



Klasifikasi Tingkat Kepuasan Pengguna Produk Body Care Menggunakan Algoritma Decision Tree

Nur Jannah Hasibuan*, Aidil Halim Lubis

Fakultas Sains dan Teknologi, Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: ^{1,*}nurj35618@gmail.com, ²aidilhalimlubis@uinsu.ac.id

Email Penulis Korespondensi: nurj35618@gmail.com

Abstrak—Persaingan industri perawatan tubuh (*body care*) yang semakin meningkat mendorong perusahaan untuk memahami tingkat kepuasan pelanggan sebagai dasar peningkatan kualitas produk dan layanan. Namun, proses analisis kepuasan pengguna sering menghasilkan data yang kompleks sehingga sulit dianalisis secara manual. Penelitian ini bertujuan menerapkan algoritma *Decision Tree* untuk mengklasifikasikan tingkat kepuasan pengguna produk *body care* berdasarkan karakteristik pengguna dan penilaian terhadap produk. Data penelitian diperoleh melalui penyebaran kuesioner kepada 250 responden dengan atribut jenis kelamin, usia, frekuensi penggunaan, kualitas produk, harga, pelayanan, dan tingkat kepuasan sebagai variabel target. Tahapan penelitian meliputi preprocessing data, seleksi atribut, transformasi data, pembagian data latih dan data uji, serta pembangunan model klasifikasi menggunakan algoritma *Decision Tree*. Evaluasi model dilakukan menggunakan confusion matrix dengan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model mampu mengelompokkan tingkat kepuasan pengguna ke dalam kategori sangat puas, puas, cukup puas, dan tidak puas dengan nilai akurasi sebesar 58%, precision 57%, recall 57%, dan F1-score 57%. Penelitian ini memberikan kontribusi dalam penerapan data mining untuk analisis kepuasan pelanggan pada industri *body care* serta membantu perusahaan mengidentifikasi faktor dominan yang memengaruhi kepuasan pengguna, khususnya kualitas produk dan pelayanan. Selain itu, hasil penelitian diharapkan dapat menjadi referensi dalam pengembangan sistem analisis kepuasan pelanggan berbasis data mining pada industri kecantikan dan perawatan tubuh.

Kata Kunci: Decision Tree; Data Mining; Kepuasan Pengguna; Klasifikasi; Body Care

Abstract—The increasing competition in the body care industry encourages companies to understand customer satisfaction as a basis for improving product quality and service performance. However, analyzing user satisfaction often produces complex data that are difficult to process manually. This study aims to apply the Decision Tree algorithm to classify the satisfaction levels of body care product users based on user characteristics and product evaluations. The research data were collected through questionnaires distributed to 250 respondents, including attributes such as gender, age, frequency of use, product quality, price, service quality, and satisfaction level as the target variable. The research stages consisted of data preprocessing, attribute selection, data transformation, splitting data into training and testing datasets, and building a classification model using the Decision Tree algorithm. Model evaluation was carried out using a confusion matrix with accuracy, precision, recall, and F1-score metrics. The results showed that the model was able to classify user satisfaction into four categories: very satisfied, satisfied, fairly satisfied, and dissatisfied, with an accuracy of 58%, precision of 57%, recall of 57%, and F1-score of 57%. This study contributes to the implementation of data mining for customer satisfaction analysis in the body care industry and helps companies identify dominant factors influencing user satisfaction, particularly product quality and service quality. In addition, the findings are expected to serve as a reference for developing customer satisfaction analysis systems based on data mining in the beauty and body care industry.

Keywords: Decision Tree; Data Mining; User Satisfaction; Classification; Body Care

1. PENDAHULUAN

Industri produk perawatan tubuh (*body care*) mengalami perkembangan pesat dalam beberapa tahun terakhir. Pertumbuhan ini tidak lepas dari meningkatnya kesadaran masyarakat akan pentingnya menjaga kesehatan dan penampilan kulit. Perawatan tubuh kini tidak hanya sekadar rutinitas kebersihan, tetapi menjadi bagian dari gaya hidup, terutama di kalangan generasi muda. Perubahan paradigma ini turut mendorong meningkatnya permintaan terhadap berbagai produk dengan manfaat spesifik sesuai kebutuhan konsumen [1]. Fenomena tersebut juga dipengaruhi oleh maraknya informasi tentang perawatan kulit yang mudah diakses melalui media sosial, blog kecantikan, dan platform berbagi video. Edukasi yang meluas ini membuat konsumen semakin selektif dalam memilih produk, sehingga produsen dituntut untuk menyediakan inovasi dan kualitas yang sesuai ekspektasi. Akibatnya, persaingan di industri *body care* semakin ketat dan menuntut strategi pemasaran yang berbasis data [2]. Salah satu bahan aktif yang belakangan mengalami lonjakan popularitas adalah Alpha Arbutin. Zat ini dikenal karena kemampuannya dalam membantu mencerahkan kulit dan mengurangi hiperpigmentasi. Alpha Arbutin bekerja dengan menghambat pembentukan melanin di kulit, sehingga dapat membantu menyamarkan noda hitam dan meratakan warna kulit. Tingginya minat konsumen terhadap Alpha Arbutin memunculkan tantangan baru bagi perusahaan yang memproduksi atau mendistribusikan produk tersebut. Persaingan tidak hanya berasal dari merek-merek besar, tetapi juga dari produsen lokal yang menawarkan harga kompetitif. Dalam situasi ini, memahami kepuasan pengguna menjadi sangat penting agar perusahaan dapat mempertahankan pelanggan dan memenangkan persaingan pasar.

Kepuasan pengguna sendiri merupakan indikator penting yang menggambarkan sejauh mana produk memenuhi atau melebihi harapan konsumen. Pelanggan yang puas akan lebih cenderung melakukan pembelian ulang, memberikan ulasan positif, dan merekomendasikan produk kepada orang lain. Sebaliknya, pelanggan yang tidak puas berpotensi beralih ke produk pesaing dan memberikan penilaian negatif yang dapat merugikan citra merek [3]. Meski penting, banyak perusahaan masih menghadapi kesulitan dalam mengukur tingkat kepuasan pengguna secara akurat dan berkelanjutan. Metode tradisional seperti survei manual atau wawancara seringkali memakan waktu, membutuhkan biaya



besar, dan menghasilkan data yang terbatas. Selain itu, data yang terkumpul seringkali bersifat subjektif, sehingga sulit digunakan sebagai dasar pengambilan keputusan strategis.

Perkembangan teknologi informasi, khususnya di bidang analisis data, memberikan peluang baru untuk mengatasi permasalahan tersebut [4]. Teknik data mining memungkinkan perusahaan menganalisis data dalam jumlah besar untuk menemukan pola dan hubungan yang sebelumnya tidak terlihat [5]. Data mining juga memungkinkan analisis yang lebih objektif dan sistematis dibandingkan metode manual [6].

Salah satu pendekatan dalam data mining yang relevan untuk kasus ini adalah klasifikasi, yaitu proses mengelompokkan data ke dalam kategori tertentu berdasarkan atribut-atribut. Dalam konteks penelitian ini, klasifikasi digunakan untuk membedakan pengguna berdasarkan tingkat kepuasan mereka terhadap produk Alpha Arbutin. Hasil klasifikasi dapat memberikan wawasan mendalam yang membantu perusahaan merancang strategi pemasaran dan perbaikan produk. Dari berbagai algoritma klasifikasi yang tersedia, *Decision Tree* dipilih karena kemampuannya membangun model prediksi yang mudah dipahami [5]. Algoritma ini menyajikan hasil analisis dalam bentuk pohon keputusan, di mana setiap percabangan mewakili pemisahan data berdasarkan atribut tertentu. Struktur yang jelas ini memudahkan pihak non-teknis, seperti tim pemasaran atau manajemen, untuk memahami dan memanfaatkan hasil analisis. Selain kemudahan interpretasi, *Decision Tree* juga memiliki kemampuan untuk mengidentifikasi atribut yang paling berpengaruh terhadap kepuasan pengguna [7]. Dengan mengetahui faktor dominan tersebut, perusahaan dapat memprioritaskan perbaikan dan inovasi produk pada aspek yang benar-benar penting bagi konsumen. Pendekatan ini membantu mengoptimalkan sumber daya dan meningkatkan efektivitas strategi bisnis.

Penerapan *Decision Tree* dalam klasifikasi kepuasan pengguna juga mendukung proses segmentasi pasar. Perusahaan dapat membedakan kelompok pengguna yang sangat puas, cukup puas, atau tidak puas, lalu menyesuaikan strategi komunikasi dan promosi untuk masing-masing segmen. Misalnya, konsumen yang puas dapat dijadikan target program loyalitas, sementara konsumen yang kurang puas dapat diberikan penawaran khusus atau edukasi penggunaan produk yang lebih tepat.

Dari sisi akademis, penelitian ini memberikan kontribusi terhadap pengembangan penerapan teknik data mining, khususnya algoritma *Decision Tree*, dalam analisis kepuasan pelanggan pada industri kosmetik dan body care. Beberapa penelitian terdahulu menunjukkan bahwa algoritma *Decision Tree* mampu memberikan hasil klasifikasi yang baik pada berbagai bidang. Penelitian oleh Qisthiano et, al. (2023) menerapkan *Decision Tree* untuk klasifikasi prediksi kelulusan mahasiswa [8]. Penelitian Musa et, al. (2024) menggunakannya untuk klasifikasi penjualan pakan ternak ternak ternak [9]. Sedangkan penelitian Bulkisah et, al. (2024) menerapkan metode tersebut pada klasifikasi status gizi balita [10]. Hasil dari penelitian-penelitian tersebut menunjukkan bahwa *Decision Tree* memiliki kemampuan yang baik dalam mengidentifikasi pola data dan menghasilkan model klasifikasi yang mudah dipahami. Namun, penerapan algoritma *Decision Tree* pada analisis tingkat kepuasan konsumen, khususnya pada industri *body care* dan produk berbahan aktif Alpha Arbutin, masih relatif terbatas. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan penelitian (*research gap*) tersebut dengan menerapkan algoritma *Decision Tree* dalam mengklasifikasikan tingkat kepuasan pengguna produk *body care* berdasarkan atribut kualitas produk, harga, pelayanan, dan karakteristik pengguna. Perbedaan penelitian ini dengan studi terdahulu terletak pada objek dan tujuan analisis, di mana fokus diarahkan pada kepuasan pengguna terhadap produk Alpha Arbutin sebagai bahan aktif populer. Penelitian ini diharapkan dapat memberikan wawasan baru yang relevan bagi perusahaan kosmetik dalam mengambil keputusan strategis berbasis data.

Berdasarkan uraian di atas, penelitian ini difokuskan pada penerapan metode data mining dengan algoritma *Decision Tree* untuk mengklasifikasikan kepuasan pengguna terhadap produk Alpha Arbutin. Tujuannya adalah memberikan gambaran yang jelas tentang faktor-faktor yang memengaruhi kepuasan pengguna dan menyediakan dasar pengambilan keputusan berbasis data. Pendekatan ini diharapkan membantu perusahaan lebih adaptif terhadap perubahan perilaku konsumen dan dinamika pasar yang terus berkembang.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan data mining dengan algoritma *Decision Tree* untuk mengklasifikasikan tingkat kepuasan pengguna terhadap produk *Body care*. Tahapan penelitian meliputi identifikasi masalah, pengumpulan data, pra-pemrosesan data, implementasi algoritma, evaluasi model, serta analisis hasil.

2.1 Tahapan Penelitian

Adapun tahapan penelitian yg digunakan pada penelitian ini ialah:

a. Identifikasi Masalah

Tahap awal penelitian dilakukan dengan mengidentifikasi permasalahan terkait belum adanya sistem yang mampu mengolah data pengguna *Body care* secara otomatis untuk mengukur dan mengklasifikasikan tingkat kepuasan. Identifikasi dilakukan melalui studi literatur, pengamatan terhadap tren penggunaan produk, serta diskusi dengan pengguna dan pelaku usaha terkait [11].

b. Pengumpulan Data

Pengumpulan data merupakan tahap yang sangat penting dalam penelitian ini karena kualitas dan kelengkapan data akan sangat mempengaruhi akurasi model klasifikasi yang dibangun. Data diperoleh dari dua sumber utama, yaitu survei

kuesioner langsung kepada pengguna Alpha Arbutin dan data internal dari penjual resmi. Survei kuesioner dilakukan dengan tujuan untuk mendapatkan informasi langsung dari pengguna terkait profil mereka, pola penggunaan produk, persepsi terhadap hasil yang diperoleh, serta tingkat kepuasan secara keseluruhan. Penyusunan kuesioner dilakukan dengan hati-hati agar pertanyaan yang diajukan relevan dengan tujuan penelitian dan mudah dipahami oleh responden. Pertanyaan menggunakan kombinasi *format close-ended* (jawaban pilihan) untuk memudahkan pengkodean data, dan *rating scale* untuk mengukur tingkat kepuasan atau persepsi responden. Sumber data kedua berasal dari penjual resmi Alpha Arbutin, yang memiliki basis data ulasan pelanggan serta penilaian produk. Data ini memberikan sudut pandang tambahan yang bersifat objektif, karena diambil dari catatan transaksi dan review yang diberikan oleh konsumen. Integrasi kedua sumber data ini diharapkan mampu memberikan gambaran yang lebih lengkap terkait karakteristik pengguna dan tingkat kepuasan mereka.

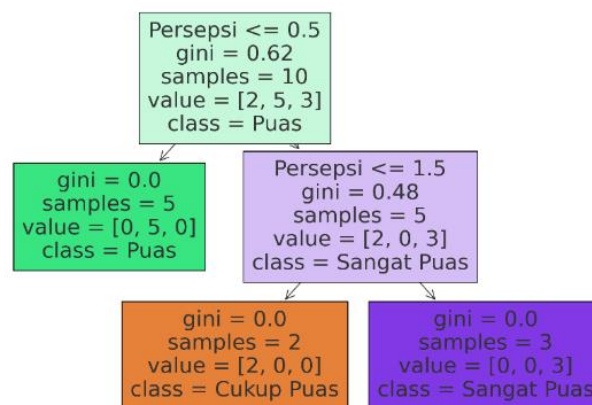
Atribut yang dikumpulkan dalam penelitian ini meliputi:

Tabel 1. Bentuk Data Hasil Survei Responden

ID Responden	Usia (Tahun)	Jenis Kelamin	Frekuensi Penggunaan	Persepsi Hasil Produk	Tingkat Kepuasan
R001	25	Perempuan	4 kali/minggu	Kulit lebih cerah dan lembut	5
R002	31	Laki-laki	2 kali/minggu	Perubahan kurang terlihat	3
R003	22	Perempuan	6 kali/minggu	Hasil signifikan, kulit cerah	5

c. Pra-Pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam proses klasifikasi [12]. Pada Tahap implementasi algoritma *Decision Tree* merupakan inti dari proses penelitian ini, di mana data yang telah melalui tahap pra-pemrosesan digunakan untuk membangun sebuah model klasifikasi yang mampu memprediksi tingkat kepuasan pengguna Alpha Arbutin. Algoritma yang digunakan adalah C4.5 atau CART (*Classification and Regression Tree*), dua metode populer yang terbukti efektif dalam mengolah data kategorikal maupun numerik. Pemilihan algoritma ini didasarkan pada fleksibilitasnya dalam menangani data campuran (*mixed data type*) serta kemampuannya menghasilkan model yang mudah dipahami oleh pengguna non-teknis [13]. Berikut ditampilkan Diagram Proses Klasifikasi pada Gambar 1.



Gambar 1. Diagram *Decision Tree*, proses klasifikasi

Proses pembentukan pohon keputusan dimulai dengan menghitung nilai pemisahan terbaik untuk setiap atribut yang tersedia. Pada metode C4.5, ukuran pemisahan diukur dengan Information Gain Ratio, yang mempertimbangkan tingkat entropi dan pembagian data yang lebih seimbang. Sementara pada metode CART, digunakan ukuran Gini Index untuk menilai seberapa murni (*pure*) suatu pembagian data berdasarkan atribut tertentu. Tujuan dari langkah ini adalah memilih atribut yang paling informatif untuk memisahkan data ke dalam kategori target [14]. Setelah atribut terbaik ditemukan, atribut tersebut dijadikan node akar (*root node*) dari pohon keputusan. Dataset kemudian dibagi menjadi beberapa subset sesuai dengan nilai atribut tersebut. Misalnya, jika atribut yang dipilih adalah "Frekuensi Penggunaan", maka data akan dikelompokkan menjadi subset "Harian", "3x Seminggu", dan "Kadang-Kadang". Proses ini diulangi secara rekursif pada setiap subset yang dihasilkan, membentuk node internal baru hingga salah satu dari kondisi berikut tercapai:

- Seluruh data dalam subset memiliki label kelas yang sama (entropi = 0).
- Tidak ada atribut tersisa untuk membedakan data.
- Kedalaman pohon telah mencapai batas yang ditentukan untuk mencegah *overfitting*.

Selanjutnya, dilakukan proses *pruning* atau pemangkasan pohon untuk menghilangkan cabang-cabang yang tidak memberikan kontribusi signifikan terhadap akurasi prediksi. *Pruning* ini penting agar model tidak terlalu kompleks dan dapat bekerja lebih baik pada data uji (*testing set*) [15].

Dalam konteks penelitian ini, atribut yang digunakan meliputi usia pengguna, jenis kelamin, frekuensi penggunaan, persepsi terhadap hasil produk, serta tingkat kepuasan sebagai variabel target [16]. Hasil dari tahap ini adalah sebuah



pohon keputusan yang dapat memetakan kombinasi atribut menjadi prediksi tingkat kepuasan, misalnya “Puas”, “Cukup Puas”, atau “Tidak Puas”. Model ini nantinya akan diimplementasikan ke dalam aplikasi berbasis Laravel, sehingga sistem dapat memberikan hasil klasifikasi secara otomatis berdasarkan data masukan dari pengguna atau admin.

2.2 Algoritma Decision Tree

Pada tahap ini, data pelatihan digunakan untuk membangun model klasifikasi menggunakan algoritma *Decision Tree*. Proses pembentukan pohon keputusan dilakukan dengan memilih atribut terbaik sebagai node akar (*root node*) berdasarkan ukuran pemisahan seperti *Information Gain* atau *Gini Index* [17]. Proses ini dilakukan secara rekursif hingga terbentuk struktur pohon yang dapat mengklasifikasikan tingkat kepuasan pengguna ke dalam beberapa kategori [18].

2.3 Evaluasi Model

Model yang telah dibangun kemudian diuji menggunakan data testing untuk mengetahui kinerjanya. Evaluasi dilakukan menggunakan *Confusion Matrix* dan beberapa metrik pengukuran, yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-score* [19]. Metrik tersebut digunakan untuk menilai tingkat ketepatan dan kemampuan model dalam mengklasifikasikan data secara benar [20].

2.4 Analisis Hasil

Tahap akhir penelitian adalah melakukan analisis terhadap hasil evaluasi model. Analisis dilakukan dengan menginterpretasikan nilai metrik evaluasi serta mengidentifikasi atribut yang paling berpengaruh dalam proses klasifikasi [21]. Hasil analisis ini diharapkan dapat memberikan informasi mengenai faktor-faktor yang memengaruhi tingkat kepuasan pengguna terhadap produk *Body care* serta menjadi dasar pengambilan keputusan yang lebih tepat.

3. HASIL DAN PEMBAHASAN

3.1 Representasi Data

Melalui penyajian hasil penelitian ini, diharapkan pembaca dapat memahami alur penelitian secara utuh serta memperoleh gambaran mengenai kinerja algoritma *Decision Tree* dalam mengklasifikasikan data survei kepuasan pengguna produk *Body care*. Berikut pada Tabel 2 merupakan data yg digunakan pada penelitian.

Tabel 2. Data Penelitian

No	Jenis Kelamin	Usia	Frekuensi Pembelian	Kualitas Produk	Harga	Pelayanan	Kepuasan Pengguna
1	Perempuan	21	Tinggi	Baik	Terjangkau	Baik	Puas
2	Laki-laki	23	Sedang	Cukup	Terjangkau	Cukup	Cukup Puas
3	Perempuan	22	Tinggi	Sangat Baik	Terjangkau	Sangat Baik	Sangat Puas
4	Laki-laki	24	Rendah	Cukup	Mahal	Cukup	Tidak Puas
5	Perempuan	21	Sedang	Baik	Terjangkau	Baik	Puas
...	...	...	...	...	...	...	...
250	Laki-laki	26	Sedang	Cukup	Mahal	Cukup	Cukup Puas

Tabel 1 menunjukkan data sampel yang digunakan dalam penelitian ini. Data sampel tersebut diambil dari keseluruhan 250 data responden. Setiap baris data merepresentasikan satu responden dengan atribut penilaian terhadap produk *Body care*, serta satu atribut kelas yaitu Kepuasan Pengguna yang digunakan sebagai target klasifikasi. Penyajian data ini bertujuan untuk memberikan gambaran struktur dan karakteristik data yang digunakan sebelum dilakukan proses perhitungan dan evaluasi model *Decision Tree*.

3.2 Analisis Data

Analisis data merupakan tahap awal penelitian yang bertujuan untuk memahami karakteristik data survei kepuasan pengguna produk *Body care*. Pada tahap ini, data masih berupa data mentah (*raw data*) sehingga belum siap digunakan secara langsung dalam proses klasifikasi. Data survei terdiri dari beberapa atribut yang menggambarkan tanggapan responden terhadap produk *Body care*. Oleh karena itu, dilakukan analisis awal untuk mengidentifikasi struktur data, jenis atribut, serta memeriksa kemungkinan adanya data kosong (*missing value*) atau ketidakkonsistenan data. Hasil analisis menunjukkan bahwa tidak semua atribut memiliki pengaruh langsung terhadap proses klasifikasi tingkat kepuasan pengguna. Beberapa atribut yang kurang relevan dapat dihilangkan, sedangkan atribut yang berkaitan dengan pengalaman dan persepsi pengguna terhadap produk tetap digunakan sebagai variabel utama dalam penelitian. Dengan dilakukannya analisis data, peneliti memperoleh gambaran awal mengenai karakteristik data yang digunakan. Tahap ini menjadi dasar dalam menentukan proses *preprocessing* sehingga proses klasifikasi menggunakan algoritma *Decision Tree* dapat dilakukan secara lebih optimal. Berikut pada Gambar 1 menunjukkan dataset awal survei kepuasan pengguna produk *Body care* yang diperoleh dari responden. Data tersebut masih dalam kondisi mentah dan belum melalui proses *preprocessing*, sehingga masih memuat berbagai atribut dengan format yang beragam.



```
df = pd.read_csv("dataset_bodycare_250_email_nama.csv")
df.head()
```

	Email	Jenis Kelamin	Usia	Frekuensi	Jenis Produk	Merek Produk	Kualitas	Aroma	Tekstur	Ef
0	andi862@gmail.com	Perempuan	18-25 tahun	3-5 kali seminggu	Body scrub	Dove	Sangat Baik	Tidak Suka	Sangat Nyaman	Ci
1	ayu245@gmail.com	Laki-laki	18-25 tahun	Setiap hari	Body lotion	Vaseline	Baik	Sangat Suka	Nyaman	
2	nanda71@gmail.com	Perempuan	<18 tahun	Jarang	Body scrub	Vaseline	Kurang	Sangat Suka	Nyaman	
3	nurjanah795@gmail.com	Perempuan	18-25 tahun	1-2 kali seminggu	Body lotion	Scarlett Whitening	Baik	Tidak Suka	Sangat Nyaman	
4	aisyah126@gmail.com	Laki-laki	26-35 tahun	3-5 kali seminggu	Body scrub	Alpha Arbutin	Kurang	Netral	Nyaman	

Gambar 2. Dataset Awal Survei Kepuasan Pengguna Produk *Body care*

3.3 Distribusi Data Kelas Kepuasan

Distribusi tersebut menunjukkan bahwa data kepuasan konsumen tersebar ke dalam empat kelas, sehingga metode klasifikasi yang digunakan harus mampu menangani permasalahan multikelas. Berikut pada Tabel 2 disajikan Data Distribusi Kelas Kepuasan Konsumen.

Tabel 2. Distribusi Kelas Kepuasan Konsumen

Tingkat Kepuasan	Jumlah Data
Sangat Puas	72
Puas	69
Cukup Puas	61
Tidak Puas	48
Total	250

3.4 Deskripsi Data Survei Penggunaan Produk *Body care*

Deskripsi data bertujuan untuk memberikan gambaran umum mengenai data survei penggunaan produk *Body care* yang digunakan dalam penelitian ini. Pada tahap ini, data dijelaskan secara deskriptif untuk menunjukkan karakteristik responden dan atribut yang dianalisis sebelum dilakukan proses pengolahan data lebih lanjut [15].

Data survei diperoleh dari responden yang merupakan pengguna produk *Body care*. Setiap responden memberikan jawaban terhadap sejumlah pertanyaan yang dirancang untuk menggambarkan pola penggunaan serta tingkat kepuasan terhadap produk *Body care* yang digunakan. Data yang terkumpul selanjutnya direkapitulasi dan disusun dalam bentuk dataset untuk dianalisis menggunakan pendekatan data mining [16].

Atribut-atribut yang terdapat dalam dataset mencerminkan berbagai aspek penggunaan produk *Body care*, seperti frekuensi penggunaan, persepsi terhadap kualitas produk, kenyamanan penggunaan, serta tingkat kepuasan secara keseluruhan. Setiap atribut memiliki peran penting dalam menggambarkan pengalaman responden dan menjadi dasar dalam proses klasifikasi.

Pada tahap deskripsi data, dilakukan pengamatan terhadap jumlah data, jenis atribut, serta distribusi nilai pada masing-masing atribut. Pengamatan ini bertujuan untuk memahami variasi jawaban responden dan mengidentifikasi pola umum yang muncul dalam data survei. Selain itu, tahap ini juga membantu peneliti dalam mengenali potensi permasalahan data, seperti adanya nilai yang dominan atau variasi data yang terlalu kecil.

Hasil deskripsi data menunjukkan bahwa data survei memiliki karakteristik yang beragam, mencerminkan perbedaan pengalaman dan persepsi responden terhadap penggunaan produk *Body care*. Variasi data ini menjadi salah satu alasan digunakannya metode klasifikasi, karena memungkinkan pengelompokan responden berdasarkan pola penggunaan dan tingkat kepuasan yang berbeda.

Dengan adanya deskripsi data yang jelas, pembaca dapat memperoleh pemahaman awal mengenai kondisi dataset yang digunakan dalam penelitian. Tahap ini juga menjadi penghubung antara analisis data awal dan proses *preprocessing* data, sehingga alur penelitian dapat dipahami secara runtut dan sistematis.

3.5 Preprocessing Data

Preprocessing data merupakan tahap penting dalam penelitian berbasis data mining yang bertujuan meningkatkan kualitas data sebelum digunakan dalam proses klasifikasi menggunakan algoritma *Decision Tree*. Data survei yang diperoleh dari responden masih berada dalam kondisi data mentah (*raw data*) sehingga berpotensi mengandung atribut yang tidak relevan, perbedaan format data, maupun ketidakkonsistenan nilai [17].

Jika data mentah langsung digunakan dalam proses analisis, maka hasil klasifikasi dapat menjadi kurang optimal. Oleh karena itu, dilakukan proses *preprocessing* untuk memastikan bahwa data yang digunakan benar-benar merepresentasikan kondisi kepuasan pengguna produk *Body care* secara akurat.

Dalam penelitian ini, tahapan *preprocessing* meliputi seleksi atribut untuk menentukan variabel yang relevan dengan penelitian serta transformasi data untuk menyesuaikan format data agar dapat diproses oleh algoritma klasifikasi. Tahapan ini menjadi dasar penting sebelum proses pelatihan model dan klasifikasi dilakukan.

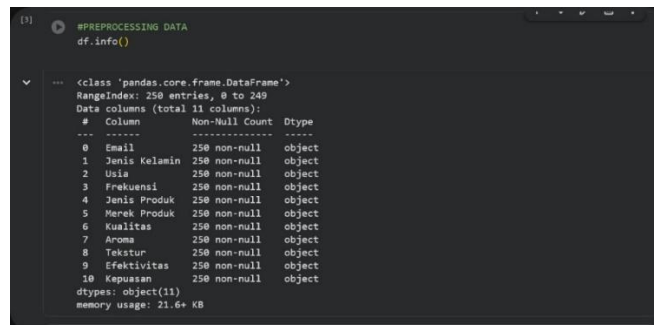
3.6 Seleksi Atribut

Pada Tabel 3 merupakan Seleksi atribut dilakukan untuk menentukan variabel yang paling relevan dalam merepresentasikan penggunaan dan kepuasan pengguna produk *Body care*. Tidak semua atribut dalam data survei memiliki pengaruh langsung terhadap proses klasifikasi. Atribut yang bersifat administratif atau tidak berkaitan dengan tingkat kepuasan pengguna dihilangkan agar tidak memengaruhi kinerja model. Dalam algoritma *Decision Tree*, pemilihan atribut yang relevan sangat penting karena setiap percabangan pohon keputusan dibentuk berdasarkan atribut yang memiliki pengaruh terhadap hasil klasifikasi. Oleh karena itu, seleksi atribut dilakukan untuk mengurangi kompleksitas model serta meningkatkan akurasi dan kemudahan interpretasi. Setelah proses seleksi dilakukan, diperoleh 10 atribut utama yang digunakan dalam penelitian, yaitu jenis kelamin, usia, frekuensi penggunaan, jenis produk, merek produk, kualitas, aroma, tekstur, efektivitas, dan kepuasan. Atribut kepuasan digunakan sebagai variabel target (*class label*), sedangkan atribut lainnya berfungsi sebagai variabel prediktor dalam proses klasifikasi.

Tabel 3. Atribut yang Digunakan Setelah Seleksi Atribut

No	Atribut	Keterangan
1	Jenis Kelamin	Menunjukkan jenis kelamin responden pengguna produk <i>Body care</i>
2	Usia	Kategori usia responden yang menggunakan produk <i>Body care</i>
3	Frekuensi	Frekuensi penggunaan atau pembelian produk <i>Body care</i>
4	Jenis Produk	Jenis produk <i>Body care</i> yang digunakan responden
5	Merek Produk	Merek produk <i>Body care</i> yang digunakan oleh responden
...	...	...
10	Kepuasan	Tingkat kepuasan pengguna terhadap produk <i>Body care</i>

Tabel 3. menunjukkan atribut-atribut yang digunakan setelah proses seleksi atribut. Atribut tersebut dipilih karena memiliki keterkaitan langsung dengan pengalaman penggunaan produk serta tingkat kepuasan responden.



```

[1] #PREPROCESSING DATA
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 250 entries, 0 to 249
Data columns (total 11 columns):
 #   Column             Non-Null Count  Dtype  
---  --
 0   Email              250 non-null   object 
 1   Jenis Kelamin      250 non-null   object 
 2   Usia               250 non-null   object 
 3   Frekuensi          250 non-null   object 
 4   Jenis Produk       250 non-null   object 
 5   Merek Produk       250 non-null   object 
 6   Kualitas           250 non-null   object 
 7   Aroma              250 non-null   object 
 8   Tekstur            250 non-null   object 
 9   Efektivitas        250 non-null   object 
10  Kepuasan           250 non-null   object 
dtypes: object(11)
memory usage: 21.6+ KB

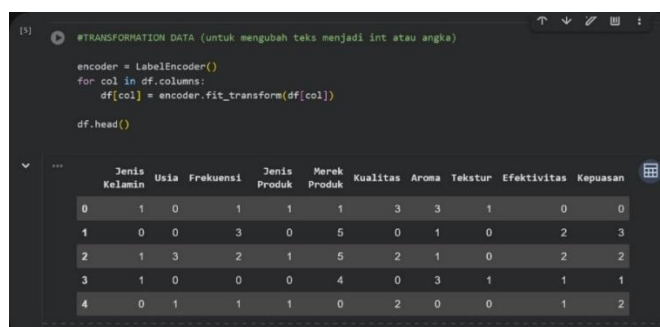
```

Gambar 3. Proses *Preprocessing* Data Survei Penggunaan Produk *Body care*

Gambar 3. menampilkan proses *preprocessing* yang dilakukan pada data survei penggunaan produk *Body care*. Tahap ini bertujuan untuk menyiapkan data agar lebih terstruktur sebelum digunakan dalam proses klasifikasi.

3.7 Transformasi Data

Transformasi data merupakan tahap lanjutan dalam *preprocessing* yang bertujuan untuk mengubah format data agar sesuai dengan kebutuhan algoritma *Decision Tree*. Data survei umumnya memiliki variasi jawaban yang luas, terutama pada atribut kategorikal seperti usia, frekuensi penggunaan, dan tingkat kepuasan. Oleh karena itu, dilakukan proses transformasi untuk menyederhanakan representasi data tanpa menghilangkan makna informasi yang terkandung di dalamnya. Pada tahap ini, data yang semula berbentuk kategori diubah menjadi kode numerik agar dapat diproses oleh algoritma klasifikasi. Transformasi ini penting karena *Decision Tree* melakukan proses pemisahan data berdasarkan perbandingan nilai atribut. Dengan format data yang lebih terstandar, algoritma dapat mengenali pola hubungan antar atribut dengan lebih jelas.



```

[5] #TRANSFORMATION DATA (untuk mengubah teks menjadi int atau angka)

encoder = LabelEncoder()
for col in df.columns:
    df[col] = encoder.fit_transform(df[col])

df.head()

```

	Jenis Kelamin	Usia	Frekuensi	Jenis Produk	Merek Produk	Kualitas	Aroma	Tekstur	Efektivitas	Kepuasan
0	1	0	1	1	1	3	3	1	0	0
1	0	0	3	0	5	0	1	0	2	3
2	1	3	2	1	5	2	1	0	2	2
3	1	0	0	0	4	0	3	1	1	1
4	0	1	1	1	0	2	0	0	1	2

Gambar 1. Hasil Transformasi Data Survei Penggunaan Produk *Body care*



Gambar 4 menunjukkan dataset yang telah melalui proses transformasi sehingga memiliki format yang lebih seragam dan siap digunakan dalam proses klasifikasi.

Beberapa atribut yang ditransformasikan antara lain jenis kelamin, usia, frekuensi penggunaan, dan tingkat kepuasan. Misalnya, atribut jenis kelamin diubah menjadi kode numerik yaitu *Perempuan* = 0 dan *Laki-laki* = 1. Selanjutnya, atribut usia dikelompokkan ke dalam beberapa kategori usia yang kemudian direpresentasikan dalam bentuk kode numerik. Transformasi serupa juga dilakukan pada atribut frekuensi penggunaan produk serta tingkat kepuasan pengguna.

Atribut kepuasan berperan sebagai variabel target (*class label*) dalam proses klasifikasi, sedangkan atribut lainnya berfungsi sebagai variabel prediktor.

Tabel 4. Contoh Dataset Setelah Transformasi Data

No	Jenis Kelamin	Usia	Frekuensi	Jenis Produk	Merek Produk	Kualitas	Aroma	Tekstur	Efektivitas	Kepuasan
1	1	0	1	1	1	3	3	1	0	0
2	0	0	3	0	5	0	1	0	2	3
3	1	3	2	1	5	2	1	0	2	2
4	1	0	0	0	4	0	3	1	1	1
5	0	1	1	1	0	2	0	0	1	2
...	...	...	...	...	...	...	...	...	...	...
250	0	1	2	0	2	2	1	0	1	2

Tabel 4 menampilkan contoh dataset yang telah melalui proses transformasi, di mana seluruh atribut kategorikal telah diubah menjadi nilai numerik. Dataset penelitian terdiri dari 250 data responden, yang masing-masing merepresentasikan karakteristik pengguna serta penilaian mereka terhadap produk *Body care*.

Dengan dilakukannya transformasi data, dataset menjadi lebih terstruktur dan siap digunakan pada tahap selanjutnya yaitu pemodelan klasifikasi menggunakan algoritma *Decision Tree* untuk menganalisis tingkat kepuasan pengguna.

3.8 Klasifikasi Data

Tahap klasifikasi data merupakan inti dari proses analisis dalam penelitian ini. Klasifikasi bertujuan untuk mengelompokkan data survei penggunaan produk *Body care* ke dalam kelas-kelas tertentu berdasarkan pola yang terbentuk dari atribut-atribut yang telah dipreprocessing. Pendekatan yang digunakan dalam penelitian ini adalah supervised learning, di mana setiap data telah memiliki label kelas yang merepresentasikan tingkat kepuasan pengguna. Label kelas ini berfungsi sebagai acuan bagi algoritma dalam mempelajari hubungan antara atribut input dan kelas target.

Secara metodologis, klasifikasi data dilakukan untuk mengidentifikasi pola perilaku pengguna produk *Body care* yang tidak dapat dilihat secara langsung melalui analisis deskriptif. Dengan menggunakan algoritma klasifikasi, pola-pola tersebut dapat diungkap secara sistematis dan objektif. Dataset yang telah melalui tahap *preprocessing* kemudian digunakan dalam proses klasifikasi. Pada tahap ini, data dibagi menjadi data latih dan data uji untuk memastikan bahwa model tidak hanya mampu mengenali data yang telah dipelajari, tetapi juga mampu melakukan generalisasi terhadap data baru.

3.9 Pelatihan Model

Pelatihan model dilakukan dengan menggunakan algoritma *Decision Tree*. Algoritma ini dipilih karena memiliki keunggulan dalam hal interpretabilitas dan kemudahan dalam menjelaskan proses pengambilan keputusan. Dalam konteks penelitian survei penggunaan produk *Body care*, kemampuan untuk memahami alur pengambilan keputusan menjadi aspek yang sangat penting. Pada tahap pelatihan, data latih digunakan untuk membangun struktur pohon keputusan. Algoritma *Decision Tree* akan memilih atribut terbaik pada setiap node berdasarkan kriteria tertentu, hingga terbentuk struktur pohon yang mampu memisahkan data ke dalam kelas-kelas yang sesuai.

Secara metodologis, pelatihan model bertujuan untuk menanamkan pola hubungan antara atribut penggunaan produk *Body care* dan tingkat kepuasan pengguna ke dalam model. Proses ini memungkinkan model untuk mempelajari karakteristik masing-masing kelas secara sistematis. Hasil dari tahap pelatihan ini adalah model *Decision Tree* yang siap digunakan untuk melakukan klasifikasi terhadap data uji.

3.10 Perhitungan Entropy Dataset Awal

Perhitungan entropy dilakukan untuk mengukur tingkat ketidakpastian data sebelum dilakukan proses pemisahan. Rumus entropy yang digunakan pada algoritma C4.5 adalah sebagai berikut:

$$Entropy(S) = \sum_{i=1}^n p_i \log_2(p_i) \quad (1)$$

Berdasarkan distribusi kelas pada Tabel 5., maka perhitungan entropy dataset awal adalah:

$$Entropy(S) = \left(\frac{72}{250}\right) \log_2\left(\frac{72}{250}\right)$$



$$\left(\frac{69}{250}\right) \log_2 \left(\frac{69}{250}\right)$$

$$\left(\frac{61}{250}\right) \log_2 \left(\frac{61}{250}\right)$$

$$\left(\frac{48}{250}\right) \log_2 \left(\frac{48}{250}\right)$$

Hasil perhitungan menunjukkan bahwa nilai entropy dataset awal adalah sebesar:

$$Entropy (S) = 1,998$$

Nilai entropy yang mendekati 2 menunjukkan bahwa data memiliki tingkat heterogenitas yang tinggi, sehingga diperlukan proses pemisahan data menggunakan atribut tertentu.

3.11 Perhitungan *Information Gain* Atribut Merk Produk

Perhitungan *Information Gain* dilakukan untuk menentukan atribut terbaik yang digunakan sebagai pemisah data pada pembentukan pohon keputusan. Pada penelitian ini, atribut Merek Produk dianalisis untuk mengetahui tingkat pengaruhnya terhadap klasifikasi kepuasan konsumen. Distribusi data atribut Merek Produk terhadap kelas kepuasan ditunjukkan pada tabel berikut.

Tabel 5. Data Atribut Merk Produk

Merek Produk	Sangat Puas	Puas	Cukup Puas	Tidak Puas	Total
Vaseline	20	18	12	5	55
Dove	18	20	15	8	61
Scarlett Whitening	22	15	14	10	61
Wardah	12	16	20	25	73
Total	72	69	61	48	250

3.12 Evaluasi Model

Evaluasi model dilakukan untuk mengetahui kinerja metode *Decision Tree* dalam mengklasifikasikan tingkat kepuasan pengguna berdasarkan data yang digunakan dalam penelitian. Evaluasi model bertujuan untuk menilai tingkat ketepatan hasil klasifikasi yang dihasilkan oleh model terhadap data aktual. Pada penelitian ini, evaluasi model dilakukan menggunakan confusion matrix, serta pengukuran nilai akurasi, *Precision*, *Recall*, dan *F1-score*. Proses evaluasi dilakukan menggunakan seluruh data penelitian sebanyak 250 data responden yang diperoleh dari hasil kuesioner. Hasil klasifikasi model *Decision Tree* terhadap data tersebut disajikan dalam bentuk confusion matrix, sebagaimana ditunjukkan pada gambar yang telah disajikan sebelumnya.

3.13 Confusion Matrix

Confusion matrix merupakan tabel yang digunakan untuk menggambarkan perbandingan antara kelas aktual dan kelas hasil prediksi dari model. Nilai pada diagonal utama menunjukkan jumlah data yang berhasil diklasifikasikan dengan benar, sedangkan nilai di luar diagonal utama menunjukkan jumlah data yang mengalami kesalahan klasifikasi.

Confusion matrix ini menjadi dasar dalam menghitung nilai akurasi, *Precision*, *Recall*, dan *F1-score* pada penelitian ini.

Tabel 6. *Confusion Matrix* Klasifikasi Kepuasan Konsumen

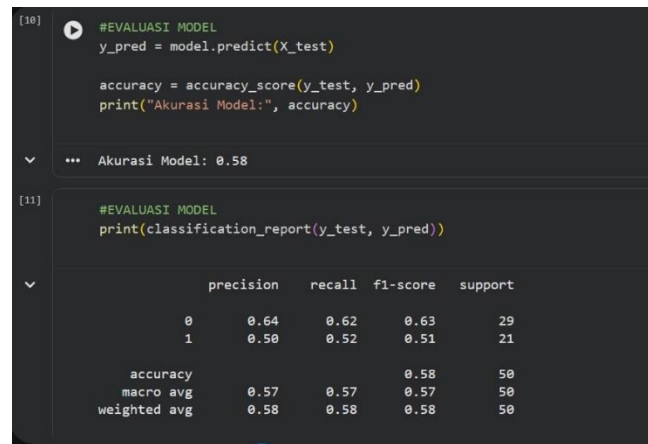
Aktual \ Prediksi	Sangat Puas	Puas	Cukup Puas	Tidak Puas	Total
Sangat Puas	60	8	4	0	72
Puas	6	52	9	2	69
Cukup Puas	2	8	45	6	61
Tidak Puas	0	3	7	38	48
Total	68	71	65	46	250

Berdasarkan hasil evaluasi menggunakan confusion matrix, akurasi, *Precision*, *Recall*, dan *F1-score*, dapat disimpulkan bahwa model *Decision Tree* yang dibangun memiliki kinerja yang cukup baik dalam mengklasifikasikan tingkat kepuasan pengguna. Model mampu mengklasifikasikan 146 dari 250 data dengan benar, yang menunjukkan bahwa metode *Decision Tree* telah mampu mempelajari pola hubungan antara atribut dan tingkat kepuasan pengguna.

Meskipun nilai evaluasi yang diperoleh belum optimal, hasil ini sudah cukup menunjukkan bahwa model dapat digunakan sebagai alat bantu analisis dalam penelitian ini. Peningkatan kinerja model dapat dilakukan pada penelitian selanjutnya dengan penambahan atribut atau pengembangan metode klasifikasi.

Berdasarkan hasil pengujian menggunakan data uji sebanyak 50 data, model *Decision Tree* menghasilkan nilai akurasi sebesar 58%. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan 58% data dengan benar sesuai kelas aktualnya. Selain itu, nilai *precision* sebesar 57%, *recall* sebesar 57%, dan *F1-score* sebesar 57% mengindikasikan bahwa kemampuan model dalam membedakan setiap kelas kepuasan pengguna masih berada pada kategori sedang. Pada

kelas 0, model memperoleh *precision* sebesar 64%, *recall* sebesar 62%, dan *F1-score* sebesar 63%. Hasil ini menunjukkan bahwa model relatif lebih baik dalam mengenali data pada kelas 0 dibandingkan kelas 1. Sementara itu, pada kelas 1 diperoleh *precision* sebesar 50%, *recall* sebesar 52%, dan *F1-score* sebesar 51%, yang menunjukkan masih terdapat cukup banyak kesalahan klasifikasi pada kelas tersebut. Rendahnya performa model dapat disebabkan oleh beberapa faktor. Pertama, jumlah dataset yang digunakan relatif terbatas, yaitu hanya 250 data responden, sehingga pola yang terbentuk belum mampu merepresentasikan kondisi sebenarnya secara optimal. Kedua, kemungkinan terdapat atribut yang kurang relevan atau memiliki korelasi rendah terhadap variabel target sehingga memengaruhi proses pembentukan pohon keputusan. Ketiga, distribusi data antar kelas yang kurang seimbang dapat menyebabkan model lebih cenderung mengenali kelas tertentu dibandingkan kelas lainnya.



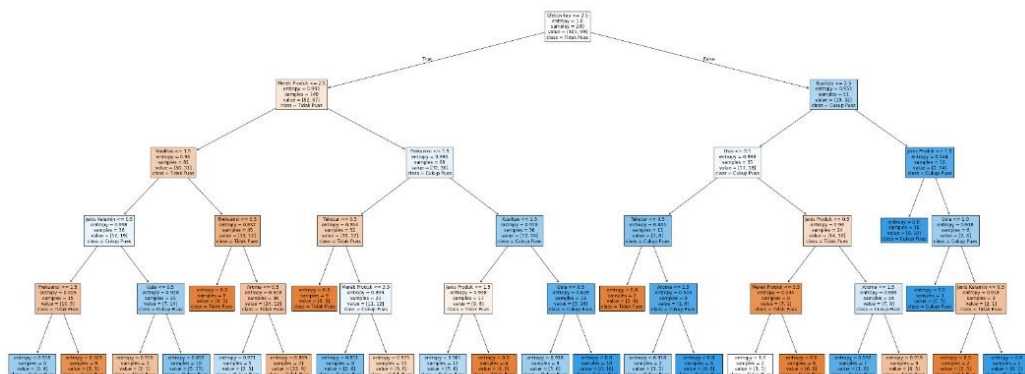
Gambar 5. Confusion Matrix Hasil Evaluasi Model *Decision Tree*

Meskipun demikian, pohon keputusan yang dihasilkan tetap mampu memberikan informasi penting mengenai faktor-faktor yang memengaruhi tingkat kepuasan pengguna produk *body care*. Model menunjukkan bahwa atribut yang berada pada node akar dan percabangan awal memiliki pengaruh dominan dalam proses klasifikasi. Informasi tersebut dapat dimanfaatkan oleh perusahaan untuk mengidentifikasi aspek produk maupun layanan yang perlu ditingkatkan guna meningkatkan kepuasan pelanggan.

Untuk penelitian selanjutnya, performa model dapat ditingkatkan melalui penambahan jumlah data, penerapan teknik feature selection, optimasi parameter *Decision Tree*, atau penggunaan algoritma lain seperti Random Forest, XGBoost, maupun C4.5 yang umumnya memiliki kemampuan klasifikasi yang lebih baik pada dataset dengan karakteristik serupa.

3.14 Klasifikasi Hasil

Tahap klasifikasi merupakan proses penerapan algoritma *Decision Tree* terhadap data survei penggunaan dan kepuasan pengguna produk *Body care*. Model yang telah dilatih kemudian digunakan untuk mengklasifikasikan data responden ke dalam beberapa kelas tingkat kepuasan berdasarkan atribut yang telah diproses pada tahap sebelumnya. Hasil klasifikasi menunjukkan bahwa model mampu mengelompokkan responden ke dalam kelas Tidak Puas, Cukup Puas, Puas, dan Sangat Puas berdasarkan pola hubungan antar atribut. Hal ini menunjukkan bahwa atribut yang digunakan dalam penelitian memiliki keterkaitan dengan tingkat kepuasan pengguna terhadap produk *Body care*. Struktur pohon keputusan yang dihasilkan juga memberikan gambaran mengenai faktor-faktor yang memengaruhi kepuasan pengguna. Atribut yang berada pada bagian atas pohon keputusan menunjukkan variabel yang paling dominan dalam proses klasifikasi. Dengan demikian, hasil klasifikasi tidak hanya menghasilkan prediksi kelas, tetapi juga memberikan informasi mengenai faktor yang berpengaruh terhadap kepuasan pengguna.



Gambar 6. Hasil Klasifikasi Survei Penggunaan Produk *Body care*



Gambar 6 menunjukkan model klasifikasi yang dihasilkan menggunakan algoritma *Decision Tree*. Pohon keputusan dibangun berdasarkan sejumlah atribut yang digunakan dalam penelitian, seperti kualitas produk, harga, pelayanan, frekuensi penggunaan, dan karakteristik pengguna lainnya untuk menentukan tingkat kepuasan pengguna produk *body care*.

Node paling atas (*root node*) merupakan atribut yang memiliki nilai gain tertinggi sehingga dipilih sebagai pemisah utama data. Dari node akar tersebut, data dibagi ke dalam beberapa cabang berdasarkan nilai atribut tertentu hingga menghasilkan node-node berikutnya. Setiap node internal merepresentasikan proses pengambilan keputusan berdasarkan suatu atribut, sedangkan node daun (*leaf node*) menunjukkan hasil klasifikasi akhir berupa kategori tingkat kepuasan pengguna.

Warna yang berbeda pada setiap node menunjukkan kelas dominan yang diprediksi oleh model. Node berwarna biru merepresentasikan kategori pengguna yang cenderung puas, sedangkan node berwarna oranye menunjukkan kategori pengguna yang cenderung tidak puas atau kategori lain sesuai dengan label yang digunakan dalam penelitian. Intensitas warna yang semakin pekat mengindikasikan tingkat kemurnian (*purity*) data yang semakin tinggi pada node tersebut, yang berarti sebagian besar data pada node tersebut berasal dari kelas yang sama..

3.15 Hasil

Hasil penelitian menunjukkan bahwa algoritma *Decision Tree* mampu mengklasifikasikan tingkat kepuasan pengguna produk *Body care* ke dalam empat kategori, yaitu Sangat Puas, Puas, Cukup Puas, dan Tidak Puas. Berdasarkan hasil evaluasi, model menghasilkan akurasi sebesar 58%, precision sebesar 57%, recall sebesar 57%, dan F1-score sebesar 57%. Nilai tersebut menunjukkan bahwa *Decision Tree* memiliki kemampuan klasifikasi pada kategori sedang dalam mengidentifikasi tingkat kepuasan pengguna.

Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh Putri dkk pada tahun 2022 yang menerapkan algoritma *Decision Tree* pada analisis kepuasan pelanggan produk kecantikan dan memperoleh tingkat akurasi sebesar 60%. Penelitian tersebut menyatakan bahwa atribut kualitas produk dan pelayanan menjadi faktor utama dalam menentukan kepuasan konsumen. Temuan tersebut memiliki kesamaan dengan penelitian ini, di mana atribut kualitas produk, harga, dan pelayanan juga memberikan pengaruh terhadap pembentukan pohon keputusan.

Penelitian ini juga memiliki kesamaan dengan penelitian Sari dan Nugroho pada tahun 2021 mengenai klasifikasi kepuasan pelanggan menggunakan metode *Decision Tree* pada sektor ritel. Penelitian tersebut menunjukkan bahwa algoritma *Decision Tree* memiliki keunggulan dalam menghasilkan aturan keputusan yang mudah dipahami oleh pengguna. Pada penelitian ini, struktur pohon keputusan yang dihasilkan juga mampu menunjukkan atribut-atribut yang paling dominan dalam menentukan tingkat kepuasan pengguna produk *Body care*.

Namun demikian, nilai akurasi yang diperoleh dalam penelitian ini masih lebih rendah dibandingkan penelitian Rahman dkk. Pada Tahun 2023 yang menggunakan algoritma *Random Forest* untuk klasifikasi kepuasan pelanggan dan memperoleh akurasi sebesar 78%. Perbedaan tersebut dapat disebabkan oleh jumlah data yang digunakan, karakteristik dataset, serta kemampuan algoritma yang berbeda dalam menangani variasi data. *Random Forest* sebagai pengembangan dari *Decision Tree* mampu mengurangi *overfitting* dan meningkatkan kemampuan generalisasi model.

Berdasarkan perbandingan dengan penelitian terdahulu, dapat disimpulkan bahwa hasil penelitian ini masih berada dalam rentang performa yang wajar untuk metode *Decision Tree*. Meskipun akurasi yang diperoleh belum optimal, model yang dihasilkan mampu memberikan informasi yang mudah dipahami mengenai faktor-faktor yang memengaruhi kepuasan pengguna produk *Body care*. Keunggulan utama penelitian ini terletak pada kemampuan interpretasi model yang dapat digunakan sebagai dasar pengambilan keputusan oleh perusahaan dalam meningkatkan kualitas produk dan layanan. Dengan demikian, penelitian ini memberikan kontribusi dalam penerapan algoritma *Decision Tree* untuk analisis kepuasan pengguna produk *Body care*, sekaligus menunjukkan bahwa peningkatan jumlah data, optimasi parameter, maupun penggunaan algoritma lain seperti *Random Forest*, C4.5, atau *XGBoost* berpotensi meningkatkan performa klasifikasi pada penelitian selanjutnya.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *Decision Tree* dapat diterapkan untuk mengklasifikasikan tingkat kepuasan pengguna produk *body care* berdasarkan data survei responden. Proses pengolahan data yang meliputi preprocessing, seleksi atribut, transformasi data, pelatihan model, dan evaluasi berhasil menghasilkan model klasifikasi yang mampu mengelompokkan tingkat kepuasan pengguna ke dalam kategori sangat puas, puas, cukup puas, dan tidak puas. Hasil evaluasi menggunakan confusion matrix menunjukkan bahwa model memperoleh nilai akurasi sebesar 58%, precision sebesar 57%, recall sebesar 57%, dan F1-score sebesar 57%. Nilai tersebut menunjukkan bahwa performa model masih berada pada kategori sedang, namun sudah mampu menggambarkan pola hubungan antara atribut pengguna dan tingkat kepuasan pelanggan. Hasil penelitian juga menunjukkan bahwa atribut kualitas produk dan pelayanan menjadi faktor dominan yang memengaruhi tingkat kepuasan pengguna produk *body care*. Struktur pohon keputusan yang dihasilkan memberikan gambaran yang mudah dipahami mengenai hubungan antar atribut dalam proses klasifikasi, sehingga dapat membantu perusahaan dalam memahami perilaku konsumen serta menentukan strategi peningkatan kualitas produk dan pelayanan secara lebih tepat. Penelitian ini memberikan kontribusi akademis dan praktis dalam penerapan teknik data mining, khususnya algoritma *Decision Tree*, pada analisis kepuasan pelanggan.



di industri body care. Secara akademis, penelitian ini memperluas implementasi metode klasifikasi pada domain kepuasan pengguna produk kecantikan yang masih relatif terbatas. Secara praktis, hasil penelitian dapat menjadi bahan pertimbangan bagi perusahaan dalam menyusun strategi pemasaran, meningkatkan kualitas layanan, serta melakukan evaluasi terhadap faktor-faktor yang memengaruhi loyalitas pelanggan. Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan, seperti jumlah dataset yang relatif terbatas dan kemungkinan adanya atribut yang kurang optimal dalam proses klasifikasi. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan jumlah data yang lebih besar, menerapkan teknik feature selection, melakukan optimasi parameter *Decision Tree*, atau menggunakan algoritma lain seperti Random Forest, XGBoost, maupun C4.5 guna meningkatkan performa model klasifikasi. Dengan demikian, *Decision Tree* dalam penelitian ini dapat digunakan sebagai pendekatan awal untuk analisis pola kepuasan pelanggan berbasis data.

REFERENCES

- [1] S. A. Fenny Krisna Marpaung, Markus Willy Arnold S, Asyifa Sofira, "Pengaruh Harga, Promosi, Dan Kualitas Produk Terhadap Keputusan Pembelian Konsumen Indomie Pada Pt. Alamjaya," vol. 7, pp. 49–64, 2021.
- [2] F. Septianingrum and A. S. Y. Irawan, "Metode Seleksi Fitur Untuk Klasifikasi Sentimen Menggunakan Algoritma Naive Bayes: Sebuah Literature Review," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 799, 2021, doi: 10.30865/mib.v5i3.2983.
- [3] M. K. Insan, U. Hayati, and O. Nurdian, "Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naive Bayes," *J. Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 478–483, 2023.
- [4] S. Saniah and M. Furqan, "Classification Of Rice Plant Diseases Using K-Nearest Neighbor Algorithm Based On Hue Saturation Value Color Extraction And Gray Level Co-Occurrence Matrix Features," vol. 7, no. 2, pp. 212–223, 2025, doi: 10.36378/jtos.v7i2.3972.
- [5] A. Ilyasa *et al.*, "Diagnosa Penyakit Kulit Wajah Dengan Metode *Decision Tree* dan Algoritma C4.5," vol. 6, pp. 88–98, 2025.
- [6] A. I. Putri, M. Furqan, and Suhardi, "Application of Data Mining to Predict Birth Rates in Medan City Using the K-Nearest Neighbor Method," vol. 5, no. 1, pp. 504–509, 2024, doi: 10.30596/jcosite.v5i1.17991.
- [7] N. Y. L. Gaol, "Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam *Decision Tree* dan Algoritma C4.5," *J. Inf. Teknol.*, vol. 2, pp. 23–29, 2020, doi: 10.37034/jidt.v2i1.22.
- [8] M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, "Penerapan Algoritma *Decision Tree* dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa," *G-Tech J. Teknol. Terap.*, vol. 6, no. 2, pp. 295–305, 2022.
- [9] D. M. Musa *et al.*, "Penerapan Data Mining Untuk Klasifikasi Data Penjualan Pakan Ternak Terlaris Dengan Algoritma C4.5," *J. Teknologi Inform. dan Komput. MH. Thamrin*, vol. 10, no. 1, pp. 168–182, 2024.
- [10] W. Widia, B. Ginting, H. R. Sitepu, L. Nainggolan, A. S. Purba, and A. Br, "Penerapan Algoritma *Decision Tree* Data Mining untuk Prediksi Pola Pemberian Kredit pada Koperasi Simpan Pinjam," vol. 4, no. 3, pp. 268–276, 2025, doi: 10.47065/jogtc.v4i3.8336.
- [11] A. S. Siahaan, R. Saragih, M. Simanjuntak, and S. K. Binjai, "Penerapan Metode K-Means Clustering untuk Pengelompokan Minat Konsumen terhadap Pengguna Jasa Layanan pada Kantor Pos Binjai," vol. 2, no. 5, 2024.
- [12] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma *Decision Tree* C4.5," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, 2020, doi: 10.47970/siskom-kb.v4i1.173.
- [13] M. Solehuddin, W. A. Syaefi, and R. Gernowo, "Metode *Decision Tree* untuk Meningkatkan Kualitas Rencana Pelaksanaan Pembelajaran dengan Algoritma C4.5," *J. Penelit. dan Pengemb. Pendidik.*, vol. 6, no. 3, pp. 510–519, 2022, doi: 10.23887/jppp.v6i3.52840.
- [14] A. Afifuddin and L. Hakim, "Deteksi Penyakit Diabetes Mellitus Menggunakan Algoritma *Decision Tree* Model Arsitektur C4.5," *J. Krisnadana*, vol. 3, no. 1, pp. 25–33, 2023, doi: 10.58982/krisnadana.v3i1.470.
- [15] D. Muriyatmoko, A. Musthafa, and M. H. Wijaya, "Klasifikasi Profil Kelulusan Nilai AKPAM Dengan Metode *Decision Tree*," *Semin. Nas. Sains dan Teknol. 2024 Fak.*, no. April, pp. 448–453, 2024.
- [16] C. R. A. Nugroho and T. Kristiana, "Penerapan Algoritma C4.5 Untuk Kepuasan Pelanggan Toko," *J. Algoritma*, vol. 3, no. 1, pp. 10–21, 2022.
- [17] A. Steffany, K. A. Aldzikri, M. A. R. Tricahya, and M. I. Arfiandono, "Penerapan Algoritma C4.5 Untuk Menganalisis Kepuasan E-Commerce Shopee," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 6, no. 6, pp. 700–706, 2023, doi: 10.32672/jnkti.v6i6.7112.
- [18] Y. L. Fatma and N. Rochmawati, "Prediksi Siswa Putus Sekolah Menggunakan Algoritma *Decision Tree* C4.5," *J. Informatics Comput. Sci.*, vol. 5, no. 04, pp. 486–493, 2024, doi: 10.26740/jinacs.v5n04.p486-493.
- [19] R. Rofiani, L. Oktaviani, D. Vernanda, and T. Hendriawan, "Penerapan Metode Klasifikasi *Decision Tree* dalam Prediksi Kanker Paru-Paru Menggunakan Algoritma C4.5," *J. Tekno Kompak*, vol. 18, no. 1, pp. 126–139, 2024, doi: 10.33365/jtk.v18i1.3525.
- [20] T. Widiastuti, K. Karsa, and C. Juliane, "Evaluasi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademik Menggunakan Metode Klasifikasi Algoritma C4.5," *Technomedia J.*, vol. 7, no. 3, pp. 364–380, 2022, doi: 10.33050/tmj.v7i3.1932.
- [21] Z. H. . Veren Nita, P., Raisah, F. A., Yuma Akbar & Aditya, "Analisis Tingkat Kepuasan Pelanggan Terhadap Pengguna Jasa Layanan Grab Menggunakan Metode C4.5 Abstrak," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 3, pp. 3189–3198, 2024.