

Penerapan Model Decision Tree untuk Klasifikasi Rating dan Rekomendasi Produk Pakaian pada Platform E-Commerce

**^{1*}Putri Ameliya, ²Anis Faizul Laila, ³M. Abdilah Saputra, ⁴Rizal Syihab
Saputra Adam**

¹⁻⁴Sistem Informasi, Universitas Nusantara PGRI Kediri

E-mail : ^{*1}pameliya977@gmail.com, ²anisfaizul879@gmail.com, ³abdilahsaputra@gmail.com,
⁴rizalsyihab07@gmail.com

Penulis Korespondens: Anis Faizul Laila

Abstrak—Pertumbuhan pesat industri fashion di platform *e-commerce* menuntut adanya sistem rekomendasi yang mampu menyesuaikan dengan preferensi pengguna. Rekomendasi berbasis *rating* sering kali tidak mempertimbangkan hubungan antara fitur produk dan preferensi subjektif konsumen. Penelitian ini mengembangkan sistem rekomendasi produk fashion berbasis klasifikasi *rating* menggunakan algoritma *Decision Tree* yang dikombinasikan dengan metode *content-based filtering*. *Dataset* yang digunakan terdiri dari berbagai fitur seperti nama produk, merek, kategori, warna, ukuran, dan harga. Proses klasifikasi dilakukan untuk memprediksi *rating*, dilanjutkan dengan penyaringan produk berdasarkan kemiripan konten menggunakan *cosine similarity* dan pencarian berbasis kata kunci melalui *TF-IDF*. Hasil penelitian menunjukkan bahwa sistem mampu memberikan rekomendasi yang lebih relevan dan sesuai dengan kebutuhan pengguna. Pendekatan ini penting untuk meningkatkan efisiensi pencarian produk, pengalaman pengguna, dan mendukung pengambilan keputusan yang lebih tepat dalam platform *e-commerce*.

Kata Kunci—Pohon Keputusan, Klasifikasi Rating, Sistem Rekomendasi

Abstract—The rapid growth of the fashion industry on e-commerce platforms demands a recommendation system that is able to adapt to user preferences. Rating-based recommendations often do not consider the relationship between product features and subjective consumer preferences. This study develops a fashion product recommendation system based on rating classification using the Decision Tree algorithm combined with the content-based filtering method. The dataset used consists of various features such as product name, brand, category, color, size, and price. The classification process is carried out to predict the rating, followed by product filtering based on content similarity using cosine similarity and keyword-based search using TF-IDF. The results of the study show that the system is able to provide recommendations that are more relevant and in accordance with user needs. This approach is important to improve product search efficiency, user experience, and support more precise decision making on e-commerce platforms.

Keywords—Decision Tree, Rating Classification, Recommendation system

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Perkembangan *e-commerce* mendorong transformasi digital dalam industri *fashion*, di mana pakaian kini tidak hanya dianggap sebagai kebutuhan pokok, tetapi juga sebagai bagian dari gaya hidup dan ekspresi diri. Banyaknya pilihan produk seperti merek, warna, ukuran, dan harga

sering kali membuat konsumen kesulitan memilih produk yang benar-benar sesuai dengan preferensi dan kebutuhan pribadi [1]. Untuk membantu konsumen dalam proses ini, sistem rekomendasi menjadi salah satu solusi penting agar pengalaman berbelanja menjadi lebih mudah, cepat, dan personal.

Beberapa penelitian sebelumnya telah memanfaatkan algoritma *machine learning* untuk membangun sistem rekomendasi. Penelitian oleh Suranda dan Nugroho (2024) [2] menggunakan *Decision Tree ID3* untuk memprediksi penjualan pakaian muslim dan berhasil mengidentifikasi produk populer dengan akurasi 88,24%, sementara Hayati et al [3] menerapkan ID3 untuk memetakan preferensi pengguna dengan akurasi 85%. Penelitian oleh Yang (2022) Pendekatan yang lebih kompleks seperti *multilabel classification* dan kombinasi *Decision Tree* dengan *deep learning* juga telah digunakan untuk mengklasifikasikan gaya pakaian [4]. Selain itu, Oktavirahani dan Maharesi (2022) [5] menggunakan metode *CART* untuk merekomendasikan ukuran pakaian berdasarkan data fisik pengguna. Sedangkan Mukhsinin et al. (2024) menerapkan *Decision Tree* juga terbukti efektif di berbagai platform. Studi pada *Netflix* menunjukkan peningkatan personalisasi layanan [6], dan efektivitas lintas sektor terlihat dari penerapannya di bidang *e-commerce* dan hiburan [7]. Dalam skala yang lebih kecil seperti UMKM, [8] menggunakan *Decision Tree* untuk segmentasi pelanggan berdasarkan usia dan pendapatan, dan [9] membangun sistem rekomendasi buku dengan akurasi tinggi. Penelitian oleh [10] mengaplikasikan C4.5 dalam analisis kepuasan pelanggan di platform Shopee, dengan hasil menunjukkan bahwa atribut produk dan kecepatan pengiriman berpengaruh signifikan terhadap rating pengguna.

Namun demikian, masih terdapat beberapa kekurangan dari penelitian-penelitian sebelumnya. Sebagian besar belum mempertimbangkan bahwa penilaian terhadap produk (rating) bisa sangat subjektif tergantung pada preferensi masing-masing pengguna. Selain itu, hubungan langsung antara fitur produk (seperti warna, bahan, atau ukuran) dengan hasil rating sering kali belum dimanfaatkan secara maksimal dalam proses klasifikasi dan rekomendasi. Berdasarkan hal tersebut, penelitian ini bertujuan membangun sistem rekomendasi produk fashion dengan dua pendekatan utama: pertama, melakukan klasifikasi rating produk menggunakan algoritma *DecisionTreeClassifier*, dan kedua, menerapkan metode *content-based filtering* dengan pendekatan *cosine similarity* dan pencarian kata kunci menggunakan *TF-IDF*. Sistem ini juga dilengkapi dengan filter rating minimal 4 untuk memastikan produk yang direkomendasikan memiliki kualitas baik.

Kebaruan dari penelitian ini terletak pada integrasi antara klasifikasi rating produk dan sistem rekomendasi berbasis konten, dengan kombinasi teknik *machine learning* dan pencocokan kata kunci secara simultan. Pendekatan ini belum banyak dikaji dalam konteks produk fashion di platform e-commerce, terutama dengan penerapan filter kualitas berbasis rating dan fitur produk yang relevan.

II. METODE

Penelitian ini menggunakan pendekatan klasifikasi dengan algoritma *Decision Tree* sebagai bagian dari sistem rekomendasi produk fashion. Klasifikasi merupakan teknik untuk membagi data ke dalam kelompok tertentu berdasarkan karakteristik atau atribut yang dimilikinya. Pendekatan ini membantu sistem dalam mengidentifikasi kategori yang sesuai untuk setiap data masukan dengan merujuk pada pola yang telah ada sebelumnya[11]. Analisis dilakukan menggunakan Google Colab dengan bahasa pemrograman Python serta pustaka (*library*) seperti

Scikit-learn, Pandas, dan Numpy [12][13]. Proses penelitian mengacu pada model CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yaitu model bertahap yang bersifat berulang dan fleksibel, yang digunakan untuk menggali pengetahuan dari data[14].

2.1 Bussiness Understanding

Pada tahap ini, peneliti memahami apa yang ingin dicapai dan dibutuhkan oleh proyek dari sisi bisnis. Setelah itu, pemahaman tersebut diterjemahkan menjadi masalah yang bisa diselesaikan dengan data mining, selanjutnya dibuat rencana awal untuk mencapainya. Dalam konteks penelitian ini, tujuan utamanya adalah untuk memberikan rekomendasi produk yang sesuai.

2.2 Data Understanding

Tahap ini diawali dengan mengumpulkan data, lalu dilanjutkan dengan berbagai kegiatan untuk memahami data lebih dalam, seperti mengecek apakah ada masalah pada kualitas data dan mencari gambaran awal dari isi data tersebut. Proses ini mirip dengan *Exploratory Data Analysis (EDA)*, yang bertujuan membantu mengenali data secara menyeluruh, menemukan pola-pola yang mungkin tersembunyi, serta menganalisis data tanpa langsung menggunakan teknik statistik atau pemodelan. Selain itu, tahap ini juga digunakan untuk mendeteksi kesalahan, mengenali jenis data, dan melihat hubungan antar variabel. Dataset diambil dari situs Kaggle dan berisi 1001 data produk fashion. Setiap data mencakup informasi seperti ID pengguna, nama produk, merek, kategori, harga, penilaian (*rating*), warna, dan ukuran. Data ini menjadi dasar dalam membangun model klasifikasi dan sistem rekomendasi.

2.3 Data Preparation

Data preparation atau *data preprocessing* adalah tahap di mana data disiapkan sebelum digunakan dalam pemodelan. Proses ini mencakup semua langkah yang diperlukan untuk mengubah data mentah menjadi dataset akhir yang siap digunakan oleh model. Tahapan ini sering kali harus dilakukan berulang kali dan tidak selalu mengikuti urutan yang tetap. Sebelum membangun model, dilakukan proses encoding terhadap data kategorikal seperti brand, warna, dan ukuran menggunakan *Label Encoder*. Selanjutnya, data dipisahkan menjadi fitur (X) dan target (y), dengan rating sebagai target klasifikasi. Data kemudian dibagi menjadi data latih dan data uji menggunakan *train_test_split*.

2.4 Modeling

Pada tahap ini, memilih dan menerapkan beberapa teknik pemodelan dipilih dan mengkalibrasi parameter sehingga memperoleh nilai optimal. Pemodelan pada penelitian ini fokus pada penggunaan metode berbasis Decision Tree.

2.4.1 Algoritma Decision Tree

Decision Tree atau pohon keputusan adalah metode pembelajaran dengan panduan yang membentuk struktur bercabang seperti pohon. Setiap titik cabang mewakili atribut, jalur menunjukkan keputusan, dan ujungnya menunjukkan hasil klasifikasi. Metode ini mudah dipahami dan cukup efektif dalam berbagai kasus klasifikasi [15][16]. Persamaan umum dalam *Decision tree* dari penelitian ini sebagai berikut :

$$f(x) = \sum_{i=1}^n c_i \cdot 1(x \in R_i)^{(l)}$$

Keterangan:

- $f(x)$ = prediksi yang dihasilkan oleh decision tree untuk input x.
- N = jumlah simpul daun.
- c_i = nilai kelas yang diberikan pada simpul daun ke-i.
- R_i = wilayah yang diwakili oleh simpul daun ke-i.

2.5 Evaluasi Model

Model klasifikasi dievaluasi menggunakan ukuran seperti akurasi, *precision*, *recall*, dan *F1-score*. Nilai-nilai ini dihitung dari hasil pengujian model terhadap data uji. Evaluasi dilakukan dengan bantuan fungsi *classification_report* dari pustaka *sklearn.metrics* untuk mengetahui kinerja model dalam setiap kelas rating. Akurasi keseluruhan digunakan sebagai indikator utama keberhasilan klasifikasi.

2.6 Sistem Rekomendasi

Setelah klasifikasi, hanya produk dengan rating ≥ 4 yang dipilih sebagai kandidat rekomendasi. Sistem rekomendasi dibangun dengan dua pendekatan:

- Rekomendasi berbasis konten menggunakan *cosine similarity* pada fitur seperti nama produk, merek, kategori, warna, dan ukuran.
- Pencarian berbasis kata kunci menggunakan *TF-IDF vectorizer* untuk mencocokkan input pengguna dengan deskripsi produk.

Pendekatan ini menggabungkan klasifikasi dengan penyaringan berbasis konten untuk meningkatkan kualitas rekomendasi yang diberikan kepada pengguna.

III. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi produk fashion menggunakan algoritma *Decision Tree* untuk klasifikasi *rating* dan *Content-Based Filtering* untuk rekomendasi berdasarkan fitur produk serta kata kunci pengguna. Implementasi dilakukan menggunakan *Python* dengan pustaka *Scikit-learn* dan *Pandas* pada dataset *fashion_products.csv*. Bagian ini menyajikan hasil *pra-pemrosesan data*, klasifikasi *rating*, evaluasi *model*, performa sistem rekomendasi, serta analisis implikasinya.

3.1 Data Understanding

Dataset *fashion_products.csv* berisi 1.000 baris data dengan 9 kolom: *User ID*, *Product ID*, *Product Name*, *Brand*, *Category*, *Price*, *Rating*, *Color*, dan *Size*. Informasi dari *df.info()* menunjukkan tidak ada data yang hilang, sehingga dataset siap digunakan. Statistik deskriptif dari *df.describe()* mengungkapkan:

- Harga produk berkisar antara 10 hingga 100 (rata-rata 55,78).
- Rating produk berkisar dari 1,00 hingga 4,99 (rata-rata 2,99).
- *User ID* dan *Product ID* memiliki distribusi merata, menunjukkan variasi pengguna dan produk yang representatif.

Fitur kategorikal (*Product Name, Brand, Category, Color, Size*) dienkode menggunakan *LabelEncoder* untuk mengubahnya menjadi nilai numerik. *Rating* diklasifikasikan menjadi kelas diskrit (1 hingga 5) dengan pembulatan. Penggunaan *LabelEncoder* efektif untuk pemrosesan, tetapi dapat menyebabkan *asumsi ordinalitas* pada *fitur* seperti *Color*, yang mungkin memengaruhi model. Alternatif seperti *One-Hot Encoding* dapat dipertimbangkan di penelitian lanjutan.

3.2 Modeling

Model *Decision Tree Classifier* dilatih untuk memprediksi kelas rating berdasarkan fitur *Product Name, Brand, Category, Color, Size*, dan *Price*. Dataset dibagi menjadi 80% data pelatihan dan 20% data pengujian menggunakan *train_test_split*. Model menggunakan parameter default (*max_depth=None*, *min_samples_split=2*) untuk mengeksplorasi struktur data secara maksimal. Hasil awal menunjukkan akurasi keseluruhan sebesar 81%, mengindikasikan performa yang baik dalam memprediksi rating produk.

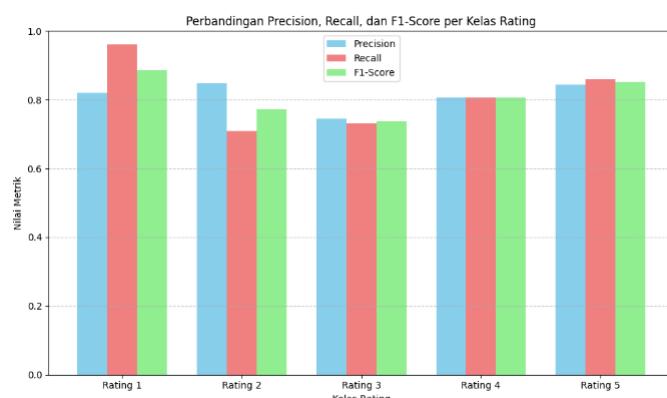
3.3 Evaluasi Model

Performa model *Decision Tree* dievaluasi menggunakan metrik *precision*, *recall*, dan *F1 score* untuk setiap kelas rating (1 hingga 5). Tabel 1 menyajikan laporan klasifikasi lengkap berdasarkan data pengujian.

Tabel 1. Evaluasi Model Klasifikasi

precision	recall	f1-score	support
1	0.79	0.96	0.87
2	0.79	0.69	0.74
3	0.76	0.73	0.75
4	0.80	0.79	0.80
5	0.88	0.86	0.87
accuracy		0.81	273
macro avg	0.80	0.81	273
weighted avg	0.81	0.81	0.80

Akurasi keseluruhan 81% menunjukkan bahwa model mampu memprediksi *rating* produk dengan baik. Untuk memvisualisasikan *performa model* secara lebih jelas, Gambar 1 membandingkan *precision*, *recall*, dan *F1-score* untuk setiap kelas *rating*.



Gambar 1 : Perbandingan Metrik Klasifikasi

Gambar 1 menunjukkan bahwa model memiliki *recall* tertinggi (0,96) untuk *Rating* 1, yang berarti model sangat efektif dalam mengidentifikasi produk dengan rating rendah, penting untuk menghindari rekomendasi produk berkualitas buruk. Sebaliknya, *recall* terendah (0,69) untuk *Rating* 2 mengindikasikan kesulitan model dalam memprediksi *rating* menengah, mungkin karena pola fitur yang kurang jelas atau ketidakseimbangan data pada kelas ini. *Rating* 5 memiliki *precision* tertinggi (0,88), menunjukkan bahwa model jarang salah mengklasifikasikan produk berkualitas tinggi, yang mendukung filter rating minimal 4 pada sistem rekomendasi. *F1-Score* tertinggi (0,87) pada *Rating* 1 dan 5 menegaskan performa kuat model untuk rating ekstrem dibandingkan rating menengah (2 dan 3).

Kelemahan pada *Rating* 2 dapat diatasi dengan teknik seperti *oversampling* untuk kelas minoritas atau penyesuaian parameter *Decision Tree* (misalnya, membatasi *max_depth* untuk mencegah *overfitting*). Hasil ini menunjukkan bahwa model dapat diandalkan untuk mendukung sistem rekomendasi, terutama dalam memilih produk berkualitas tinggi berdasarkan rating yang diprediksi.

3.4 Sistem Rekomendasi Berbasis Content-Based Filtering

Sistem *Content-Based Filtering* menggunakan *cosine similarity* untuk merekomendasikan produk berdasarkan kemiripan fitur, yaitu *Product Name*, *Brand*, *Category*, *Color*, *Size*. Rekomendasi hanya ditampilkan jika produk memiliki rating minimal 4 untuk menjaga kualitas hasil. Contoh hasil rekomendasi ditunjukkan pada Tabel 2, dengan produk acuan diambil dari indeks ke-5 pada dataset uji (*X_test*).

Tabel 2. Detail Data Produk

Product Details						
No	Product Nama	Brand	Category	Color	Size	Rating Class
1	Shoes	Gucci	Women's Fashion	Red	XL	4
2	Jeans	Gucci	Men's Fashion	White	XL	5
3	Sweater	Adidas	Kids' Fashion	White	XL	5
4	Dress	Adidas	Women's Fashion	Yellow	XL	4
5	Dress	Gucci	Women's Fashion	Yellow	S	4

Sistem ini menghasilkan rekomendasi yang relevan berdasarkan fitur seperti *Brand* dan *Color*, tetapi keberagaman kategori (misalnya, *Shoes* dan *Dress*) menunjukkan bahwa bobot fitur perlu disesuaikan untuk meningkatkan kecocokan kategori produk.

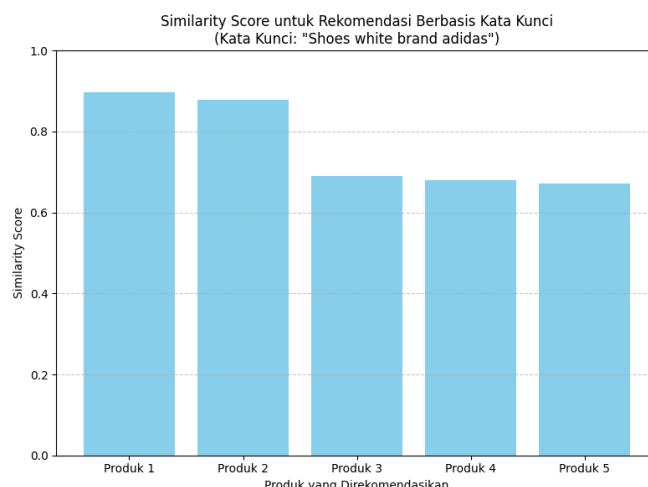
3.5 Sistem Rekomendasi Berbasis Kata Kunci

Sistem berbasis kata kunci menggunakan *TF-IDF Vectorizer* untuk mengubah deskripsi produk menjadi vektor numerik, dengan *cosine similarity* untuk mencocokkannya dengan *keyword* pengguna. Contoh hasil untuk *keyword* “Shoes white brand adidas” ditunjukkan pada Tabel 3.

Tabel 3. Rekomendasi Sepatu

Product Sepatu						
Product Name	Brand	Category	Color	Size	Rating Class	Similarity Score
Shoes	Adidas	Kids' Fashion	White	S	5	0.896202
Shoes	Adidas	Men's Fashion	White	S	4	0.878616
Shoes	H&M	Kids' Fashion	White	L	4	0.689488
Shoes	H&M	Women's Fashion	White	M	5	0.678810
Shoes	H&M	Men's Fashion	White	S	5	0.670953

Sistem ini efektif dalam menampilkan produk relevan dengan kemiripan tinggi terhadap kata kunci/keyword, khususnya dari brand Adidas berwarna putih.



Gambar 2 : Grafik Skor Rekomendasi

IV. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem rekomendasi produk fashion yang mengintegrasikan algoritma *Decision Tree* untuk klasifikasi rating dan pendekatan *Content-Based Filtering* untuk rekomendasi berbasis fitur produk serta kata kunci pengguna, sesuai dengan tujuan penelitian untuk meningkatkan pengalaman pengguna dalam platform *e-commerce* fashion. Sistem ini mampu memprediksi rating produk dengan akurasi 81%, menunjukkan kemampuan yang baik dalam mengidentifikasi produk berkualitas tinggi, yang mendukung rekomendasi relevan dengan filter rating minimal 4. Pendekatan *Content-Based Filtering* menghasilkan rekomendasi yang sesuai berdasarkan kemiripan fitur, sementara sistem berbasis kata kunci secara efektif mencocokkan input pengguna dengan produk yang relevan, meskipun memerlukan perbaikan untuk menghilangkan duplikasi hasil. Dalam konteks teknik industri, penelitian ini berkontribusi pada optimalisasi sistem rekomendasi *e-commerce* dengan

memanfaatkan teknik *machine learning* untuk meningkatkan efisiensi pencarian produk dan kepuasan pelanggan. Secara umum, pendekatan ini memperkaya *sains data* dengan menunjukkan integrasi klasifikasi dan rekomendasi berbasis konten dalam aplikasi praktis. Untuk pengembangan lebih lanjut, disarankan untuk mengintegrasikan data preferensi pengguna, menerapkan filter duplikasi berbasis *Product ID*, dan menggunakan algoritma seperti *Random Forest* untuk meningkatkan akurasi klasifikasi, serta menguji sistem pada *dataset* yang lebih besar untuk memastikan skalabilitas di platform *e-commerce* skala besar.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada Prodi Sistem Informasi UNP Kediri, dosen mata kuliah Data Mining, serta semua pihak yang telah memberikan dukungan selama proses penyusunan artikel ini. Harapannya, tulisan ini dapat memberikan manfaat dan inspirasi bagi para pembaca.

DAFTAR PUSTAKA

- [1] L. Rajput and S. Singh, *Customer Segmentation of E-commerce data using K-means Clustering Algorithm*. 2023. doi: 10.1109/Confluence56041.2023.10048834.
- [2] D. I. Suranda and A. Nugroho, “KLASIFIKASI DATA PENJUALAN UNTUK MEMPREDIKSI TINGKAT PENJUALAN PRODUK MENGGUNAKAN METODE DECISION TREE,” *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, vol. 7, no. 1, p. 370, Jun. 2024, doi: 10.37600/tekinkom.v7i1.1269.
- [3] P. H. S. A. I. H. Nur Hayati, “Penerapan Metode Algoritma ID3 Decision Tree untuk Rekomendasi Pemilihan Baju,” *JICTE (Journal of Information and Computer Technology Education)*, vol. 4, no. 2, Dec. 2020, doi: 10.21070/jict.e.v4i2.1019.
- [4] B. Yang, “Clothing Design Style Recommendation Using Decision Tree Algorithm Combined with Deep Learning,” *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/5745457.
- [5] F. A. Oktavirahani and R. Maharesi, “Implementasi Algoritma Decision Tree Cart Untuk Merekendasikan Ukuran Baju,” *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 1, p. 138, Feb. 2022, doi: 10.30865/jurikom.v9i1.3838.
- [6] D. A. Mukhsinin, M. Rafliansyah, S. A. Ibrahim, R. Rahmaddeni, and D. Wulandari, “Implementasi Algoritma Decision Tree untuk Rekomendasi Film dan Klasifikasi Rating pada Platform Netflix,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 570–579, Mar. 2024, doi: 10.57152/malcom.v4i2.1255.
- [7] Fadil Indra Sanjaya and Anna Dina Kalifia, “Precision Marketing Model using Decision Tree on e-Commerce Case Study Orebae.com,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 5, pp. 1033–1039, Sep. 2023, doi: 10.29207/resti.v7i5.4531.
- [8] M. Maulidah *et al.*, “Algoritma Klasifikasi Decision Tree Untuk Rekomendasi Buku Berdasarkan Kategori Buku,” *JURNAL ILMIAH EKONOMI DAN BISNIS*, vol. 13, no. 2, pp. 89–96, 2020, doi: <https://doi.org/10.51903/e-bisnis.v13i2.251>.
- [9] A. Steffany, K. Aldzikri, M. Amar, R. Tricahya, and M. I. Arfiandono, “Penerapan Algoritma C4.5 Untuk Menganalisis Kepuasan Pelanggan E-Commerce Shopee,” *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 6, no. 6, 2023, doi: <https://doi.org/10.32672/jnkti.v6i6.7112>.
- [10] B. H. Mawaridi and M. Faisal, “Rekomendasi Merk Mobil Untuk Calon Pembeli Menggunakan Algoritma Decision Tree,” *Jurnal Informatika*, vol. 10, no. 2, pp. 157–162, Oct. 2023, doi: 10.31294/inf.v10i2.16000.
- [11] D. Putriani, A. Puspa, A. Prayogi, A. I. Shofyana, A. Ristyawan, and E. Daniati, “Prediksi Customer Churn Menggunakan Algoritma Decision Tree,” Online, 2024. doi: <https://doi.org/10.29407/inotek.v8i1.4914>.

- [12] A. Krisna Theo, A. G. Candra Andy, R. Achmad, E. Daniati, and A. Ristyawan, “Analisa Perbandingan Algoritma Classification Berdasarkan Komposisi Label,” Online, 2024. doi: <https://doi.org/10.29407/inotek.v8i1.4906>.
- [13] N. K. A. Ni Wayan Wardani, “Analisa Komparasi Algoritma Decision Tree C4.5 dan Naïve Bayes untuk Prediksi Churn Berdasarkan Kelas Pelanggan Retail,” *International Journal of Natural Sciences and Engineering.*, 2019, doi: <https://doi.org/10.23887/ijnse.v3i3.23113>.
- [14] A. Ristyawan, A. Nugroho, and T. K. Amarya, “Optimasi Preprocessing Model Random Forest Untuk Prediksi Stroke,” vol. 12, no. 1, pp. 29–44, 2025.
- [15] S. Haryati, A. Sudarsono, and E. Suryana, “IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA C4.5 (STUDI KASUS: UNIVERSITAS DEHASEN BENGKULU),” 2015. doi: <https://doi.org/10.37676/jmi.v11i2.260>.
- [16] N. A. Mustakim, M. Abdul Aziz, and S. A. Rahman, “Predicting Consumer Behavior in E-Commerce Using Decision Tree: A Case Study in Malaysia,” 2024. doi: [https://doi.org/10.22610/imbr.v16i3\(I\).3965](https://doi.org/10.22610/imbr.v16i3(I).3965).