
Penerapan *Data Mining* untuk Klasifikasi Tingkat Kepuasan Pelanggan Café Menggunakan Metode *Decision Tree C4.5*

Atta Tha Ariq^{*1}, Sucipto², Rachmat Wahid Saleh Insani³

Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer,
Universitas Muhammadiyah Pontianak

Email: ^{*1}attathaariq@gmail.com, ²sucipto@unmuhpnk.ac.id, ³rachmat.wahid@unmuhpnk.ac.id

(Naskah masuk: 22 Agustus 2025, diterima untuk diterbitkan: 30 Desember 2025)

Abstrak: Kepuasan pelanggan adalah faktor utama dalam meningkatkan reputasi bisnis, loyalitas pelanggan, dan efisiensi operasional. Penelitian ini bertujuan mengembangkan sistem yang memberikan informasi akurat tentang tingkat kepuasan dan ketidakpuasan pelanggan di sebuah café. Harapannya, temuan dari penelitian ini dapat memberikan dampak baik kepada café guna untuk meningkatkan kualitas pelayanan dengan mengetahui apa saja indikator-indikator yang mempengaruhi tingkat kepuasan pelanggan café. Metode yang digunakan adalah *Decision Tree C4.5*, yang membangun pohon keputusan untuk klasifikasi. Proses meliputi penanganan *missing value*, pengecekan *duplicate data*, label encoding, penanganan data imbalance dengan *SMOTE*, pemodelan *Decision Tree C4.5*, pengecekan akurasi, dan visualisasi aturan keputusan. Evaluasi model dilakukan menggunakan metrik *confusion matrix*. Hasil evaluasi menunjukkan bahwa model klasifikasi memiliki performa sangat baik, dengan akurasi 98% pada data latih dan 93% pada data uji. Nilai recall, precision, dan F1-score masing-masing adalah 94%, 97%, dan 95%.

Kata Kunci – *Confusion Matrix; Data Mining; Decision Tree C4.5; Kepuasan Pelanggan; Klasifikasi*

Application of Data Mining for Classification of Café Customer Satisfaction Levels using The C4.5 Decision Tree Method

Abstract: Customer satisfaction is a key factor in enhancing business reputation, customer loyalty, and operational efficiency. This study aims to develop a system that provides accurate information about the level of customer satisfaction and dissatisfaction at a café. The findings of this study are expected to help the café improve service quality and maintain customer loyalty, as well as provide added value for researchers and practitioners. The method used is the *Decision Tree C4.5*, which constructs a decision tree for classification. The process includes handling missing values, checking for duplicate data, label encoding, addressing data imbalance with *SMOTE*, *Decision Tree C4.5* modeling, accuracy checking, and decision rule visualization. Model evaluation is conducted using the *confusion matrix* metrics. The evaluation results indicate that the classification model performs very well, with 98% accuracy on training data and 93% on test data. The recall, precision, and F1-score are 94%, 97%, and 95%, respectively.

Keywords – *Customer Satisfaction; Classification; Confusion Matrix; Data Mining; Decision Tree C4.5*

1. PENDAHULUAN

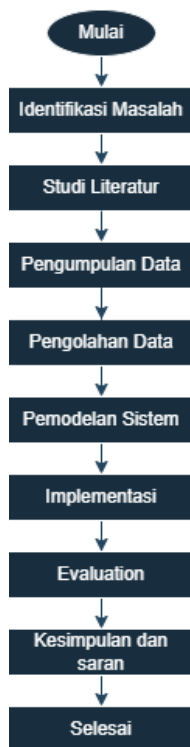
Kepuasan pelanggan pada dasarnya mencakup perbedaan antara tingkat kepentingan dan hasil yang dirasakan, serta merupakan evaluasi setelah pembeliandimana alternatif yang dipilih minimal dapat memenuhi atau melebihi harapan konsumen. Sebaliknya jika ketidakpuasan terjadi maka hasil yang diterima pun tidak memenuhi harapan konsumen. Masalah pelayanan sebenarnya tidaklah sulit atau rumit, tetapi jika diabaikan, dapat menimbulkan masalah bagi perusahaan [1]. Pelanggan yang puas akan terus datang kembali, sedangkan pelanggan yang tidak puas akan berhenti berkunjung.

Hal ini dapat menyebabkan kehilangan para pelanggan [2]. Kepuasan pelanggan merupakan salah satu faktor utama yang mendukung peningkatan reputasi bisnis, loyalitas pelanggan, pengurangan biaya produksi, penurunan elastisitas harga, dan peningkatan efisiensi serta

produktivitas karyawan [3]. Hal ini tentu saja berdampak pada tingkat kepuasan pelanggan cafe. Oleh karena itu, perusahaan perlu melakukan evaluasi untuk mengidentifikasi faktor-faktor yang mempengaruhi kepuasan pelanggan, sehingga dapat mencapai keunggulan kompetitif melalui peningkatan efisiensi pelayanan. Dalam menilai kualitas layanan, indikator kepuasan pelanggan sangat penting dalam bidang bisnis. Dengan memanfaatkan teknik *data mining*, pemilik usaha dapat mengidentifikasi faktor utama yang mempengaruhi kepuasan atau ketidakpuasan pelanggan, seperti kualitas makanan, ketepatan waktu pelayanan, dan kinerja karyawan.

2. METODE PENELITIAN

Proses dalam penelitian ini dimulai dengan mengidentifikasi masalah dan berlanjut hingga tahap kesimpulan dan saran.



Gambar 1. Tahapan Penelitian

2.1 Identifikasi Masalah

Pada tahap ini, penulis mengidentifikasi berbagai permasalahan yang muncul dalam data mining terkait klasifikasi kepuasan pelanggan café. Hal ini dilakukan agar penulis dapat menentukan data apa saja yang dibutuhkan untuk penelitian. Langkah ini sangat penting karena pemahaman yang mendalam mengenai masalah yang dihadapi akan membantu dalam mengumpulkan data yang relevan dan memastikan bahwa analisis yang dilakukan tepat sasaran, menghasilkan temuan yang akurat dan bermanfaat.

2.2 Studi Literatur

Pada tahap ini, peneliti mencari teori dari berbagai sumber, termasuk jurnal dan referensi lainnya, untuk digunakan sebagai dasar dalam penelitian. Langkah ini penting karena menyediakan landasan teori yang kuat, yang membantu dalam memahami konteks penelitian dan mengarahkan analisis serta interpretasi data secara lebih akurat.

2.3 Pengumpulan Data

Pengumpulan data dalam penelitian ini menggunakan *Google Form kuesioner* sebagai alat survei untuk mendapatkan data dari responden. data yang diperoleh dari pembagian kuesioner

sebanyak 200 *kuesioner*. Responden diminta untuk memberikan tanggapan terkait faktor-faktor kunci, seperti harga, pelayanan yang diberikan, kualitas produk, loyalitas, serta tempat.

Tabel 1. Atribut Dalam Dataset

No	Atribut	Indikator	Pernyataan
1.	Harga	Harga terjangkau	Tarif harga relatif murah
		Harga sesuai dengan produk	Pelayanan yang diberikan sesuai dengan harga
		Daya saing harga	Mutu yang diperoleh pas dengan harga
		Manfaat sesuai harga	Perbandingan harga dari tempat lain
2.	Pelayanan	Keandalan	Kecepatan dalam penyajian
		Berwujud	Melayani dengan profesional
		Sikap	Pelayanan sopan dan ramah
		Perhatian	Menerima kritik dan saran
		Tindakan	Cepat dan tanggap dalam melayani permintaan konsumen
3.	Kualitas produk	Kinerja	Kinerja karyawan cepat dalam menyelesaikan produk
		Kesesuaian dalam spesifikasi	Produk sesuai dengan yang dipesan
		Daya tahan	Produk makanan dan minuman masih fresh
		Kemampuan diperbaiki	Mampu mengatasi produk makanan yang bermasalah atau yang tidak sesuai dengan yang dipesan
4.	Loyalitas	Persepsi pada kualitas	Memiliki rasa yang enak
			Secara keseluruhan anda lebih sering kesini dari pada tempat lain
			Jika ada seseorang yang meminta saran anda, maka café mana yang sebaiknya dikunjungi, anda akan menyarankan Popina bistro and cafe
			Anda akan mengatakan hal yang baik tentang popina bistro and cafe
			Harga bukan masalah bagi anda untuk mengunjungi popina bistro and cafe
5.	Tempat	Letak lokasi	Tempat lokasi sangat strategis
		Keamanan	Suasana tempat yang aman
		Kenyamanan	Suasana tempat yang nyaman
		Kebersihan	Lingkungan yang bersih
		Area parkir	Memiliki tempat lokasi parkir yang luas
6	Kepuasan		Bagaimana kepuasan anda mengenai harga, pelayanan, kualitas produk, loyalitas dan tempat yang diberikan Popina bistro and cafe?

Untuk jawaban responden dapat dilihat pada tabel berikut:

Tabel 2. Bobot Skor

Jawaban responden	Bobot skor
Sangat tidak puas	1
Tidak puas	2
Puas	3
Sangat puas	4

2.4 Pengolahan Data

Proses pengolahan data dalam penelitian ini mengikuti tahapan-tahapan yang terdapat dalam kerangka kerja Knowledge Discovery in Databases (KDD). Proses ini meliputi beberapa langkah penting, pembersihan data, transformasi data, hingga penerapan algoritma untuk menemukan pola dan informasi berharga dalam database.

1) Pembersihan data (*Data Cleaning*)

Pada bagian ini, data akan diperiksa agar menemukan nilai yang hilang, duplikat, atau tidak valid. Langkah ini penting untuk memastikan bahwa data tetap bersih dan akurat sebelum digunakan dalam pemodelan. Nilai yang tidak lengkap atau tidak valid akan diperbaiki atau dihapus agar tidak mempengaruhi hasil analisis. Proses ini sangat penting untuk menjaga integritas data dan meningkatkan keandalan hasil analisis yang akan dilakukan.

2) *Data Selection*

Sebelum memasuki tahap penggalian informasi dalam proses KDD, dilakukan terlebih dahulu seleksi data dari kumpulan data operasional. Data yang telah dipilih tersebut akan digunakan dalam proses data mining dan disimpan secara terpisah dari database operasional utama. Langkah ini penting untuk memastikan bahwa hanya data relevan yang dianalisis, sehingga meningkatkan efisiensi dan akurasi dalam proses *data mining*.

3) *Data Transformation*

Proses transformasi data melibatkan perubahan struktur atau format dalam dataset untuk memenuhi kebutuhan pemodelan atau analisis tertentu. Ini dapat mencakup pengkodean ulang variabel kategorikal menjadi variabel numerik, normalisasi, skala data, atau penanganan *outlier*. Transformasi data bertujuan untuk meningkatkan kualitas data, membuatnya lebih mudah diinterpretasikan, dan mempersiapkannya untuk analisis atau model tertentu. Langkah ini penting untuk memastikan data siap digunakan dalam proses analisis yang lebih lanjut.

2.5 Pemodelan Sistem

Pemodelan sistem merupakan salah satu langkah penting dalam pengembangan sistem kepuasan pelanggan. Proses ini dapat mencakup berbagai aspek, termasuk desain struktur sistem. Dengan memodelkan sistem secara menyeluruh, kita dapat memastikan bahwa semua elemen berfungsi dengan baik dan efisien untuk mencapai tujuan utama, yaitu meningkatkan kepuasan pelanggan. Memahami setiap komponen dalam desain struktur sistem membantu dalam menciptakan solusi yang lebih responsif dan adaptif terhadap kebutuhan pelanggan.

2.5.1 UML Diagram

Perancangan sistem dalam pembuatan studi ini menggunakan diagram *Unified Modeling Language (UML)*. Diagram UML ini membantu memvisualisasikan struktur dan perilaku sistem secara menyeluruh. Dengan demikian, pemahaman dan komunikasi antara pengembang dan pemangku kepentingan menjadi lebih mudah dan efektif.

2.5.2 Perancangan Antarmuka

Langkah ini mencakup perancangan antarmuka pengguna (UI) yang memfasilitasi interaksi pengguna dengan sistem. Proses ini bertujuan untuk menciptakan UI yang intuitif dan efisien guna mendukung penggunaan sistem secara optimal. Dengan demikian, pengguna akan merasa lebih nyaman dan tidak mengalami kesulitan saat menggunakan sistem, yang pada akhirnya meningkatkan kepuasan dan produktivitas mereka.

2.6 Implementasi

Pada tahap ini, model yang telah dibangun menggunakan metode *decision tree C4.5* akan diimplementasikan melalui *coding* dengan bahasa pemrograman *Python*. Selain itu, *framework Streamlit* akan digunakan untuk mengembangkan aplikasi *web* interaktif. Langkah ini penting untuk memastikan model dapat diaplikasikan secara praktis dan memberikan kemudahan bagi pengguna dalam berinteraksi dengan hasil penelitian.

2.7 Evaluasi

Tahap ini adalah proses evaluasi kinerja algoritma dengan menghitung *Confusion Matrix* untuk menentukan nilai *accuracy*, *recall*, dan *precision*. Evaluasi ini penting untuk memahami seberapa baik algoritma dalam mengklasifikasikan data dan memastikan hasil yang dihasilkan dapat diandalkan.

2.8 Kesimpulan dan Saran

Pada tahap ini, hasil penelitian akan dijelaskan secara rinci, termasuk kekurangan dan kelebihan, untuk memberikan panduan bagi penelitian selanjutnya. Penjabaran ini penting agar peneliti berikutnya dapat memanfaatkan temuan ini sebagai dasar untuk memperbaiki dan mengembangkan penelitian lebih lanjut.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Explore Data

Penelitian ini memanfaatkan statistik deskriptif dan visualisasi data untuk menganalisis dan memahami dataset yang digunakan. Statistik deskriptif berfungsi memberikan ringkasan umum mengenai data, seperti jumlah, rata-rata, median, dan ukuran statistik lainnya. Visualisasi data digunakan untuk membuat grafik dan diagram yang membantu menampilkan pola dan hubungan dalam data.

3.1.1 Statistik deskriptif

Statistik deskriptif merupakan teknik analisis data yang digunakan untuk merangkum dan menyajikan karakteristik dasar dari dataset melalui ukuran statistik seperti total frekuensi, mean, minimum, maksimum, dan nilai-nilai sentral lainnya. Pendekatan ini bertujuan untuk memberikan representasi ringkas yang dapat memperjelas distribusi serta kecenderungan data sebelum dilakukan analisis lebih lanjut.

Hasil statistik deskriptif dari *dataset* dapat dilihat pada gambar 2.

	berwujud	kinerja	kesesuaian	kualitas	rekomendasi	opini_pribadi	keamanan	kenyamanan	label
count	200.000000	200.000000	200.000000	200.000000	200.000000	200.000000	200.000000	200.000000	200.000000
mean	3.065000	2.900000	3.060000	3.155000	2.910000	3.000000	3.205000	3.220000	0.820000
std	0.634657	0.501255	0.623171	0.680655	0.577553	0.584558	0.696708	0.695716	0.385152
min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000
25%	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	1.000000
50%	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	1.000000
75%	3.000000	3.000000	3.000000	4.000000	3.000000	3.000000	4.000000	4.000000	1.000000
max	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	1.000000

Gambar 2. Statistik Deskriptif

Gambar 2 merupakan hasil *dataset describe*. berikut ini merupakan penjelasan dari gambar 2:

- 1) *Count*: Jumlah baris data dalam *dataset*.
- 2) *Mean* (rata-rata): Nilai rata-rata dari suatu *dataset*.
- 3) *Std* (Standar deviasi): Pengukuran seberapa jauh nilai-nilai dalam kumpulan data tersebar dari rata-rata (*mean*).
- 4) *25%*: Nilai yang membagi 25% data terendah dari data yang diurutkan. Juga dikenal sebagai kuartil bawah atau *Q1*.
- 5) *50%*: Nilai tengah dari dataset ketika data diurutkan dari yang terkecil hingga yang terbesar atau dikenal sebagai kuartil tengah atau median.
- 6) *75%*: Nilai yang membagi 25% data tertinggi dari data yang diurutkan. Juga dikenal sebagai kuartil atas atau *Q3*.

7) *Max* (Nilai Maksimum): Nilai terbesar dalam *dataset*.

8) *Min* (Nilai Minimum): Nilai terkecil dalam *dataset*.

3.2 Hasil Preprocessing

Tahapan *preprocessing* merupakan proses untuk menyiapkan data sehingga dapat digunakan dalam implementasi metode *decision tree* C4.5.

3.2.1 Missing Value

Nilai yang hilang (*Missing value*) adalah ketika hilangnya data dalam *dataset*. Penanganan *missing value* penting dilakukan untuk memastikan kualitas *dataset*.

Hasil mengecek *missing value* dalam *dataset* dapat dilihat pada gambar 5.2.

```
berwujud      0
kinerja       0
kesesuaian    0
kualitas      0
rekomendasi   0
opini_pribadi 0
keamanan     0
kenyamanan   0
label        0
dtype: int64
```

Gambar 3. Missing Value

Nilai 0 menunjukkan bahwa *dataset* tidak memiliki nilai yang hilang dalam *dataset*.

3.2.2 Duplicated Data

Duplicated data atau data yang terduplikat, adalah ketika terdapat entri yang sama dalam *dataset*. Ini dapat mempengaruhi analisis dan pemodelan data, sehingga perlu mendeteksi dan menghapus data yang terduplikat.

Gambar hasil untuk mengecek *duplicated data* dapat dilihat pada gambar 4.

```
0      False
1      False
2      False
3      False
4      False
...
195    False
196    False
197    False
198    False
199    False
Length: 200, dtype: bool
```

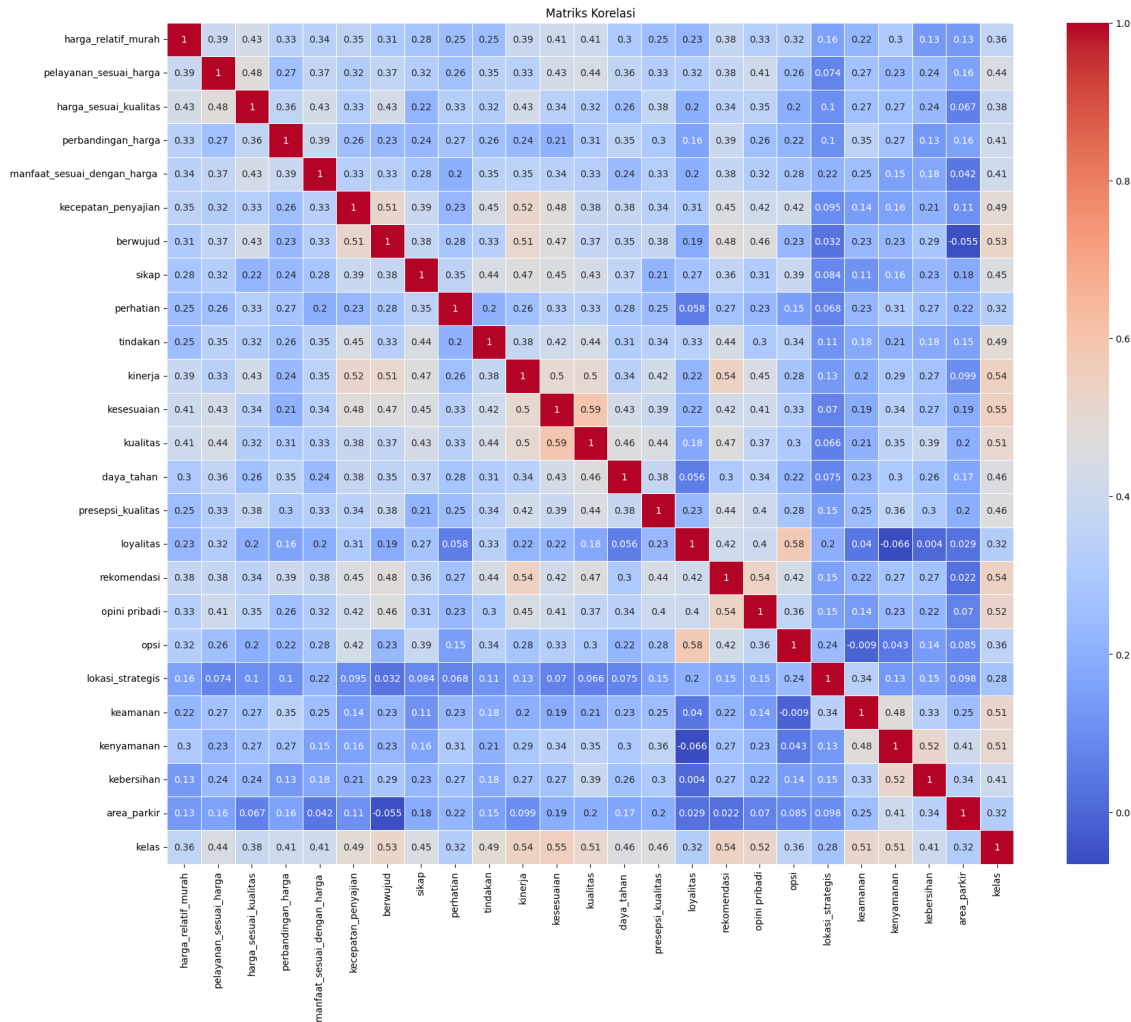
Gambar 4. Duplicated Data

Hasil pengecekan menunjukkan nilai *false* untuk setiap data yang menunjukkan tidak ada data yang terduplikat dalam *dataset* tersebut.

3.2.3 Correlation Matrix

Correlation matrix adalah metrik digunakan untuk mengecek korelasi antara setiap fitur pada dataset.

Gambar correlation matrix pada dataset dapat dilihat pada gambar 5.



Gambar 5. Correlation Matrix

Pada correlation matrix, terdapat delapan atribut dengan nilai korelasi tertinggi terhadap label. Dengan demikian, penelitian ini akan menerapkan 8 dari 25 atribut dalam dataset dengan nilai korelasi lebih dari lima puluh (>50).

3.3 Training Test Split

Pada penelitian ini, dataset dibagi menjadi 80% data pelatihan (training set) dan 20% data pengujian (testing set). Pembagian ini bertujuan agar model dapat dilatih dengan mayoritas data dan dievaluasi menggunakan data yang tidak pernah dilihat sebelumnya untuk mengukur generalisasi model secara objektif. Sebelum pembagian, dilakukan pemisahan antara variabel fitur (input) dan variabel target (output) pada dataset.

Hasil pembagian data latih dan uji dapat dilihat pada gambar 6.

```
x_train shape: (160, 8)
x_test shape: (40, 8)
y_train shape: (160,)
y_test shape: (40,)
```

Gambar 6. Splitting Data

3.4 Imbalanced Data

Imbalanced data merupakan keadaan kelas pada *dataset* mengalami ketimpangan jumlah yang signifikan sehingga perlunya dilakukan *balancing data*. Penelitian ini menerapkan teknik SMOTE untuk mengatasi masalah ketidakseimbangan kelas.

Hasil teknik *SMOTE* dapat dilihat pada gambar 3.6.

```
X_train_resampled shape: (262, 8)
y_train_resampled shape: (262,)
```

Gambar 7. Balancing Data

3.5 Pemodelan Decision Tree C4.5

Proses pemodelan decision tree C4.5 dilakukan sebagai langkah untuk menerapkan metode tersebut pada dataset. Metode decision tree C4.5 diterapkan pada data latih dan data uji untuk membangun serta mengevaluasi model.

Hasil dari penggunaan metode *decision tree C4.5* pada data latih dapat dilihat pada gambar 8.

```
Accuracy Data Latih: 0.9809160305343512
Confusion Matrix:
[[128  3]
 [ 2 129]]
Classification Report:

```

	precision	recall	f1-score	support
0	0.98	0.98	0.98	131
1	0.98	0.98	0.98	131
accuracy			0.98	262
macro avg	0.98	0.98	0.98	262
weighted avg	0.98	0.98	0.98	262

Gambar 8. Penerapan Metode Decision Tree C4.5 Pada Data Latih

Classification report menunjukkan bahwa model berhasil melakukan pembelajaran dari dataset dengan akurasi 98%. Setelah itu, dilakukan Implementasi metode decision tree C4.5 terhadap data uji untuk mengetahui apakah model berhasil melakukan klasifikasi dengan baik atau tidak.

Hasil dari penerapan metode *decision tree C4.5* pada data latih dapat dilihat pada gambar 9.

```

Accuracy Data Uji: 0.925
Confusion Matrix:
[[ 6  1]
 [ 2 31]]
Classification Report:

```

	precision	recall	f1-score	support
0	0.75	0.86	0.80	7
1	0.97	0.94	0.95	33
accuracy			0.93	40
macro avg	0.86	0.90	0.88	40
weighted avg	0.93	0.93	0.93	40

Gambar 9. Penerapan Metode Decision Tree C4.5 Pada Data Uji

Classification report menunjukkan bahwa model berhasil melakukan pembelajaran dari dataset dengan sangat baik yaitu 93% yang artinya model menunjukkan performa yang sangat baik dalam melakukan klasifikasi.

3.6 Hasil Pengujian

Pengujian metode decision tree C4.5 dalam penelitian ini dilakukan dengan menggunakan *confusion matrix* untuk menghitung nilai akurasi, presisi, *recall*, dan F1-score.

1) Akurasi

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$$

$$Accuracy = \frac{31 + 6}{(31 + 6 + 1 + 2)}$$

$$Accuracy = 0,925$$

Hasil akurasi dikali dengan 100% untuk mendapatkan nilai dalam bentuk persen(%). Hasil akurasi adalah 92,5% dibulatkan menjadi 93%.

2) Presisi

$$Precision = \frac{TP}{(TP + FP)}$$

$$Precision = \frac{31}{(31 + 1)}$$

$$Precision = 0,968$$

Hasil presisi dikali dengan 100% untuk mendapatkan nilai dalam bentuk persen(%). Hasil presisi adalah 96,8% dibulatkan menjadi 97%.

3) Recall

$$Recall = \frac{TP}{(TP + FN)}$$

$$Recall = \frac{31}{(31 + 2)}$$

$$Recall = 0,939$$

Hasil *recall* dikali dengan 100% untuk mendapatkan nilai dalam bentuk persen(%). Hasil *recall* adalah 93,9% dibulatkan menjadi 97%.

4) F1-Score

$$F1-Score = \frac{2 \times recall \times presisi}{recall + presisi}$$

$$F1-Score = \frac{2 \times 0,939 \times 0,968}{0,939 + 0,968}$$

$$F1-Score = 0,953$$

Hasil *f1-score* dikali dengan 100% untuk mendapatkan nilai dalam bentuk persen(%). Hasil *f1-score* adalah 95,3% dibulatkan menjadi 95%.

Hasil dari pengujian bisa dilihat pada Tabel 3.

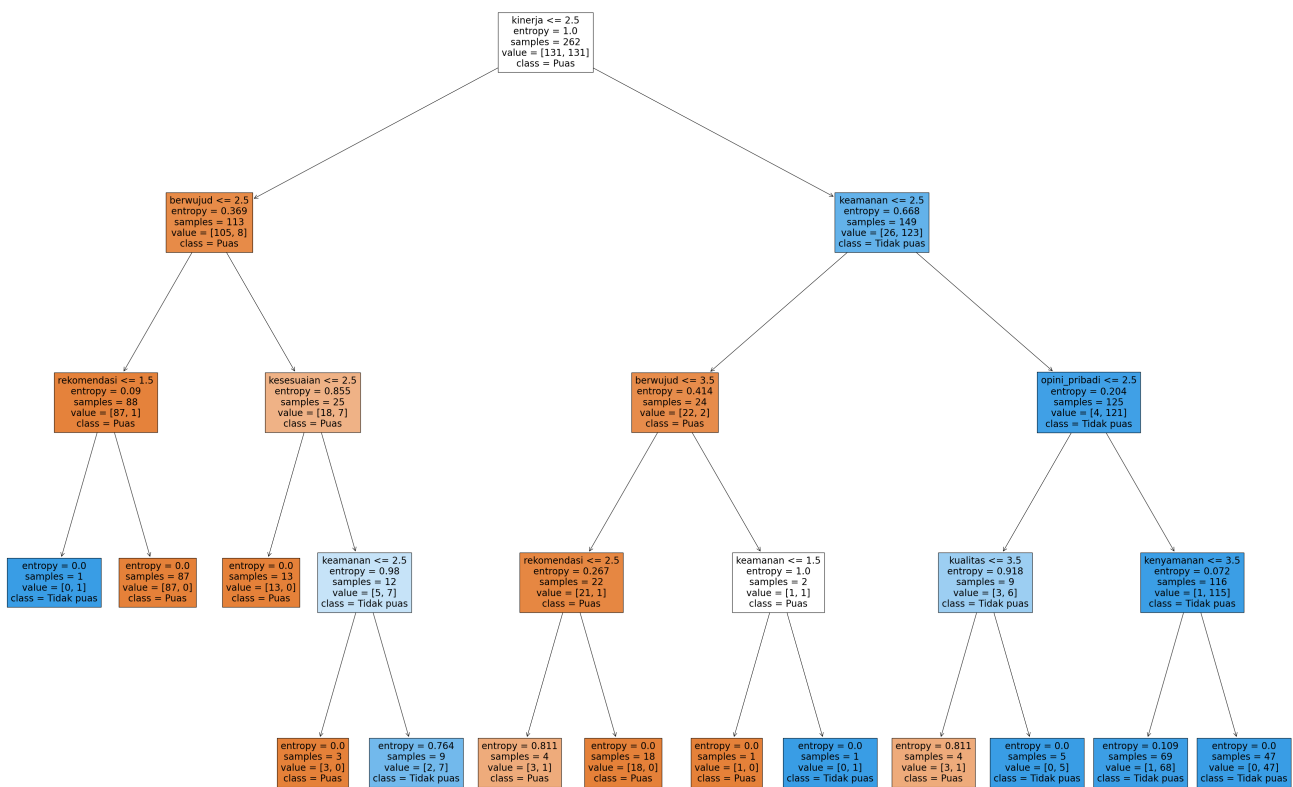
Tabel 3. Hasil pengujian

No	Data Latih (%)				Data Uji (%)			
	Akurasi	Presisi	Recall	F1-Score	Akurasi	Presisi	Recall	F1-Score
1.	98	98	98	98	93	97	94	95

3.7 Visualisasi Pohon Keputusan

Visualisasi pohon keputusan adalah hasil penjelasan bagaimana metode membuat keputusan berdasarkan atribut pada *dataset*.

Hasil pohon keputusan pada pada klasifikasi yang dilakukan dapat dilihat pada gambar 10.



Gambar 10. Visualisasi Pohon Keputusan

Merujuk pada visualisasi pohon keputusan yang ditampilkan pada Gambar 10, diperoleh sejumlah aturan keputusan sebagai berikut:

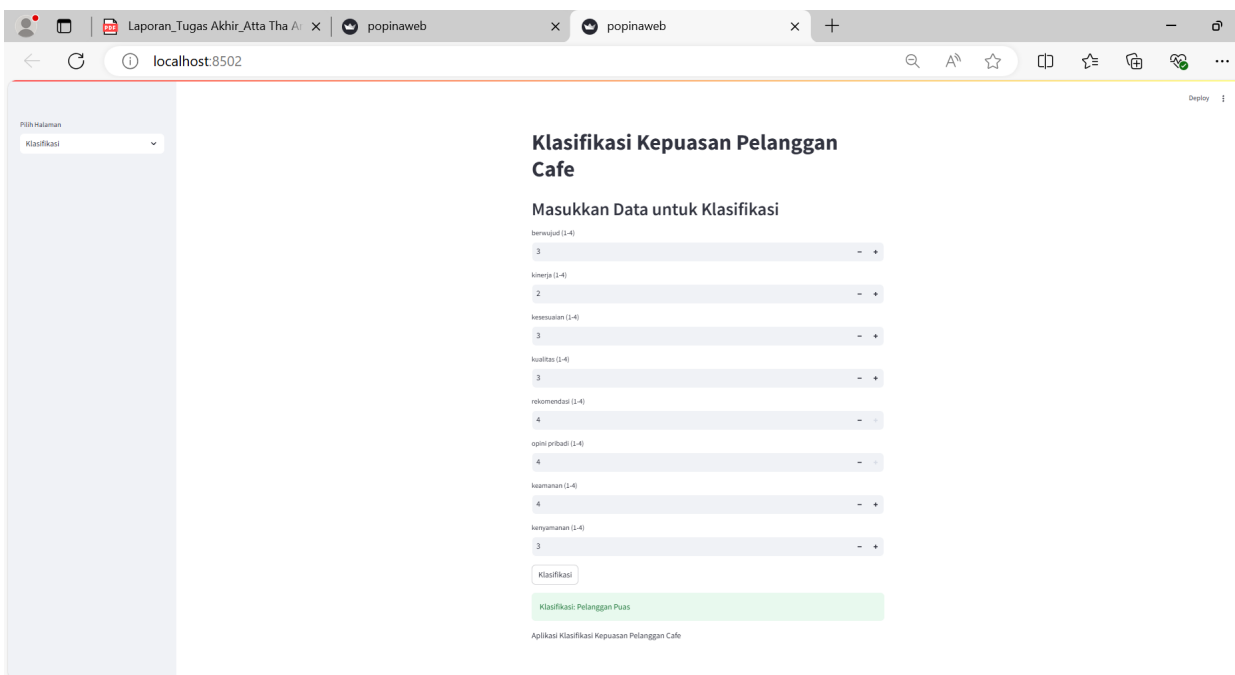
Jika nilai kinerja kurang dari atau sama dengan 2.50, kita lakukan pengecekan lebih lanjut:

- Jika nilai berwujud kurang dari atau sama dengan 2.50, kita periksa kembali rekomendasi:
 - Jika rekomendasi kurang dari atau sama dengan 1.50, hasil klasifikasi adalah 1 (puas).
 - Jika rekomendasi lebih besar dari 1.50, hasil klasifikasi adalah 0 (tidak puas).
- Jika nilai berwujud lebih besar dari 2.50, kita periksa kembali kesesuaian:
 - Jika kesesuaian kurang dari atau sama dengan 2.50, hasil klasifikasi adalah 0 (tidak puas).
 - Jika kesesuaian lebih besar dari 2.50, kita periksa kembali keamanan:
 - Jika keamanan kurang dari atau sama dengan 2.50, hasil klasifikasi adalah 0 (tidak puas).
 - Jika keamanan lebih besar dari 2.50, hasil klasifikasi adalah 1 (puas).
- Jika nilai kinerja lebih besar dari 2.50, kita lakukan pengecekan lebih lanjut:
 - Jika nilai keamanan kurang dari atau sama dengan 2.50, kita periksa kembali berwujud:
 - Jika berwujud kurang dari atau sama dengan 3.50, kita periksa kembali rekomendasi:
 - Jika rekomendasi kurang dari atau sama dengan 2.50, hasil klasifikasi adalah 0 (tidak puas).
 - Jika rekomendasi lebih besar dari 2.50, hasil klasifikasi adalah 0 (tidak puas).
 - Jika berwujud lebih besar dari 3.50, kita periksa kembali keamanan:
 - Jika keamanan kurang dari atau sama dengan 1.50, hasil klasifikasi adalah 0 (tidak puas).
 - Jika keamanan lebih besar dari 1.50, hasil klasifikasi adalah 1 (puas).
 - Jika nilai keamanan lebih besar dari 2.50, kita periksa kembali opini_pribadi:
 - Jika opini_pribadi kurang dari atau sama dengan 2.50, kita periksa kembali kualitas:
 - Jika kualitas kurang dari atau sama dengan 3.50, hasil klasifikasi adalah 0 (tidak puas).
 - Jika kualitas lebih besar dari 3.50, hasil klasifikasi adalah 1 (puas).
 - Jika opini_pribadi lebih besar dari 2.50, kita periksa kembali kenyamanan:
 - Jika kenyamanan kurang dari atau sama dengan 3.50, hasil klasifikasi adalah 1 (puas).
 - Jika kenyamanan lebih besar dari 3.50, hasil klasifikasi adalah 1 (puas).

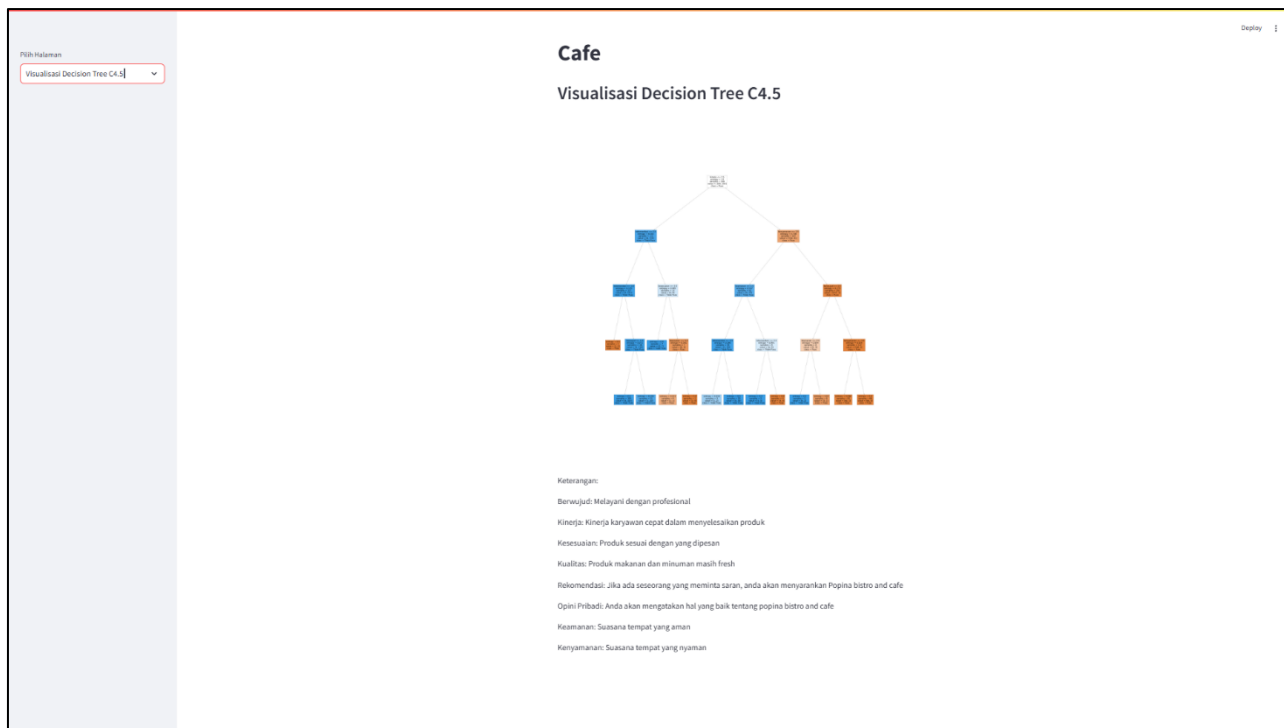
3.8 Hasil Aplikasi Streamlit

Model yang dibangun menggunakan metode decision tree C4.5 kemudian diinterpretasikan ke dalam bentuk web interaktif menggunakan Streamlit. Interpretasi tersebut disajikan secara visual dan interaktif melalui aplikasi web.

Dapat dilihat pada gambar 11. untuk tampilan halaman klasifikasi dan halaman visualisasi.



Gambar 11. Halaman Klasifikasi



Gambar 12. Tampilan Visualisasi

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan beberapa hal sesuai dengan rumusan masalah yang telah ditetapkan sebagai berikut:

1. Penelitian mengenai klasifikasi kepuasan pelanggan dengan metode decision tree C4.5 dilakukan melalui serangkaian tahapan mulai dari menangani *missing value*, mengecek *duplicate data*, melakukan *label encoding*, menangani *data imbalance* dengan menggunakan metode SMOTE, pemodelan *decision tree* C4.5, pengecekan akurasi, serta visualisasi aturan keputusan dengan pohon keputusan.
2. Hasil dari pengujian metode *decision tree* C4.5 serta tahapan model dievaluasi untuk mengetahui pengaruh setiap tahapan terhadap kemampuan model dalam melakukan klasifikasi. Metrik evaluasi yang digunakan berupa hasil dari *confusion matrix*. Dari evaluasi terhadap model menunjukkan bahwa model mampu melakukan klasifikasi dengan tingkat akurasi yaitu dengan begitu baik, dengan akurasi pada data latih 98% dan data uji 93%. Hasil *confusionmatrix* adalah *recall* 94%, *precision* 97%, dan *F1-score* 95%.

UCAPAN TERIMA KASIH

Saya mengucapkan rasa syukur kepada Allah SWT, serta terima kasih yang sebesar-besarnya kepada kedua orang tua, keluarga, dosen pembimbing, dan teman-teman atas segala dukungan dan motivasi yang telah diberikan hingga saya berhasil menyelesaikan skripsi dan penulisan jurnal ini.

DAFTAR PUSTAKA

- [1] S. Takalapeta, "Penerapan Data Mining Untuk Menganalisis Kepuasan Konsumen Menggunakan Metode Algoritma C4.5," *J I M P - J. Inform. Merdeka Pasuruan*, vol. 3, no. 3, pp. 34–38, 2018, doi: [10.37438/jimp.v3i3.186](https://doi.org/10.37438/jimp.v3i3.186).
- [2] M. Kalam Mollah and R. Prabowo, "Analisa Kepuasan Pelanggan dan Pengembangan Desain Proses Pelayanan pada Swalayan Pondok Pesantren Hidayatullah Surabaya," *Pros. SENASTITAN Semin. Nas. Teknol. Ind. Berkelanjutan*, vol. 1, no. 1, pp. 337–348, 2021, [Online].

Available: <https://ejournal.itats.ac.id/senastitan/article/view/1658>

- [3] M. D. Pamungkas, R. Ikha, and P. Sari, "Jurnal Optimasi Teknik Industri Analisis Kepuasan Pelanggan pada PT . Fairco Agung Kencana Menggunakan Metode Algoritma C4 . 5 dan Customer Satisfaction Index," pp. 1-6, 2024.
- [4] Yoga Religia and A. Amali, "Perbandingan Optimasi Feature Selection pada Naïve Bayes untuk Klasifikasi Kepuasan Airline Passenger," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 5, no. 3, pp. 527-533, 2021, doi: [10.29207/resti.v5i3.3086](https://doi.org/10.29207/resti.v5i3.3086).
- [5] S. F. Damanik, A. Wanto, and I. Gunawan, "Penerapan Algoritma DecisionTree C4.5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok," J. Krisnadana, vol. 1, no. 2, pp. 21-32, 2022, doi: [10.58982/krisnadana.v1i2.108](https://doi.org/10.58982/krisnadana.v1i2.108).
- [6] F. Fauziah, D. Hartama, and I. S. Damanik, "Analisa Kepuasan Pelanggan Menggunakan Klasifikasi Data Mining," Brahmana J. Penerapan Kecerdasan Buatan, vol. 2, no. 1, pp. 41-48, 2020, doi: [10.30645/brahmana.v2i1.47](https://doi.org/10.30645/brahmana.v2i1.47).
- [7] Y. Bastian *et al.*, "Analisis Penerapan Algoritma C4.5 Dalam Mengukur Tingkat Kepuasan Pelanggan Indihome Pada Kota Pematangsiantar," Januari, vol. 2, no. 1, pp. 62-69, 2021.
- [8] T. Widiastuti, K. Karsa, and C. Juliane, "Evaluasi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademik Menggunakan MetodeKlasifikasi Algoritma C4.5," Technomedia J., vol. 7, no. 3, pp. 364-380, 2022, doi: [10.33050/tmj.v7i3.1932](https://doi.org/10.33050/tmj.v7i3.1932).
- [9] Nurfitriyani, Islamiyah, and A. P. A. Masa, "Penerapan Klasifikasi Algoritma C4.5 Dan Algoritma C5.0 Untuk Mengetahui Tingkat Kepuasan Mahasiswa Terhadap Website Sistem Informasi Terpadu Layanan Program Studi (SIPLO)," J. Media Inform. Budidarma, vol. 7, pp. 1877-1887, 2023,doi: [10.30865/mib.v7i4.6433](https://doi.org/10.30865/mib.v7i4.6433).
- [10] P. Algoritma, C. Dalam, R. Sulistiani, D. Yusuf, and V. Terisia, "(STUDI KASUS : NILLA WEDDING GALLERY)," vol. 4, no. 2, pp. 89-100.
- [11] E. Prasetyaningrum and P. Susanti, "Analisa Tingkat Kepuasan Pelanggan Pada Percetakan Cv. Mega Media Menggunakan Algoritma C4.5," Sisfotenika, vol. 13, no. 1, pp. 65-75, 2023.
- [12] S. M. Sinaga, J. T. Hardinata, and M. Fauzan, "Implementasi Data Mining Clustering Tingkat Kepuasan Konsumen Terhadap Pelayanan Go-Jek," KSATRIA J. Penerapan Sist. Inf. (Komputer Manajemen), vol. 2, no. 2, pp.118-124, 2021.
- [13] S. Khaerunisa, T. Nur Padilah, and J. Haerul Jaman, "Implementasi Data Mining Menggunakan Metode Regresi Data Panel Untuk Memprediksi Capaian Indeks Pembangunan Manusia," JATI (Jurnal Mhs. Tek. Inform., vol. 7, no. 5, pp. 3399-3406, 2024, doi: [10.36040/jati.v7i5.7260](https://doi.org/10.36040/jati.v7i5.7260).