



## Perancangan Sistem Rekomendasi Produk Menggunakan Metode Decision Tree Berdasarkan Karakteristik Pelanggan Dan Riwayat Transaksi Studi Kasus PT Diamond Interest

Ray Yoga Tantra<sup>1\*</sup>, Yogasetya Suhandi<sup>2</sup>, Ike Kurniati<sup>3</sup>

<sup>1-3</sup>Program Studi Sistem Informasi, Institut Teknologi dan Bisnis Swadharma, Indonesia

Alamat : Jl. Malaka No.3, RT.6/RW.2, Roa Malaka, Kec. Tambora, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta

Email: [raytantra99@gmail.com](mailto:raytantra99@gmail.com)<sup>1\*</sup>, [yogasety@yahoo.com](mailto:yogasety@yahoo.com)<sup>2</sup>, [ikekurniati@swadharma.ac.id](mailto:ikekurniati@swadharma.ac.id)<sup>3</sup>

Korespondensi penulis : [raytantra99@gmail.com](mailto:raytantra99@gmail.com)

**Abstract.** *Diamond Interest as a herbal product distribution company has customer characteristic data and transaction history that has the potential to be used to support data-driven marketing strategies. However, the data has not been optimized to generate product recommendations that suit customer needs. This research aims to design a product recommendation system using the Decision Tree method based on customer characteristics and transaction history to produce accurate and relevant recommendations. The research uses a quantitative approach with the CRISP-DM framework which includes Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The dataset used comes from data on customers, transactions, and herbal products of PT Diamond Interest for the period May 17, 2025 to May 17, 2026. The Decision Tree model was compared with Random Forest and Support Vector Machine (SVM), then evaluated using accuracy, precision, recall, and F1-score metrics. The results showed that Decision Tree provided the best performance with an accuracy value of 80.77%, accuracy of 80.92%, recall of 80.77%, and an F1-score of 80.58%, and was in a good fit condition with no indication of overfitting or underfitting. The model is implemented into a web-based information system using Streamlit which provides features of uploading datasets, data visualization, classification of customer health preferences, and herbal product recommendations. The resulting system is able to provide more targeted product recommendations to support data-driven marketing decision-making and increase the effectiveness of the company's marketing strategy.*

**Keywords:** *CRISP-DM; customer characteristics; Decision Tree; recommendation system; transaction history.*

**Abstrak.** Diamond Interest sebagai perusahaan distribusi produk herbal memiliki data karakteristik pelanggan dan riwayat transaksi yang berpotensi dimanfaatkan untuk mendukung strategi pemasaran berbasis data. Namun, data tersebut belum dioptimalkan untuk menghasilkan rekomendasi produk yang sesuai dengan kebutuhan pelanggan. Penelitian ini bertujuan merancang sistem rekomendasi produk menggunakan metode *Decision Tree* berdasarkan karakteristik pelanggan dan riwayat transaksi guna menghasilkan rekomendasi yang akurat dan relevan. Penelitian menggunakan pendekatan kuantitatif dengan kerangka kerja *CRISP-DM* yang meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. *Dataset* yang digunakan berasal dari data pelanggan, transaksi, dan produk herbal PT Diamond Interest periode 17 Mei 2025 hingga 17 Mei 2026. Model *Decision Tree* dibandingkan dengan *Random Forest* dan *Support Vector Machine* (SVM), kemudian dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa *Decision Tree* memberikan kinerja terbaik dengan nilai *accuracy* sebesar 80,77%, *precision* 80,92%, *recall* 80,77%, dan *F1-score* 80,58%, serta berada pada kondisi *good fit* tanpa indikasi *overfitting* maupun *underfitting*. Model tersebut diimplementasikan ke dalam sistem informasi berbasis web menggunakan *Streamlit* yang menyediakan fitur unggah *dataset*, visualisasi data, klasifikasi preferensi kesehatan pelanggan, dan rekomendasi produk herbal. Sistem yang dihasilkan mampu memberikan rekomendasi produk yang lebih tepat sasaran sehingga mendukung pengambilan keputusan pemasaran berbasis data dan meningkatkan efektivitas strategi pemasaran perusahaan.

**Kata kunci:** *CRISP-DM; Decision Tree; karakteristik pelanggan; riwayat transaksi; sistem rekomendasi.*

## 1. LATAR BELAKANG

Perkembangan teknologi informasi telah mendorong perusahaan di berbagai sektor, termasuk industri produk kesehatan herbal, untuk mengoptimalkan strategi pemasaran berbasis data guna meningkatkan daya saing dan loyalitas pelanggan (Denyka et al., 2022), mengingat karakteristik konsumen dan pola perilaku pembelian menjadi aspek penting yang perlu dipahami agar strategi penjualan dapat disesuaikan dengan kebutuhan pasar yang terus berubah (Utami et al., 2025). PT. Diamond Interest, sebagai perusahaan produk kesehatan herbal berskala menengah, menghadapi tantangan dalam memanfaatkan data karakteristik pelanggan dan riwayat transaksi untuk menghasilkan rekomendasi produk yang tepat sasaran, sehingga dibutuhkan suatu sistem yang mampu mengklasifikasikan preferensi pelanggan secara otomatis dan akurat (Mahariani et al., 2026).

Berbagai metode telah digunakan dalam penelitian terdahulu untuk membangun sistem rekomendasi produk, mulai dari *collaborative filtering* pada *platform marketplace* (Hariri & Rochim, 2022) dan *content-based filtering* untuk produk *skincare* (Nouvalina & Hati, 2024), hingga algoritma berbasis *data mining* seperti *K-Nearest Neighbor* untuk prediksi penjualan produk terlaris (Gautama & Arijanto, 2024; Sri Puspita Dewi et al., 2022) dan *Single Moving Average* untuk prediksi penjualan obat (Bela & Bhakti, 2022). Teknik *clustering* seperti *K-Means* juga banyak dimanfaatkan untuk segmentasi pelanggan berbasis *Recency, Frequency, Monetary* (RFM) (Anshary et al., 2022), pengelompokan perilaku konsumen (Nurhidayati et al., 2025), maupun identifikasi pola data pelanggan pada sektor jasa publik (Karin Annisa et al., 2022).

Algoritma *Decision Tree* menjadi salah satu metode klasifikasi yang banyak diminati karena kemampuannya menghasilkan aturan keputusan yang mudah diinterpretasikan, sebagaimana terbukti pada klasifikasi *rating* dan rekomendasi produk pakaian (Ameliya et al., 2025), rekomendasi film pada platform Netflix (Dimas et al., 2024), prediksi risiko kredit nasabah bank (Marsianus et al., 2025), dan sistem informasi perpustakaan (Yosua Partogi et al., 2021). Meskipun demikian, sebagian besar penelitian tersebut masih terbatas pada satu domain aplikasi tertentu dan belum banyak yang membandingkan kinerja *Decision Tree* secara langsung dengan algoritma klasifikasi lain seperti *Random Forest* dan *Support Vector Machine* (SVM), padahal perbandingan

tersebut penting untuk memastikan pemilihan model yang paling optimal sebelum diimplementasikan pada sistem nyata.

Penerapan sistem rekomendasi produk juga terbukti berdampak positif terhadap peningkatan minat beli pelanggan (Andriko et al., 2025), dan pemahaman terhadap pola transaksi digital dapat membantu perusahaan merancang strategi pemasaran yang lebih tepat (Juliansyah & Sutabri, 2024; Sitopu, 2025). Namun demikian, sebagian besar penelitian terdahulu masih berfokus pada analisis deskriptif atau statistik sederhana (Sasono & Ulfatun Nikmah, 2021) tanpa mengintegrasikan hasil analisis ke dalam sistem aplikatif seperti *dashboard* interaktif berbasis web, padahal kerangka kerja *Streamlit* telah terbukti efektif digunakan untuk membangun aplikasi visualisasi data interaktif (Farid et al., 2025).

Berdasarkan permasalahan tersebut, PT. Diamond Interest membutuhkan sistem rekomendasi produk berbasis klasifikasi yang tidak hanya akurat, tetapi juga dapat divisualisasikan dalam bentuk *dashboard* strategis yang mendukung pengambilan keputusan manajerial. Tujuan penelitian ini adalah merancang sistem rekomendasi produk herbal berbasis algoritma *Decision Tree* yang mempertimbangkan karakteristik pelanggan dan riwayat transaksi, serta membandingkan kinerjanya dengan *Random Forest* dan *SVM* guna memperoleh model klasifikasi dengan performa terbaik, menggunakan metodologi *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*). Model terpilih selanjutnya diimplementasikan ke dalam aplikasi berbasis *Streamlit* dengan tiga halaman utama, yaitu Beranda, Visualisasi, dan Rekomendasi, sehingga dapat digunakan langsung oleh manajemen PT. Diamond Interest.

Penelitian ini memberikan tiga kontribusi utama: pertama, menghasilkan model klasifikasi *Decision Tree* untuk sistem rekomendasi produk herbal berdasarkan karakteristik pelanggan dan riwayat transaksinya; kedua, menyajikan perbandingan kinerja antara *Decision Tree*, *Random Forest*, dan *SVM* dalam konteks klasifikasi rekomendasi produk; dan ketiga, menghasilkan prototipe *dashboard* berbasis *Streamlit* yang dapat diadopsi PT. Diamond Interest sebagai alat bantu pengambilan keputusan strategis.

## 2. KAJIAN TEORITIS

Klasifikasi dalam *data mining* merupakan proses menempatkan data ke dalam kelompok tertentu berdasarkan pola yang dipelajari dari data berlabel, sehingga sistem dapat memprediksi kelas dari data baru. Pendekatan ini berbeda dari *clustering*, yang mengelompokkan data tanpa label kelas, sebagaimana diterapkan pada segmentasi pelanggan berbasis *Recency, Frequency, Monetary* (RFM) (Anshary et al., 2022) dan pengelompokan perilaku konsumen berbelanja pada *platform e-commerce* (Nurhidayati et al., 2025). Berbagai algoritma klasifikasi dan rekomendasi telah dikembangkan, mulai dari *collaborative filtering* (Hariri & Rochim, 2022), *content-based filtering* (Nouvalina & Hati, 2024), hingga *K-Nearest* (Gautama & Arijanto, 2024; Sri Puspita Dewi et al., 2022), yang masing-masing memiliki karakteristik berbeda dan tidak selalu cocok pada kasus dengan atribut kategorikal yang beragam seperti karakteristik pelanggan dan riwayat transaksi.

*Decision Tree* membangun model klasifikasi dalam bentuk struktur pohon, di mana setiap *node* merepresentasikan pengujian atribut, cabang merepresentasikan hasil pengujian, dan daun merepresentasikan label kelas, dengan pemisahan pada tiap *node* umumnya ditentukan melalui ukuran *Gini Index* atau *Entropy*. Keunggulan utamanya adalah kemudahan interpretasi hasil, sehingga banyak diterapkan pada prediksi risiko kredit nasabah bank (Marsianus et al., 2025) dan sistem informasi perpustakaan (Yosua Partogi et al., 2021), meskipun rentan mengalami *overfitting* apabila pohon tumbuh terlalu dalam. Sebagai pembanding, *Random Forest* menggabungkan banyak pohon keputusan melalui *majority voting* sehingga lebih stabil namun kurang interpretatif, sedangkan *Support Vector Machine* (SVM) mencari *hyperplane* pemisah optimal antar kelas dan efektif pada data berdimensi tinggi, tetapi memerlukan penyesuaian parameter kernel agar tidak *underfitting*.

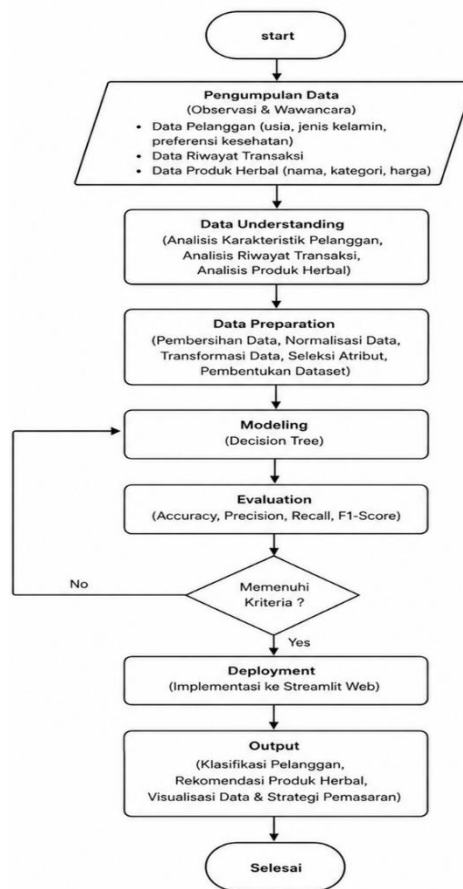
Dua penelitian terdahulu menjadi acuan langsung karena sama-sama menerapkan *Decision Tree* untuk klasifikasi sekaligus rekomendasi. Ameliya et al. (2025) menguji *Decision Tree* untuk mengklasifikasikan *rating* dan merekomendasikan produk pakaian pada *platform e-commerce* menggunakan *dataset* 1.000 data produk *fashion*, dan memperoleh akurasi 81% dengan *precision* tertinggi 0,88 (kelas *rating* 5) dan *recall* tertinggi 0,96 (kelas *rating* 1), di mana rekomendasi dihasilkan melalui *content-based*

*filtering* berbasis *cosine similarity* dan *TF-IDF*. Dimas et al. (2024) menerapkan *Decision Tree* pada domain berbeda, yaitu rekomendasi film dan klasifikasi *rating* Netflix, dengan skema data latih 80% dan data uji 20%, namun hanya mencapai akurasi 39,46%, yang diduga akibat distribusi *rating* tidak seimbang (*rating* 4 mendominasi 34% data) tanpa penyeimbangan kelas. Kedua penelitian ini menghasilkan rekomendasi murni berdasarkan kemiripan atribut produk atau kata kunci, belum mempertimbangkan karakteristik pelanggan dan riwayat transaksi, serta belum membandingkan *Decision Tree* dengan algoritma klasifikasi lain seperti *Random Forest* atau *SVM*.

Berdasarkan kajian teoritis dan tinjauan penelitian terdahulu tersebut, penelitian ini diarahkan pada dugaan bahwa penerapan *Decision Tree* dengan mempertimbangkan karakteristik pelanggan dan riwayat transaksi secara bersamaan, serta dibandingkan dengan *Random Forest* dan *SVM*, akan menghasilkan model klasifikasi preferensi kesehatan pelanggan yang lebih akurat dan aplikatif dibandingkan pendekatan yang hanya mengandalkan kemiripan atribut produk maupun kata kunci semata.

### 3. METODE PENELITIAN

Penelitian ini menggunakan desain penelitian kuantitatif dengan pendekatan eksperimen komparatif, yang bertujuan membandingkan kinerja tiga algoritma klasifikasi, yaitu *Decision Tree*, *Random Forest*, dan *Support Vector Machine* (*SVM*), dalam memprediksi preferensi kesehatan pelanggan berdasarkan karakteristik dan riwayat transaksinya. Pendekatan eksperimen komparatif dipilih karena penelitian ini tidak hanya bertujuan membangun satu model klasifikasi, tetapi juga menguji dan membandingkan performa beberapa algoritma pada dataset yang sama, sehingga dapat diperoleh model dengan tingkat akurasi dan kestabilan terbaik untuk diimplementasikan pada sistem nyata. Tahapan penelitian mengikuti kerangka kerja *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*), yaitu suatu metodologi standar dalam proyek *data mining* yang bersifat iteratif dan terdiri atas enam tahap, meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Pemilihan kerangka kerja ini didasarkan pada sifatnya yang sistematis dan fleksibel, memungkinkan peneliti untuk kembali ke tahap sebelumnya apabila hasil evaluasi belum memenuhi kriteria yang ditetapkan, sebagaimana disajikan pada gambar berikut.



**Gambar 1.** Diagram alir penelitian berbasis kerangka kerja CRISP-DM

(Sumber: dokumentasi penelitian, 2026)

Sebagaimana ditunjukkan pada Gambar 1, penelitian diawali dengan pengumpulan data melalui observasi dan wawancara, mencakup data pelanggan, riwayat transaksi, dan data produk herbal. Data tersebut kemudian dianalisis pada tahap *Data Understanding*, diproses pada tahap *Data Preparation*, dimodelkan menggunakan algoritma *Decision Tree* pada tahap *Modeling*, dan dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Apabila hasil evaluasi belum memenuhi kriteria yang ditetapkan, proses akan kembali ke tahap *Modeling* untuk dilakukan penyesuaian parameter, hingga akhirnya model diimplementasikan pada tahap *Deployment* dalam bentuk aplikasi berbasis web menggunakan *Streamlit*.

Populasi penelitian adalah seluruh data transaksi pelanggan PT. Diamond Interest selama periode 17 Mei 2025 hingga 17 Mei 2026, sedangkan sampel yang digunakan berjumlah 12.635 transaksi dari 4.577 pelanggan unik, mencakup 23 jenis produk herbal

yang tergolong dalam tiga kategori, yaitu Cair, Kapsul, dan Tablet. Teknik pengumpulan data dilakukan melalui pencatatan transaksi penjualan perusahaan, dengan instrumen berupa 14 atribut awal meliputi identitas pelanggan, usia, jenis kelamin, preferensi kesehatan, tanggal transaksi, nomor invoice, nama dan kategori produk, harga satuan, jumlah pembelian, total nilai transaksi, total belanja, serta kategori favorit. Data kemudian diperiksa validitasnya melalui pemeriksaan *missing value* dan duplikasi, yang hasilnya menunjukkan dataset bersih tanpa nilai kosong maupun data ganda.

Sebelas atribut hasil seleksi digunakan sebagai variabel prediktor, sebagaimana disajikan pada Tabel 1, dengan preferensi kesehatan sebagai variabel target yang terdiri atas lima kelas, yaitu Energi & Vitalitas, Imunitas & Stamina, Kulit & Kecantikan, Nutrisi & Gizi, dan Tulang & Sendi.

Tabel 1. Atribut prediktor yang digunakan dalam penelitian  
(Sumber: hasil pengolahan data, 2026)

| No. | Atribut                   | Jenis Data  | Keterangan                          |
|-----|---------------------------|-------------|-------------------------------------|
| 1   | Usia                      | Numerik     | Usia pelanggan                      |
| 2   | Kelompok Usia             | Kategorikal | Hasil pengelompokan usia            |
| 3   | Jenis Kelamin             | Kategorikal | Jenis kelamin pelanggan             |
| 4   | Nama Produk               | Kategorikal | Nama produk herbal yang dibeli      |
| 5   | Kategori Produk           | Kategorikal | Cair, Kapsul, atau Tablet           |
| 6   | Harga Satuan              | Numerik     | Harga per unit produk               |
| 7   | Jumlah Pembelian          | Numerik     | Kuantitas produk yang dibeli        |
| 8   | Total Nilai Transaksi     | Numerik     | Nilai transaksi per pembelian       |
| 9   | Total Belanja             | Numerik     | Akumulasi belanja pelanggan         |
| 10  | Frekuensi Transaksi       | Numerik     | Jumlah transaksi pelanggan          |
| 11  | Rata-rata Nilai Transaksi | Numerik     | Nilai transaksi rata-rata pelanggan |

Data dibagi menjadi data latih sebesar 80% dan data uji sebesar 20% secara *stratified*, kemudian teknik *Synthetic Minority Oversampling Technique* (SMOTE) diterapkan pada data latih untuk mengatasi ketimpangan proporsi antar kelas. Fitur numerik selanjutnya dinormalisasi menggunakan metode *z-score* agar seluruh fitur berada pada skala yang setara, dengan formulasi sebagaimana Persamaan (1):

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

dengan  $x$  adalah nilai aktual fitur,  $\mu$  adalah rata-rata fitur pada data latih, dan  $\sigma$  adalah standar deviasinya.

Ketiga algoritma klasifikasi dikonfigurasi dengan parameter yang telah disesuaikan melalui proses eksperimen, sebagaimana disajikan pada Tabel berikut.

Tabel 2. Konfigurasi hyperparameter model klasifikasi

(Sumber: hasil pengolahan data, 2026)

| Model         | Parameter Utama                                                                  |
|---------------|----------------------------------------------------------------------------------|
| Decision Tree | criterion = entropy; max_depth = 8; min_samples_split = 15; min_samples_leaf = 7 |
| Random Forest | n_estimators = 40; max_depth = 7                                                 |
| SVM           | kernel = radial basis function; C = 5                                            |

Ketiga model dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* berdasarkan matriks konfusi, yang merupakan metrik evaluasi standar untuk klasifikasi multi-kelas, sehingga rincian rumus tidak dituliskan ulang secara rinci pada bagian ini. Performa model juga dibandingkan antara data latih dan data uji untuk mendeteksi kemungkinan *overfitting* maupun *underfitting*, yang diinterpretasikan berdasarkan besarnya selisih akurasi antara kedua kelompok data tersebut. Model dengan performa terbaik selanjutnya disimpan dan diintegrasikan ke dalam aplikasi berbasis *Streamlit* yang terdiri atas tiga halaman utama, yaitu Beranda, Visualisasi, dan Rekomendasi, sehingga hasil klasifikasi dapat langsung dimanfaatkan oleh manajemen PT. Diamond Interest.

#### 4. HASIL DAN PEMBAHASAN

Data penelitian dikumpulkan dari sistem transaksi penjualan PT. Diamond Interest yang salah satu cabangnya berlokasi di Jakarta, mencakup periode 17 Mei 2025 hingga 17 Mei 2026. Proses analisis dilakukan menggunakan lingkungan *Google Colaboratory* dengan bahasa pemrograman Python, memanfaatkan pustaka *pandas*, *scikit-learn*, dan *imbalanced-learn*, sedangkan implementasi sistem menggunakan *framework Streamlit*.



#### 4.1 Karakteristik dan Distribusi Data

Dataset yang digunakan berjumlah 12.635 transaksi dari 4.577 pelanggan unik, dengan 23 jenis produk herbal yang tergolong dalam tiga kategori, yaitu Cair, Kapsul, dan Tablet. Data dibagi menjadi 10.108 data latih dan 2.527 data uji secara *stratified*. Analisis distribusi kelas target pada data latih menunjukkan proporsi yang tidak sepenuhnya seimbang, sebagaimana disajikan pada Tabel berikut.

Tabel 3. Distribusi kelas preferensi kesehatan sebelum dan sesudah SMOTE

(Sumber: hasil pengolahan data, 2026)

| Kelas Preferensi Kesehatan | Jumlah Sebelum SMOTE | Jumlah Sesudah SMOTE |
|----------------------------|----------------------|----------------------|
| Energi & Vitalitas         | 2.317                | 2.317                |
| Imunitas & Stamina         | 2.311                | 2.317                |
| Nutrisi & Gizi             | 1.929                | 2.317                |
| Kulit & Kecantikan         | 1.889                | 2.317                |
| Tulang & Sendi             | 1.662                | 2.317                |

Sebagaimana ditunjukkan pada Tabel 3, kelas Tulang & Sendi memiliki jumlah data paling sedikit sebelum penyeimbangan, sehingga berisiko menyebabkan model bias terhadap kelas mayoritas apabila tidak ditangani. Teknik *Synthetic Minority Oversampling Technique* (SMOTE) diterapkan khusus pada data latih untuk menyeimbangkan kelima kelas menjadi masing-masing 2.317 data, sementara data uji tetap mempertahankan distribusi aslinya agar evaluasi model merepresentasikan kondisi data yang sesungguhnya.

#### 4.2 Perbandingan Performa Model Klasifikasi

Ketiga algoritma, yaitu *Decision Tree*, *Random Forest*, dan SVM, dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* pada data uji, sebagaimana disajikan pada Tabel berikut.

Tabel 4. Perbandingan performa model pada data uji

(Sumber: hasil pengolahan data, 2026)

| Model         | Accuracy | Precision | Recall | F1-Score |
|---------------|----------|-----------|--------|----------|
| Decision Tree | 80,77%   | 80,92%    | 80,77% | 80,58%   |
| Random Forest | 80,29%   | 80,46%    | 80,29% | 80,11%   |
| SVM           | 78,08%   | 79,09%    | 78,08% | 77,94%   |

Berdasarkan Tabel 4, *Decision Tree* memperoleh *F1-score* tertinggi (80,58%), sedikit di atas *Random Forest* (80,11%), sementara *SVM* memperoleh performa terendah (77,94%).

#### a. Deteksi Overfitting dan Underfitting

Selisih akurasi antara data latih dan data uji tercatat kecil pada *Decision Tree* (0,22%) dan *Random Forest* (0,42%), sehingga keduanya dikategorikan berada pada kondisi *good fit* tanpa indikasi *overfitting* maupun *underfitting*. Sebaliknya, *SVM* menunjukkan indikasi *underfitting*, kemungkinan karena kernel *radial basis function* dengan parameter yang digunakan belum sepenuhnya menangkap kompleksitas hubungan antar fitur kategorikal yang telah di-*encode* secara numerik. Hasil ini menunjukkan bahwa kompleksitas tambahan dari pendekatan *ensemble* tidak selalu memberikan keuntungan berarti dibandingkan model pohon tunggal pada dataset dengan skala menengah seperti pada penelitian ini, sejalan dengan prinsip dasar bahwa performa algoritma klasifikasi sangat bergantung pada karakteristik data yang digunakan, bukan semata pada kompleksitas model. Hasil ini turut mengonfirmasi dugaan penelitian pada Bab 2, bahwa penerapan *Decision Tree* dengan mempertimbangkan karakteristik pelanggan dan riwayat transaksi secara bersamaan mampu menghasilkan model klasifikasi yang akurat dan kompetitif dibandingkan algoritma pembandingan.

### 4.3 Evaluasi Kinerja per Kelas

Evaluasi lebih lanjut dilakukan terhadap performa *Decision Tree* sebagai model terbaik pada masing-masing kelas target, sebagaimana disajikan pada Tabel berikut.

Tabel 5. Evaluasi kinerja *Decision Tree* per kelas preferensi kesehatan

(Sumber: hasil pengolahan data, 2026)

| Kelas              | Precision | Recall | F1-Score |
|--------------------|-----------|--------|----------|
| Energi & Vitalitas | 0,87      | 0,83   | 0,85     |
| Imunitas & Stamina | 0,81      | 0,75   | 0,78     |
| Kulit & Kecantikan | 0,77      | 0,95   | 0,85     |
| Nutrisi & Gizi     | 0,78      | 0,69   | 0,73     |
| Tulang & Sendi     | 0,80      | 0,83   | 0,82     |

#### **a. Analisis Kesalahan Klasifikasi**

Pola pada Tabel 5 menunjukkan performa terbaik dicapai pada kelas Kulit & Kecantikan, dengan *recall* mencapai 0,95 meskipun *precision*-nya relatif lebih rendah (0,77), yang mengindikasikan model cenderung terlalu sering mengklasifikasikan data ke kelas ini. Sebaliknya, performa terlemah konsisten muncul pada kelas Nutrisi & Gizi (*F1-score* 0,73), dengan *recall* hanya 0,69, yang menunjukkan model masih kesulitan membedakan kelas ini dari kelas yang berdekatan karakteristiknya, terutama Imunitas & Stamina dan Tulang & Sendi. Temuan ini mengindikasikan adanya tumpang tindih pola pembelian antar kategori preferensi kesehatan pada data penelitian, yang kemungkinan berasal dari karakteristik data itu sendiri, bukan dari pemilihan algoritma.

#### **4.4 Perbandingan dengan Penelitian Sejenis**

Hasil penelitian ini menunjukkan kesesuaian sekaligus perbedaan dengan penelitian terdahulu yang telah diuraikan pada Bab 2. (Ameliya et al., 2025) melaporkan akurasi *Decision Tree* sebesar 81% pada klasifikasi *rating* produk pakaian, relatif sebanding dengan akurasi *Decision Tree* pada penelitian ini (80,77%), yang menunjukkan konsistensi performa algoritma tersebut pada permasalahan klasifikasi multi-kelas dengan jumlah kategori yang relatif sedikit. Sebaliknya, penelitian (Dimas et al., 2024) hanya memperoleh akurasi 39,46% pada klasifikasi *rating* Netflix, jauh di bawah capaian pada penelitian ini, yang diduga disebabkan oleh dominasi satu kelas *rating* tanpa penyeimbangan data sebelum pemodelan. Perbedaan ini menegaskan pentingnya penerapan teknik SMOTE pada penelitian ini untuk mengatasi ketimpangan kelas, sehingga model yang dihasilkan relatif lebih stabil dalam mengenali kelas-kelas dengan jumlah data yang lebih sedikit.

#### **4.5 Implementasi Sistem dan Implikasi Penelitian**

Model *Decision Tree* yang terpilih sebagai model terbaik selanjutnya diintegrasikan ke dalam aplikasi berbasis *Streamlit* dengan tiga halaman utama, yaitu Beranda, Visualisasi, dan Rekomendasi, yang memungkinkan pengguna memasukkan karakteristik pelanggan dan riwayat transaksi untuk memperoleh rekomendasi produk herbal secara langsung.

### **a. Implikasi Teoritis**

Penelitian ini memperkuat pemahaman bahwa performa algoritma *Decision Tree* tidak selalu berada di bawah algoritma *ensemble* seperti *Random Forest*, khususnya pada dataset dengan skala menengah dan jumlah kelas target yang terbatas, sehingga memberikan kontribusi pada literatur mengenai perbandingan algoritma klasifikasi pada domain rekomendasi produk kesehatan herbal yang sebelumnya belum banyak diteliti.

### **b. Implikasi Terapan**

Secara praktis, hasil penelitian ini memberikan manfaat langsung bagi PT. Diamond Interest berupa sistem rekomendasi yang dapat dimanfaatkan untuk menyusun strategi pemasaran yang lebih personal dan sesuai pola pembelian pelanggan, sekaligus menjawab kesenjangan penelitian terkait integrasi karakteristik pelanggan dan riwayat transaksi ke dalam proses rekomendasi produk yang telah diidentifikasi pada Bab 2.

## **5. KESIMPULAN DAN SARAN**

Penelitian ini bertujuan merancang sistem rekomendasi produk kesehatan herbal pada PT. Diamond Interest menggunakan algoritma *Decision Tree*, yang dibandingkan dengan *Random Forest* dan SVM melalui kerangka kerja *CRISP-DM*. Hasil penelitian menunjukkan bahwa *Decision Tree* menjadi model dengan performa terbaik, dengan *accuracy* sebesar 80,77% dan *F1-score* sebesar 80,58% pada data uji, sedikit unggul dibandingkan *Random Forest* dan lebih baik dibandingkan SVM yang terindikasi *underfitting*. Temuan ini menjawab tujuan penelitian sekaligus mengonfirmasi dugaan pada kajian teoritis, bahwa integrasi karakteristik pelanggan dan riwayat transaksi ke dalam klasifikasi *Decision Tree* menghasilkan model yang lebih akurat dan aplikatif dibandingkan pendekatan yang hanya mengandalkan kemiripan atribut produk atau kata kunci pada penelitian terdahulu. Model terbaik tersebut telah diimplementasikan ke dalam aplikasi berbasis *Streamlit* dengan tiga halaman utama, yaitu Beranda, Visualisasi, dan Rekomendasi, sehingga dapat langsung dimanfaatkan manajemen PT. Diamond Interest sebagai alat bantu pengambilan keputusan pemasaran berbasis data.

Meskipun demikian, hasil penelitian ini perlu dimaknai secara hati-hati dan tidak digeneralisasi di luar konteks data yang digunakan, mengingat performa klasifikasi pada kelas Nutrisi & Gizi tercatat paling rendah akibat tumpang tindih pola pembelian antar

kategori, serta cakupan penelitian yang hanya membandingkan tiga algoritma konvensional pada satu periode dan satu perusahaan.

Berdasarkan keterbatasan tersebut, penelitian selanjutnya disarankan menambahkan algoritma pembanding seperti *XGBoost* atau *Gradient Boosting*, memperluas cakupan data pada periode dan segmen pelanggan yang lebih luas, serta mengeksplorasi atribut tambahan yang dapat membedakan kelas dengan karakteristik pembelian serupa. Pengembangan sistem juga dapat dilanjutkan dengan fitur pemantauan performa model secara berkala serta pengujian penerimaan pengguna untuk memastikan kesiapan operasional aplikasi.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Sistem Informasi, Institut Teknologi dan Bisnis Swadharma Jakarta atas dukungan administratif dan teknis selama proses penelitian, serta kepada dosen pembimbing, Yogasetya Suhandi dan Ike Kurniati, yang telah memberikan arahan, masukan, dan bimbingan selama proses penyusunan penelitian ini. Penulis juga menyampaikan terima kasih kepada PT. Diamond Interest atas kesediaannya menjadi objek studi kasus dalam penelitian ini. Artikel ini merupakan bagian dari skripsi penulis pada Program Studi Sistem Informasi, Institut Teknologi dan Bisnis Swadharma Jakarta. Dalam proses penyusunan naskah, penulis menggunakan bantuan perangkat *AI* untuk mendukung perapian bahasa dan struktur penulisan, dengan seluruh isi, analisis, dan tanggung jawab akhir tetap berada pada penulis.

## DAFTAR REFERENSI

- Ameliya, P., Laila, A. F., Saputra, M. A., Syihab, R., & Adam, S. (2025). Penerapan Model Decision Tree untuk Klasifikasi Rating dan Rekomendasi Produk Pakaian pada Platform E-Commerce. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 9(1), 135–143. <https://doi.org/https://doi.org/10.29407/dcwwz815>
- Andriko, T., Juansen, M., Saputra, S. A., & Handayani, S. (2025). DAMPAK PENERAPAN SISTEM REKOMENDASI PRODUK TERHADAP PENINGKATAN MINAT BELI PELANGGAN PADA E-COMMERCE. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4442–4450. <https://doi.org/https://doi.org/10.36040/jati.v9i3.13683>
- Anshary, M. H., Soesanto, O., & Ayatullah. (2022). SEGMENTASI PELANGGAN MENGGUNAKAN METODE K-MEANS CLUSTERING BERDASARKAN MODEL RFM (RECENCY, FREQUENCY, MONETARY). *RAGAM: Journal of Statistics and Its Application*, 1(1), 1–10.

<https://doi.org/https://doi.org/10.20527/ragam.v1i1.7382>

- Bela, Z. I., & Bhakti, H. D. (2022). SISTEM PREDIKSI PENJUALAN OBAT MENGGUNAKAN METODE SINGLE MOVING AVERAGE ( STUDI KASUS : APOTEK WILUJENG KECAMATAN PANCENG KAB . GRESIK ). *INDEXIA: Informatic and Computational Intelligent Journal*, 4(1), 47–58. <https://doi.org/https://doi.org/10.30587/indexia.v4i1.3638>
- Denyka Arinda Putri, Stephanie Ceicillia, Rizky, G. A., & Farida, S. N. (2022). IMPLEMENTASI ANALISIS SWOT (STRENGTH, WEAKNESS, OPPORTUNITIES, AND THREAT) DALAM STRATEGI PEMASARAN PRODUK PADA PT ADIB GLOBAL FOOD SUPPLIES SURABAYA. *Jurnal Bisnis Indonesia*, 13(1), 78–104. <https://doi.org/https://doi.org/10.33005/jbi.v13i1.3079>
- Dimas Aditya Mukhsinin, M Rafliansyah, Adji Ibrahim Sang, Rahmadden, & Wulandari Denok. (2024). Implementasi Algoritma Decision Tree untuk Rekomendasi Film dan Klasifikasi Rating pada Platform Netflix. *Indonesian Journal of Machine Learning and Computer Science*, 4(2), 570–579. <https://doi.org/https://doi.org/10.57152/malcom.v4i2.1255>
- Farid, S. B., Firdaus, D. A., Darmawan, A. G., & Khalid. (2025). Analisis Data dan Visualisasi Interaktif Kecanduan Media Sosial pada Siswa Menggunakan Python dan Streamlit. *JURNAL TEKNOLOGI INFORMASI*, 4(2), 114–124. <https://doi.org/https://doi.org/10.35308/jti.v4i2.12442>
- Gautama, K., & Arijanto, R. (2024). PENERAPAN DATA MINING UNTUK PREDIKSI PENJUALAN PRODUK TERLARIS MENGGUNAKAN METODE K-NEAREST NEIGHBOR PADA CV. NUANSA KARUNA LESTARI. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(6), 11184–11190. <https://doi.org/https://doi.org/10.36040/jati.v8i6.11189>
- Hariri, F. R., & Rochim, L. W. (2022). Sistem Rekomendasi Produk Aplikasi Marketplace Berdasarkan Karakteristik Pembeli Menggunakan Metode User Based Collaborative Filtering. *TEKNIKA Journal of Information and Communication Technology*, 11(3), 208–217. <https://doi.org/10.34148/teknika.v11i3.538>
- Juliansyah, R., & Sutabri, T. (2024). EKSPLORASI E-COMMERCE MENEMUKAN KUNCI SUKSES MENARIK PEMBELI DI INDUSTRI ONLINE. *Scientica: Jurnal Ilmiah Sains Dan Teknologi*, 2(5), 25–29. <https://jurnal.kolibi.org/index.php/scientica/article/view/688>
- Karin Annisa, Budi Serasi Ginting, & Syari Alfhi Mili. (2022). Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada PDAM Langkat. *Jurnal Ilmu Komputer Dan Informatika*, 06(01), 112–129. <https://doi.org/http://dx.doi.org/10.30829/algoritma.v6i1.11624>
- Mahariani, Y. R., Suseno, P., Junianto, D., & Brilliantio, N. N. A. (2026). Analisis Pola Transaksi dan Perilaku Pembelian Pelanggan E-Commerce Berdasarkan Karakteristik Demografis dan Waktu Transaksi. *Jurnal Manajemen Bisnis Era Digital* Volume, 3(1), 54–68. <https://doi.org/https://doi.org/10.61132/jumabedi.v3i1.1279>

- Marsianus Gerlian Eka Mbete, Nathanael Nyala Bintang, & Victor Noviandus. (2025). PENERAPAN ALGORITMA DECISION TREE UNTUK MEMPREDIKSI RISIKO KREDIT PADA NASABAH BANK. *JURNAL TEKNOLOGI INFORMASI*, 16(2), 241–246. <https://doi.org/https://doi.org/10.52972/hoaq.vol16no2>
- Nouvalina, D. A., & Hati, K. (2024). Sistem Rekomendasi Produk Skin Care Berdasarkan Permasalahan Kulit Wajah dengan Metode Content Based Filtering. *JURNAL TEKNIK INFORMATIKA STMIK ANTAR BANGSA*, 10(2), 60–67. <https://doi.org/https://doi.org/10.51998/jti.v10i2.586>
- Nurhidayati, Alwanda, A. Y., & Fathurrahman. (2025). Penerapan Algoritma K-Means untuk Mengelompokkan Perilaku Konsumen dalam Berbelanja (Studi Kasus : Platform Lazada). *Jurnal Pengembangan Rekayasa Informatika Dan Komputer*, 3(2), 199–214. <https://doi.org/https://doi.org/10.29408/jprinter.v3i2.33140>
- Sasono, E., & Ulfatun Nikmah. (2021). KARAKTERISTIK MEREK DAN KARAKTERISTIK PELANGGAN MEREK SERTA PENGARUHNYA TERHADAP LOYALITAS MEREK. *JURNAL STIE SEMARANG*, 13(2), 1–17. <https://doi.org/10.33747>
- Sitopu, J. W. (2025). Model Prediksi Statistik untuk Mengukur Perilaku Konsumen dalam Transaksi Digital. *Journal of Economic Studies*, 1(2), 63–73. <https://ejournal.risetanakbangsa.id/jeta/article/view/42>
- Sri Puspita Dewi, Nurwati, & Rahayu, E. (2022). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, 3(4), 639–648. <https://doi.org/10.47065/bits.v3i4.1408>
- Utami, H. N., Putri Gita Pratiwi, & Wiyono, ulistyodewi N. (2025). KARAKTERISTIK KONSUMEN DAN POLA PERILAKU BERBELANJA PANGAN PREMIUM: STUDI KASUS PADA PEMASARAN PRODUK BERLABEL SEHAT. *Agricore: Jurnal Agribisnis Dan Sosial Ekonomi Pertanian Unpad*, 10(1), 319–332. <https://doi.org/https://doi.org/10.24198/agricore.v10i1.63512>
- Yosua Partogi, Pasaribu, A., & Sutrisno. (2021). PERANCANGAN METODE DECISION TREE TERHADAP SISTEM PERPUSTAKAAN STMIK KUWERA. *JURNAL SISTEM INFORMASI DAN TEKNOLOGI ( S I N T E K )*, 1(2), 20–26. <https://doi.org/https://doi.org/10.56995/sintek.v1i2.4>