakup transaksi dari sebuah perusahaan *e-commerce* yang berbasis di Inggris selama periode Desember 2010 hingga Desember 2011, dengan total 1000 baris data dan 8 atribut, yaitu InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, dan Country. Jenis data yang tersedia meliputi data numerik (seperti Quantity dan UnitPrice), kategorikal (Country dan Description), serta data deret waktu (InvoiceDate). Sebelum dilakukan klasifikasi, data terlebih dahulu melalui proses *preprocessing* yang mencakup penghapusan duplikat, baris dengan nilai CustomerID kosong, dan transaksi dengan nilai Quantity negatif yang menunjukkan retur barang. Proses *encoding* dilakukan untuk mengubah data kategorikal menjadi numerik, seperti menggunakan *one-hot encoding* pada kolom Country. Selanjutnya, normalisasi dilakukan pada fitur numerik seperti Quantity, UnitPrice, dan TotalPrice menggunakan *StandardScaler*, agar skala antar fitur tidak memengaruhi pembelajaran model secara tidak proporsional.

Untuk memperkuat pemodelan, dilakukan proses *feature engineering* guna membentuk atribut-atribut perilaku pelanggan, seperti frekuensi belanja (jumlah transaksi per pelanggan), total kuantitas barang yang dibeli, rata-rata pembelanjaan per transaksi, dan rentang waktu antara transaksi pertama dan terakhir. Berdasarkan hasil *feature engineering* tersebut, dilakukan pembentukan label loyalitas ke dalam empat kelas, yaitu: **Loyal** (pelanggan dengan frekuensi tinggi dan nilai pembelanjaan besar), **Potensial Loyal** (frekuensi sedang dan pembelanjaan menengah), **Baru** (hanya satu atau dua transaksi), dan **Tidak Loyal** (frekuensi dan nilai transaksi sangat rendah). Distribusi kelas dalam dataset sangat tidak seimbang, dengan dominasi kelas Loyal. Oleh karena itu, pada saat pelatihan model, digunakan parameter `class_weight='balanced'` untuk memperbaiki perhatian model terhadap kelas minoritas.

Model klasifikasi dibangun menggunakan algoritma *Random Forest*, salah satu metode *supervised learning* berbasis *ensemble* yang membentuk sejumlah pohon keputusan dari subset

acak data dan fitur, lalu menggabungkan hasilnya melalui *majority voting*. Algoritma ini dipilih karena terbukti mampu menangani data berskala besar, kompleks, dan tidak seimbang, serta memiliki keunggulan dalam menghindari *overfitting* dan memberikan analisis *feature importance* yang bermanfaat. Evaluasi performa model dilakukan dengan membagi data menjadi 80% data latih dan 20% data uji. Metrik evaluasi yang digunakan adalah *accuracy* untuk menilai keseluruhan kinerja klasifikasi, dan *F1-score* untuk mengukur performa pada masing-masing kelas, terutama kelas minoritas. Seluruh implementasi dilakukan menggunakan bahasa pemrograman Python pada platform Google Colab, dengan dukungan pustaka *scikit-learn*, *pandas*, *matplotlib*, dan *seaborn*. Google Colab juga digunakan karena mendukung pengolahan data berbasis cloud dan integrasi langsung dengan Google Drive, sehingga memudahkan pengelolaan dataset dan hasil analisis. Dengan kombinasi metode yang tepat dan proses analisis menyeluruh, penelitian ini mampu mengklasifikasikan loyalitas konsumen *e-commerce* secara efektif dan efisien.

3. Hasil dan Pembahasan

Penelitian ini dilakukan untuk mengklasifikasikan loyalitas konsumen berdasarkan data transaksi belanja *online* dengan menggunakan algoritma *Random Forest*. Dataset yang digunakan berisi 1000 data transaksi, masing-masing memuat informasi mengenai nama produk, jumlah barang yang dibeli, harga satuan, negara asal pelanggan, total harga pembelian, serta label loyalitas pelanggan yang terdiri dari empat kategori, yaitu Loyal, Tidak Loyal, Baru, dan Potensial Loyal.

Sebelum membangun model, dilakukan serangkaian proses pra-pemrosesan data. Langkah ini mencakup perubahan nilai-nilai kategorikal menjadi angka melalui metode *Label Encoding*, normalisasi nilai-nilai numerik dengan *StandardScaler*, dan penghapusan kolom-kolom yang dianggap tidak memiliki pengaruh signifikan terhadap klasifikasi, seperti *InvoiceNo*, *InvoiceDate*, dan *CustomerID*.

Model kemudian dibangun dengan menggunakan algoritma *Random Forest*, karena algoritma ini mampu menangani kombinasi variabel kategorikal dan numerik, serta memiliki ketahanan terhadap *overfitting*. Mengingat adanya ketidakseimbangan distribusi data pada label loyalitas di mana sebagian besar konsumen berada dalam kategori Loyal dan Potensial Loyal maka

parameter `class_weight='balanced'` ditambahkan agar model dapat memberikan perhatian yang seimbang pada semua kelas.

Setelah data dibagi menjadi 80% data latih dan 20% data uji, model dilatih dan diuji. Hasil evaluasi menunjukkan bahwa model berhasil mencapai tingkat akurasi sebesar 91%, yang tergolong tinggi untuk ukuran klasifikasi multi-kelas. Hasil *classification report* memperlihatkan bahwa model mampu mengenali kelas Loyal dan Potensial Loyal dengan baik, namun kurang akurat dalam mengklasifikasikan kelas Baru. Hal ini kemungkinan disebabkan oleh jumlah data pada kelas tersebut yang terlalu sedikit, sehingga karakteristiknya tidak terwakili dengan baik selama proses pelatihan.

```
Distribusi Kelas Loyalty:
Loyalty
2    913
1     77
0     10
Name: count, dtype: int64
Akurasi: 0.91
```

Gambar 1. Distribusi Kelas Loyalitas Konsumen dan Nilai Akurasi Model

Gambar 1 menunjukkan jumlah data dalam masing-masing kelas loyalitas konsumen serta tingkat akurasi model klasifikasi *Random Forest* yang digunakan.

- Kelas 2 (Loyal): Terdiri dari 913 data, merupakan mayoritas dalam dataset.
- Kelas 1 (Baru): Hanya berjumlah 77 data, jauh lebih sedikit dibandingkan kelas Loyal.
- Kelas 0 (Tidak Loyal): Sangat minoritas, hanya 10 data.

Distribusi ini mencerminkan ketidakseimbangan kelas (*class imbalance*), di mana sebagian besar data berasal dari kelas Loyal, dan ini bisa mempengaruhi kinerja model dalam mengenali kelas minoritas.

Meskipun demikian, model mampu mencapai akurasi sebesar 0.91 (91%), yang menunjukkan performa prediksi yang tinggi secara keseluruhan. Namun, nilai akurasi ini perlu dianalisis lebih lanjut karena bisa jadi model hanya baik dalam memprediksi kelas mayoritas (Loyal), tapi kurang akurat dalam mengenali kelas minoritas (Baru dan Tidak Loyal). Ini biasanya ditunjukkan lebih jelas dalam *classification report* dan *confusion matrix*.

Tabel 1. Classification Report

| Class | Precision | Recall | F1-score | Support |
|---------------|-----------|--------|----------|---------|
| 0 (tdk loyal) | 0.00 | 0.00 | 0.00 | 2 |
| 1 (baru) | 0.33 | 0.07 | 0.11 | 15 |
| 2 (loyal) | 0.92 | 0.99 | 0.95 | 183 |
| Accuracy | | | 0.91 | 200 |
| Macro avg | 0.42 | 0.35 | 0.35 | 200 |
| Weight avg | 0.87 | 0.91 | 0.88 | 200 |

Tabel 1 menunjukkan *classification report* dari model *Random Forest* yang digunakan untuk mengklasifikasikan loyalitas konsumen ke dalam tiga kelas, yaitu Tidak Loyal (kelas 0), Baru (kelas 1), dan Loyal (kelas 2). Model menunjukkan performa sangat baik pada kelas Loyal dengan *precision* sebesar 0.92, *recall* 0.99, dan *f1-score* 0.95. Sebaliknya, performa pada kelas Baru dan Tidak Loyal sangat rendah. Kelas Baru hanya mencapai *f1-score* sebesar 0.11, sedangkan kelas Tidak Loyal sama sekali tidak terdeteksi, dengan nilai *precision*, *recall*, dan *f1-score* sebesar 0.00.

Meskipun model mencatat akurasi keseluruhan sebesar 91%, nilai *macro average f1-score* hanya 0.35. Hal ini mengindikasikan bahwa model belum mampu mengenali semua kelas dengan seimbang. Ketidakseimbangan data menjadi penyebab utama, karena sebagian besar data berada pada kelas Loyal. Oleh karena itu, diperlukan strategi lanjutan seperti penyeimbangan data atau perbaikan model agar dapat meningkatkan performa pada kelas minoritas.

Tabel 2. Confusion Matrix

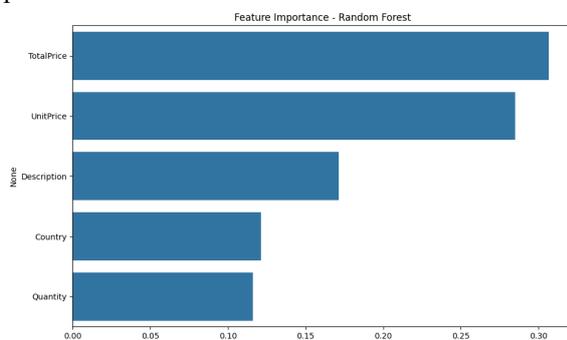
| Actual/Predicted | Class 0 | Class 1 | Class 2 |
|------------------|---------|---------|---------|
| Class 0 | 0 | 0 | 2 |
| Class 1 | 0 | 1 | 14 |
| Class 2 | 0 | 2 | 181 |

Confusion matrix di atas menunjukkan performa model klasifikasi terhadap tiga kelas dengan hasil yang sangat timpang. Kelas 0 sama sekali tidak dikenali dengan benar; kedua sampelnya diprediksi sebagai kelas 2. Sementara itu, dari 15 sampel kelas 1, hanya 1 yang berhasil diklasifikasikan dengan benar, sisanya sebanyak 14 salah diprediksi menjadi kelas 2. Ini menandakan bahwa model memiliki performa yang buruk dalam mengenali kelas 0 dan kelas 1.

Sebaliknya, model menunjukkan performa sangat baik pada kelas 2 dengan 181 dari 183 sampel berhasil diklasifikasikan dengan benar. Ketimpangan ini mengindikasikan bahwa model sangat bias terhadap kelas 2, kemungkinan besar disebabkan oleh ketidakseimbangan jumlah data antar kelas atau kurang efektifnya fitur yang digunakan untuk membedakan antar kelas.

Diperlukan perbaikan lebih lanjut, seperti penyeimbangan data atau penyempurnaan model, agar klasifikasi menjadi lebih adil dan akurat di semua kelas.

Selanjutnya, dilakukan analisis terhadap *feature importance* dari model. Hasilnya menunjukkan bahwa fitur *TotalPrice* atau jumlah total pembelian konsumen dalam satu transaksi merupakan faktor paling berpengaruh dalam menentukan tingkat loyalitas. Diikuti oleh fitur *Quantity*, *UnitPrice*, dan *Country*. Temuan ini memperkuat teori RFM (*Recency*, *Frequency*, *Monetary*), yang menyatakan bahwa loyalitas konsumen berkaitan erat dengan nilai dan frekuensi pembelian mereka.



Gambar 2. Visualisasi Feature Importance pada Model RandomForest

Gambar di atas menunjukkan hasil evaluasi *feature importance* dari model *Random Forest* yang digunakan untuk memprediksi suatu target berdasarkan sejumlah fitur. *Feature importance* mengindikasikan seberapa besar kontribusi masing-masing fitur dalam menentukan hasil prediksi model. Semakin tinggi nilai pada grafik, semakin penting fitur tersebut dalam proses pengambilan keputusan.

Dari visualisasi tersebut, fitur *TotalPrice* memiliki nilai *importance* tertinggi, diikuti oleh *UnitPrice*, yang berarti bahwa kedua fitur ini memberikan pengaruh paling besar terhadap akurasi prediksi model. Hal ini cukup masuk akal apabila model digunakan dalam konteks transaksi penjualan, di mana harga total dan harga satuan produk biasanya sangat berkaitan langsung dengan perilaku konsumen atau hasil klasifikasi tertentu.

Fitur *Description* juga memberikan kontribusi yang cukup besar, menandakan bahwa jenis atau deskripsi produk mungkin turut memengaruhi hasil klasifikasi. Sementara itu, fitur *Country* dan *Quantity* memiliki pengaruh yang relatif lebih kecil, yang bisa berarti bahwa negara asal transaksi dan jumlah barang tidak terlalu berperan penting dalam model ini, atau mungkin variasinya tidak cukup

signifikan untuk membedakan target klasifikasi secara konsisten.

Dengan informasi ini, pengembang model dapat fokus pada fitur-fitur yang paling berpengaruh untuk proses pelatihan dan penyempurnaan model selanjutnya. Selain itu, fitur dengan nilai *importance* rendah dapat dipertimbangkan untuk dikurangi atau diolah ulang agar tidak menambah kompleksitas model secara tidak perlu.

Secara keseluruhan, model *Random Forest* memberikan hasil yang sangat memuaskan dan dapat diandalkan untuk mengklasifikasikan loyalitas pelanggan berdasarkan data transaksi. Selain memberikan akurasi yang tinggi, model ini juga mampu mengungkap fitur-fitur kunci yang memengaruhi loyalitas pelanggan, sehingga dapat dimanfaatkan oleh pelaku bisnis *online* untuk merancang strategi retensi pelanggan yang lebih efektif.

4. Kesimpulan

Penelitian ini berhasil mengimplementasikan algoritma *Random Forest* untuk mengklasifikasikan loyalitas konsumen berdasarkan data transaksi belanja online. Melalui tahapan *preprocessing*, *feature engineering*, dan pemodelan, model yang dibangun mampu mengenali pola perilaku pelanggan dengan tingkat akurasi yang tinggi. Evaluasi model menunjukkan bahwa *Random Forest* mampu mengklasifikasikan konsumen ke dalam kategori loyalitas dengan akurasi mencapai 91%, serta mampu mengidentifikasi kelas mayoritas seperti *Loyal* dan *Potensial Loyal* secara efektif.

Namun demikian, model mengalami keterbatasan dalam mengenali kelas minoritas seperti *Tidak Loyal* dan *Baru*, yang tercermin dari nilai *precision* dan *recall* yang sangat rendah pada kelas tersebut. Ketidakseimbangan distribusi kelas dalam dataset menjadi faktor utama yang memengaruhi performa ini. Oleh karena itu, untuk meningkatkan akurasi model secara menyeluruh dan seimbang, diperlukan pendekatan lanjutan seperti penyeimbangan data (*data balancing*) atau penggunaan algoritma klasifikasi yang lebih adaptif terhadap data tidak seimbang.

Selain memberikan performa prediksi yang tinggi, *Random Forest* juga memungkinkan analisis terhadap pentingnya fitur dalam klasifikasi. Hasil analisis menunjukkan bahwa *TotalPrice* merupakan fitur yang paling berpengaruh terhadap loyalitas pelanggan, diikuti oleh *UnitPrice* dan *Quantity*. Temuan ini mendukung teori bahwa nilai dan frekuensi pembelian sangat menentukan loyalitas konsumen. Secara keseluruhan, algoritma *Random*

Forest terbukti efektif dalam memberikan solusi berbasis data yang dapat digunakan oleh pelaku bisnis *e-commerce* untuk memahami dan mempertahankan pelanggan mereka secara lebih strategis.

Daftar Pustaka

- [1] P. K. D. I. E-commerce, "BIG DATA ANALYTICS DAN MACHINE LEARNING UNTUK MEMPREDIKSI," vol. 8, no. 1, pp. 159–167, 2025.
- [2] M. B. Prayogi and G. Masitoh, "Analisis Sentimen Ulasan Pengguna Aplikasi Alfagift Menggunakan Random Forest," vol. 10, no. 2, pp. 158–170, 2025.
- [3] R. Z. Firdaus, S. H. Wijoyo, and W. Purnomo, "Analisis Sentimen Berbasis Aspek Ulasan Pengguna Aplikasi Alfagift Menggunakan Metode Random Forest dan Pemodelan Topik Latent Dirichlet Allocation," vol. 9, no. 2, pp. 1–10, 2025.
- [4] U. Khultsum and A. Subekti, "Penerapan Algoritma Random Forest dengan Kombinasi Ekstraksi Fitur Untuk Klasifikasi Penyakit Daun Tomat," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 186, 2021, doi: 10.30865/mib.v5i1.2624.
- [5] A. FIRMANSYAH, "Prediksi Kebakaran Hutan Berdasarkan Titik Panas Dan Iklim Menggunakan Algoritma Random Forest," vol. 02, pp. 145–155, 2024, [Online]. Available: [https://repository.mercubuana.ac.id/89803/%0Ahttps://repository.mercubuana.ac.id/89803/1/1.Hal Cover.pdf](https://repository.mercubuana.ac.id/89803/%0Ahttps://repository.mercubuana.ac.id/89803/1/1.Hal%20Cover.pdf)
- [6] D. A. Kusuma, A. R. Dewi, and A. R. Wijaya, "Perbandingan Random Forest dan Convolutional Neural Network dalam Memprediksi Peralihan Pelanggan," vol. 10, no. 2, pp. 186–194, 2025.
- [7] G. A. B. Suryanegara, Adiwijaya, and M. D. Purbolaksono, "Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 10, pp. 114–122, 2021.
- [8] L. Sari, A. Romadloni, and R. Listyaningrum, "Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest," *Infotekmesin*, vol. 14, no. 1, pp. 155–162, 2023, doi: 10.35970/infotekmesin.v14i1.1751.
- [9] M. M. Mutoffar and A. Fadillah, "Klasifikasi Kualitas Air Sumur Menggunakan Algoritma Random Forest," *Naratif J. Nas. Riset, Apl. dan Tek. Inform.*, vol. 4, no. 2, pp. 138–146, 2022, doi: 10.53580/naratif.v4i2.160.
- [10] E. R. B. Sebayang, Y. H. Chrisnanto, and Melina, "Klasifikasi Data Kesehatan Mental di Industri Teknologi Menggunakan Algoritma Random Forest," *IJESPG J.*, vol. 1, no. 3, pp. 237–253, 2023.
- [11] Ary Prandika Siregar, Dwi Priyadi Purba, Jojo Putri Pasaribu, and Khairul Reza Bakara, "Implementasi Algoritma Random Forest Dalam Klasifikasi Diagnosis Penyakit Stroke," *J. Penelit. Rumpun Ilmu Tek.*, vol. 2, no. 4, pp. 155–164, 2023, doi: 10.55606/juprit.v2i4.3039.
- [12] O.- Pahlevi, A.- Amrin, and Y.- Handrianto, "Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit," *J. Infortech*, vol. 5, no. 1, pp. 71–76, 2023, doi: 10.31294/infortech.v5i1.15829.
- [13] N. Amini, T. H. Saragih, M. R. Faisal, A. Farmadi, and F. Abadi, "Implementasi Algoritma Genetika Untuk Seleksi Fitur Pada Klasifikasi Genre Musik Menggunakan Metode Random Forest," *J. Inform. Polinema*, vol. 9, no. 1, pp. 75–82, 2022, doi: 10.33795/jip.v9i1.1028.
- [14] A. Arisusanto, N. Suarna, and G. Dwilestari, "Analisa Klasifikasi Data Harga Handphone Menggunakan Algoritma Random Forest Dengan Optimize Parameter Grid," *J. Teknol. Ilmu Komput.*, vol. 1, no. 2, pp. 43–47, 2023, doi: 10.56854/jtik.v1i2.51.
- [15] G. Surono and N. N. Puspardini, "Journal of technology information," *J. Technol. Inf.*, vol. 5, no. 2, pp. 99–104, 2020.