

## Prediksi Kepuasan Pelanggan Berdasarkan Ulasan Produk di Lazada Indonesia Menggunakan Algoritma Decision Tree C4.5

Wafiq Azizah Ramadhani<sup>1\*</sup>, Fahrur Rozi<sup>2</sup>

<sup>1,2</sup>Program Studi Pendidikan Teknologi Informasi, Universitas Bhinneka PGRI

\* wafiqazizah2411@gmail.com

### Abstrak

Perkembangan e-commerce di Indonesia terus meningkat seiring pesatnya penggunaan internet dan teknologi digital. Lazada, sebagai salah satu platform e-commerce terbesar di Indonesia, menyediakan sistem ulasan produk yang dapat mencerminkan tingkat kepuasan pelanggan. Namun, banyaknya ulasan menyulitkan penjual untuk mengevaluasi kepuasan secara manual. Penelitian ini bertujuan untuk memprediksi kepuasan pelanggan berdasarkan ulasan produk di Lazada Indonesia menggunakan algoritma Decision Tree C4.5. Penelitian ini bersifat kuantitatif dengan pendekatan data mining. Data ulasan diambil dari Kaggle dan melalui tahapan preprocessing, eksplorasi, pembangunan dan pengujian model. Atribut yang digunakan mencakup nama produk, kategori, harga, jumlah ulasan, dan rating, yang diklasifikasikan ke dalam tingkat kepuasan. Model dievaluasi menggunakan akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa algoritma C4.5 dapat memprediksi kepuasan pelanggan dengan akurasi 77,85%, presisi 76,74%, recall 83,54%, dan F1-score 80,00%. Harga produk menjadi atribut paling dominan dalam menentukan kepuasan. Temuan ini diharapkan dapat membantu penjual memahami faktor yang memengaruhi kepuasan pelanggan dan mendukung peningkatan layanan serta kualitas produk di e-commerce.

Kata kunci : Decision Tree C4.5, E-commerce, Lazada, Prediksi Kepuasan Pelanggan, Ulasan Produk

### Abstract

*The growth of e-commerce in Indonesia continues to rise with the increasing use of internet and digital technology. Lazada, one of the largest e-commerce platforms in the country, offers a product review system that reflects customer satisfaction. However, the large volume of reviews makes it difficult for sellers to manually evaluate satisfaction levels. This study aims to predict customer satisfaction based on product reviews on Lazada Indonesia using the C4.5 Decision Tree algorithm. The research is quantitative with a data mining approach. Review data was obtained from Kaggle and went through preprocessing, exploration, model building, and testing stages. The attributes used include product name, category, price, number of reviews, and average rating, which were classified into satisfaction levels. The model was evaluated using accuracy, precision, recall, and F1-score. Results show that the C4.5 algorithm can predict customer satisfaction with 77.85% accuracy, 76.74% precision, 83.54% recall, and an F1-score of 80.00%. Product price is the most influential attribute in determining satisfaction. These findings are expected to help sellers understand the key factors influencing customer satisfaction and support improvements in service and product quality on e-commerce platforms.*

Keywords : Decision Tree C4.5; E-commerce; Lazada; Prediction of Customer Satisfaction; Product Reviews.

### 1. Pendahuluan

Perkembangan teknologi informasi dan internet telah mengubah pola belanja konsumen, menjadikan e-commerce sebagai solusi populer karena kemudahan dan efisiensinya<sup>[1]</sup>. Kepuasan

pelanggan menjadi kunci sukses e-commerce karena berdampak pada loyalitas dan retensi. Ulasan produk, yang dianggap lebih tepercaya daripada iklan, mencerminkan persepsi pelanggan terhadap kualitas, harga, dan

pelayanan, serta sangat memengaruhi keputusan dan potensi pembelian ulang<sup>[2]</sup>. Lazada, sebagai salah satu platform e-commerce terbesar di Indonesia, mencatat jumlah kunjungan yang tinggi meskipun sempat mengalami penurunan pada kuartal 1 tahun 2021<sup>[3]</sup>. Meski demikian, tantangan seperti ketidaksesuaian produk, keterlambatan pengiriman, dan respons layanan pelanggan masih memengaruhi tingkat kepuasan pengguna. Prediksi kepuasan pelanggan dapat dilakukan melalui data mining menggunakan algoritma Decision Tree, khususnya C4.5, yang mampu menangani data kompleks dan menghasilkan model yang mudah dipahami. Data mining adalah metode pengolahan data yang bertujuan untuk mengungkap pola-pola tersembunyi di dalam data<sup>[4]</sup>. Atribut seperti harga produk, jumlah ulasan, dan rating digunakan sebagai indikator utama. Harga sering menjadi penentu keputusan pembelian, sementara jumlah ulasan dan rating mencerminkan tingkat kepercayaan dan kualitas produk. Beberapa penelitian sebelumnya mendukung penggunaan algoritma decision tree C4.5 dalam memprediksi kepuasan pelanggan. Prasetyaningrum & Susanti (2023) meneliti kepuasan pelanggan pada percetakan CV. Mega Media dan memperoleh akurasi sebesar 93%<sup>[5]</sup>. Sikumbang et al. (2022) menerapkan algoritma yang sama pada kepuasan pelanggan kartu Telkomsel

prabayar dan mencapai akurasi 96,55%<sup>[6]</sup>. Penelitian oleh Haekal et al. (2021) mengenai pengguna aplikasi Shopee menunjukkan akurasi 97%<sup>[7]</sup>. Selain itu, Putri & Arnomo (2020) meneliti ketidakpuasan pelanggan Hinet Batam dan menyimpulkan bahwa kualitas pelayanan yang rendah berdampak negatif terhadap kepuasan, juga menggunakan algoritma C4.5<sup>[8]</sup>.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk memprediksi kepuasan pelanggan Lazada Indonesia berdasarkan ulasan produk menggunakan algoritma Decision Tree C4.5. Data ulasan diambil dari platform Kaggle, dengan atribut seperti harga, jumlah ulasan, dan rating sebagai variabel utama. Penelitian ini diharapkan dapat membantu platform e-commerce seperti Lazada dalam memahami faktor-faktor yang memengaruhi kepuasan pelanggan dan meningkatkan strategi pelayanan agar tetap kompetitif di pasar digital yang terus berkembang

## **2. Tinjauan Pustaka**

### **2.1. Penelitian Terkait**

Penulis mengutip beberapa penelitian sebelumnya sebagai acuan untuk memperdalam dan memperkuat bahan kajian.

- Aditya Nugroho dan Kristiana (2022) Menerapkan algoritma C4.5 pada data kepuasan pelanggan toko online Parfume Chantik dengan 85 data sampel dari 591

pelanggan, dan memperoleh akurasi sebesar 71% [9].

- Budi et al. (2024) Menggunakan data kuisisioner dari 151 responden untuk menganalisis kepuasan terhadap layanan aplikasi e-commerce secara umum, tanpa menyebutkan nama e-commerce secara spesifik, dan memperoleh akurasi sebesar 93% [10].
- Steffany (2023) Meneliti kepuasan pelanggan Shopee dengan 120 responden dan berbagai atribut. Akurasi mencapai 94,17% [11].
- Soewardy dan Bororing (2022) Menganalisis kepuasan pengguna Tokopedia dengan 114 responden. Akurasi sebesar 94,77% [12].
- Safitri (2024) Meneliti aplikasi layanan publik Depok Single Window dengan 35 responden. Akurasi yang didapatkan adalah 85% [13].

## 2.2. Landasan Teori

### 1. Kepuasan Pelanggan

Kepuasan pelanggan adalah perasaan positif yang muncul saat harapan terhadap produk atau jasa terpenuhi atau melampaui ekspektasi [14]. Dalam e-commerce, kepuasan menjadi kunci mempertahankan keunggulan, memperluas pasar, dan menjaga keberlanjutan bisnis. Pelanggan mengharapkan kemudahan, pengiriman cepat, kualitas produk, dan layanan responsif. Oleh karena itu, perusahaan berfokus pada peningkatan pengalaman belanja melalui

umpan balik. Kepuasan yang terjaga mendorong pembelian ulang, rekomendasi, dan loyalitas pelanggan.

### 2. Decision Tree C4.5

Decision Tree C4.5 adalah metode machine learning yang digunakan untuk klasifikasi dan regresi, yang berfungsi untuk membagi dataset menjadi subset yang lebih kecil berdasarkan fitur yang paling signifikan. Decision tree adalah sistem buatan yang meniru cara manusia mengambil keputusan dalam menyelesaikan suatu masalah [15]. Pemilihan atribut dilakukan menggunakan Entropy dan Gain Information yang digunakan untuk mengukur ketidakpastian dalam data. Entropy mengukur tingkat ketidakpastian atau keberagaman, sedangkan Gain Information mengukur seberapa banyak informasi yang diperoleh dari pemisahan data berdasarkan atribut tertentu [16].

#### a Rumus Entropy

Rumus untuk menghitung entropy pada algoritma C4.5 sebagai berikut.

$$Entropy(S) = \sum_{i=1}^k -p_i \times \log_2 p_i$$

Keterangan :

- S adalah Himpunan (dataset) kasus.
- k adalah banyaknya partisi S.
- $p_i$  adalah probabilitas yang didapatkan dari Sum (Ya) atau Sum (Tidak) dibagi total kasus.

## b. Rumus Gain

Rumus dari Gain adalah sebagai berikut.

$Gain(A) =$

$$Entropy(S) - \sum_{i=1}^k \frac{S_i}{S} \times Entropy(S_i)$$

Keterangan :

- A adalah atribut yang dipilih
- $S_i$  adalah ukuran atau jumlah elemen dalam subset
- S adalah ukuran atau jumlah total elemen dalam dataset

## 3. Lazada Indonesia

Lazada Indonesia merupakan salah satu platform e-commerce terbesar dan terkemuka di Indonesia, yang pertama kali didirikan pada tahun 2012. Platform ini merupakan bagian dari Lazada Group, sebuah perusahaan e-commerce multinasional yang beroperasi di berbagai negara di kawasan Asia Tenggara, termasuk Malaysia, Singapura, Filipina, Thailand, dan Vietnam. Lazada telah berhasil menciptakan pengalaman belanja daring yang mudah dan aman bagi konsumennya, menjadikannya salah satu pilihan utama bagi masyarakat Indonesia dalam memenuhi kebutuhan mereka.

## 4. Atribut Kepuasan Pelanggan

Berikut adalah beberapa atribut utama kepuasan pelanggan:

- a. Nama Produk : Nama yang spesifik dapat membentuk ekspektasi dan meningkatkan kepuasan jika sesuai harapan.

- b. Kategori Produk : Kepuasan dipengaruhi oleh kemampuan produk dalam kategori tersebut memenuhi kebutuhan pelanggan.

- c. Harga Produk : Harga yang wajar dan sebanding dengan kualitas cenderung meningkatkan kepuasan..

- d. Jumlah Total Ulasan :Banyaknya ulasan mencerminkan kepercayaan dan pengalaman pelanggan lain, sehingga mendorong kepuasan.

- e. Rata-Rata Nilai Produk (Rating)

Rating tinggi menunjukkan kualitas baik dan meningkatkan potensi kepuasan. Penelitian ini menggunakan rating sebagai indikator kepuasan dengan klasifikasi "Puas" dan "Tidak Puas" untuk analisis yang lebih jelas.

## 3. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk menganalisis ulasan pelanggan dan memprediksi tingkat kepuasan terhadap produk di platform Lazada Indonesia. Penelitian ini termasuk dalam kategori predictive analysis karena bertujuan membangun model klasifikasi menggunakan algoritma Decision Tree C4.5 untuk memetakan pola-pola dalam data yang berpengaruh terhadap tingkat kepuasan pelanggan.

### 3.1. Lokasi dan Sumber Data

Penelitian ini tidak dilakukan secara langsung di lokasi fisik tertentu, tetapi bersumber pada data

sekunder yang diperoleh secara daring dari situs Kaggle ([www.kaggle.com](http://www.kaggle.com)). Dataset yang digunakan berisi informasi mengenai ulasan produk di Lazada Indonesia, termasuk atribut nama produk, kategori, harga, jumlah total ulasan, dan rating produk. Pemanfaatan sumber data terbuka ini bertujuan untuk memperoleh dataset yang kaya dan bervariasi untuk keperluan klasifikasi.

### 3.2. Teknik Pengumpulan Data

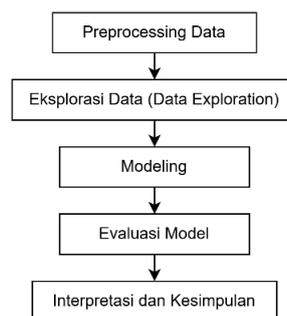
Pengumpulan data dilakukan dengan mengakses dan mengunduh dataset dari platform Kaggle. Proses ini dimulai dengan pencarian dataset menggunakan kata kunci seperti “Lazada,” “product reviews,” dan “customer satisfaction.” Dataset dipilih berdasarkan kriteria relevansi terhadap topik penelitian, kelengkapan atribut, dan kesesuaian format data. Setelah dataset diperoleh, dilakukan proses data cleaning yang mencakup:

- Penghapusan data duplikat,
- Penanganan nilai kosong,
- Seleksi atribut yang relevan untuk proses klasifikasi.
- Atribut rating digunakan sebagai variabel dependen (kepuasan pelanggan), sementara atribut seperti nama produk, kategori, harga, dan jumlah ulasan digunakan sebagai variabel independen.

### 3.3. Teknik Analisis Data

Teknik analisis data digunakan untuk

mengevaluasi performa model Decision Tree yang dibangun dalam memprediksi kepuasan pelanggan. Proses analisis data dilakukan melalui tahapan sesuai dengan gambar 1 berikut.



Gambar 1. Teknik Analisis Data

Preprocessing mencakup pembersihan data, konversi variabel numerik ke kategorik (misalnya harga menjadi murah, sedang, mahal), dan pengelompokan sesuai logika klasifikasi. Pada tahap eksplorasi data, dilakukan analisis deskriptif untuk memahami distribusi dan karakteristik data seperti nama produk, harga, jumlah ulasan, dan rating. Modeling menggunakan algoritma Decision Tree C4.5 untuk membentuk model klasifikasi yang interpretatif dan efisien. Evaluasi dilakukan dengan metrik akurasi, presisi, recall, dan F1-score. Hasil klasifikasi dianalisis untuk mengidentifikasi pola dan menyusun rekomendasi strategis guna meningkatkan layanan atau produk berdasarkan ulasan pelanggan. Atribut dalam penelitian ini diklasifikasikan untuk mendukung proses analisis prediktif. Nama produk dianggap spesifik jika mencantumkan merek dan spesifikasi secara jelas, dan tidak spesifik jika hanya

mencantumkan salah satunya [17]. Harga produk diklasifikasikan menjadi tiga kategori berdasarkan rentang nilai dalam data: murah (<Rp500.000), sedang (Rp500.000–Rp5.000.000), dan mahal (>Rp5.000.000) [18]. Jumlah total ulasan dibagi menjadi tiga kategori: sedikit (<100), sedang (100–1000), dan banyak (>1000), menggunakan pendekatan nilai minimum, tengah, dan maksimum dari data [19]. Sementara itu, kepuasan pelanggan diklasifikasikan berdasarkan rata-rata rating dengan skala Likert: tidak puas (rating 1–3) dan puas (rating 4–5), untuk memudahkan pemodelan [20]

#### 4. Hasil dan Pembahasan

##### 4.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah ulasan pelanggan dari Lazada Indonesia yang tersedia di platform Kaggle. Dataset berisi 1000 dari 4000 data ulasan. Proses pengumpulan data melibatkan pengunduhan dataset, pemahaman struktur data, serta pengkategorian data. Berikut adalah tabel 5 yang berisi 1000 data dari platform Kaggle.

Tabel 1. Data Keseluruhan

No	Nama	K	Harga (Rp)	JTU	RN	TK
1	SSD SAMSUNG 850 EVO 250GB	harddisk-eksternal	1.214.000	1	1	Tidak Puas
2	Coccaa LED TV 24 Inch - 24 jabodetabek	televisi-digital	1.125.000	8	4	Puas

3	Coccaa LED TV Hitam	smart-tv	1.475.000	2	5	Puas
100	Samsung Micro USB Flashdisk - [8 GB]	flash-drives	30.550	4	2	Tidak Puas

Keterangan Tabel:

K: Kategori

JTU : Jumlah Total Ulasan

RN : Rata-Rata Nilai

TK : Tingkat Kepuasan

##### 4.2. Preprocessing Data (Cleaning & Preparation)

Tahap preprocessing dilakukan untuk memastikan data relevan dan siap diolah. Proses cleaning mencakup penghapusan duplikasi (data ke-163, 208, 249, 389, 412, dan 751) serta penanganan missing values. Atribut rating juga dikonversi pada tahap ini. Dari 1000 data acak, data yang telah dibersihkan kemudian diklasifikasikan sesuai ketentuan pada tabel 2 hingga 5. Data akhir setelah cleaning ditampilkan pada tabel 6.

Tabel 2. Preprocessing Data (Cleaning & Preparation)

No	Nama	K	Harga	JTU	TK
1.	Spesifik	harddisk-eksternal	Sedang	Sedikit	Tidak Puas
2.	Spesifik	televisi-digital	Sedang	Sedikit	Puas
3.	Spesifik	smart-tv	Sedang	Sedikit	Puas
996.	Spesifik	flash-drives	Murah	Sedikit	Tidak Puas

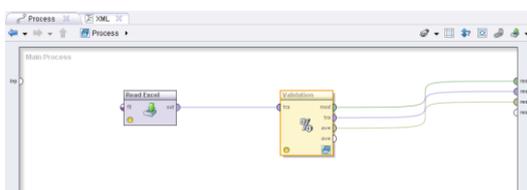
##### 4.3. Eksplorasi Data

Dataset penelitian ini berisi 994 ulasan produk teknologi yang dikategorikan berdasarkan nama produk, jenis produk, harga, dan jumlah ulasan.

Sebanyak 539 pelanggan merasa puas dan 455 tidak puas. Sebagian besar produk (841) memiliki nama spesifik, sementara 153 tidak. Jenis produk didominasi flash-drives (465), diikuti laptop (276), harddisk-eksternal (101), televisi-digital (88), dan smart-TV (64). Untuk harga, 510 ulasan berasal dari produk murah, 354 dari harga sedang, dan 130 dari harga mahal.

#### 4.4. Modelling (Decision Tree)

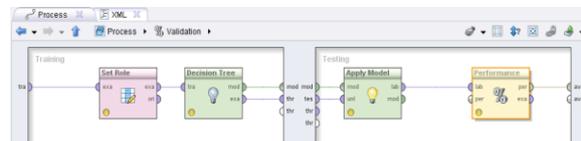
Data hasil preprocessing dimodelkan menggunakan algoritma Decision Tree C4.5 untuk menghasilkan pohon keputusan dan mengukur akurasi, presisi, recall, serta F1-score. Dari 994 data, 80% digunakan untuk pelatihan dan 20% untuk pengujian, sesuai standar agar model tidak overfitting [17]. Modeling dilakukan di RapidMiner dengan mengimpor data .xlsx menggunakan operator Read Excel, menetapkan atribut kepuasan sebagai label, dan menambahkan X-Validation untuk membagi data secara bergantian. Hubungan antar operator ditampilkan pada Gambar 2.



Gambar 2. Proses menghubungkan port

Langkah berikutnya adalah menambahkan operator Set Role untuk menetapkan label, Decision Tree untuk membangun model, Apply Model untuk pengujian, dan Performance untuk

evaluasi. Pada proses training, data dari port tra diarahkan ke Set Role lalu ke Decision Tree, dan model dikirim ke Apply Model melalui port mod. Pada tahap testing, data uji masuk ke Apply Model melalui port tes, hasil prediksi diteruskan ke Performance, lalu hasil evaluasi dikirim ke X-Validation melalui port per. Rangkaian proses ini ditunjukkan pada Gambar 3.



Gambar 3. Proses menghubungkan port antar operator

Setelah proses dijalankan, diperoleh nilai akurasi, presisi, dan recall yang menggambarkan kinerja model. Akurasi menunjukkan ketepatan model secara keseluruhan, presisi menunjukkan seberapa tepat prediksi positif, dan recall mengukur kemampuan model mendeteksi semua data positif. Hasil ini menjadi dasar awal untuk mengevaluasi apakah model sudah memenuhi tujuan penelitian atau perlu ditingkatkan. Nilai-nilai tersebut ditampilkan pada Gambar 4.

Metric	Value
Accuracy	71.54%
Precision	79.11%
Recall	71.54%
F1-Score	75.27%

Gambar 4. Nilai Akurasi, Presisi, Recall

Hasil pada Gambar 4 menunjukkan bahwa model Decision Tree memiliki akurasi 71,54%, artinya mayoritas data berhasil diprediksi dengan benar. Presisi untuk kelas “Puas” mencapai 79,11%, menandakan sebagian besar prediksi positif

sesuai dengan data aktual. Namun, recall sebesar 72,95% menunjukkan masih ada data positif yang belum terdeteksi oleh model.

Sedangkan untuk mengetahui nilai F1-score-nya, gunakan rumus dibawah ini untuk menghitungnya.

$$\begin{aligned}
 F1\ Score &= 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \\
 &= 2 \times \frac{79,11 \times 72,95}{79,11 + 72,95} \\
 &= 75,91\%
 \end{aligned}$$

Kombinasi hasil ini menunjukkan bahwa model memiliki performa yang baik secara keseluruhan, tetapi dapat ditingkatkan lebih lanjut terutama dalam kemampuan mendeteksi data positif secara menyeluruh. Adapun untuk hasil pohon keputusan terdapat pada gambar 5 berikut.



Gambar 5. Pohon Keputusan

Keterangan warna:

- Merah : Puas
- Biru : Tidak Puas

Adapun rule untuk pohon keputusan pada gambar 5 adalah sebagai berikut:

- a. Jika Harga Produk lebih dari Rp500.000 yang mana termasuk klasifikasi “Mahal” maka akan menghasilkan Tingkat Kepuasan “Puas”.
- b. Jika Harga Produk kurang dari Rp500.000 yang mana termasuk klasifikasi “Murah” harus

memperhatikan ketentuan berikut:

- Jika Jumlah Total Ulasan kurang dari 100 masuk dalam klasifikasi “Sedikit” maka akan menghasilkan Kepuasan Pelanggan “Tidak Puas”.
  - Jika Jumlah Total Ulasan berada di antara 100-1.000 masuk dalam klasifikasi “Sedang” maka akan menghasilkan Kepuasan Pelanggan “Puas”.
  - Jika Jumlah Total Ulasan lebih dari 1.000 masuk dalam klasifikasi “Banyak” maka akan menghasilkan Kepuasan Pelanggan “Puas”.
- c. Jika Harga Produk berada di antara Rp500.000,00 - Rp5.000.000 yang mana termasuk klasifikasi “Sedang” maka akan menghasilkan Kepuasan Pelanggan “Puas”.

Berdasarkan pohon keputusan pada Gambar 5, atribut harga menjadi faktor utama dalam menentukan kepuasan pelanggan. Produk dengan harga mahal atau sedang cenderung menghasilkan kepuasan lebih tinggi, memperkuat anggapan bahwa kualitas sebanding dengan harga<sup>[20]</sup>. Sementara itu, atribut nama dan kategori tidak muncul dalam pohon karena komposisi datanya tidak seimbang, sehingga kontribusi informasinya rendah dalam pemilihan atribut terbaik.

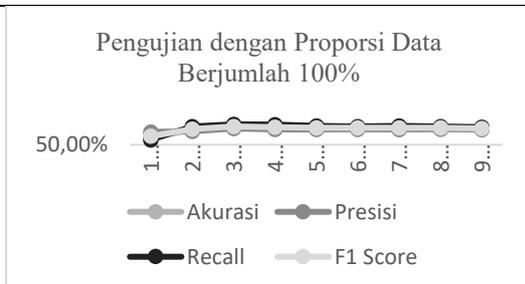
#### 4.5. Pengujian Data

Pengujian dilakukan dengan membagi data uji ke dalam proporsi 10% hingga 90% untuk

menganalisis pengaruh ukuran data uji terhadap performa model Decision Tree. Setiap proporsi diuji secara independen, sementara sisanya digunakan sebagai data latih. Pengujian ini menggunakan operator Split Validation untuk membagi data, mencegah overfitting, dan meningkatkan generalisasi model. Hasil evaluasi dari tiap proporsi dibandingkan untuk menentukan kombinasi terbaik antara akurasi, presisi, recall, dan F1-score. Detail hasil pengujian disajikan pada Tabel 3 dan grafiknya pada Gambar 6.

Tabel 3. Pengujian dengan Proporsi Data Berjumlah 100%

Data Uji	Data Latih	Akurasi	Presisi	Recall	F1-Score
10%	90%	63,64%	71,11%	58,18%	64,00%
20%	80%	73,37%	72,17%	79,81%	75,80%
30%	70%	77,85%	76,74%	83,54%	80,00%
40%	60%	76,63%	75,43%	82,94%	79,01%
50%	50%	75,65%	75,17%	81,13%	78,04%
60%	40%	75,34%	75,59%	80,06%	77,76%
70%	30%	75,43%	75,25%	80,70%	77,88%
80%	20%	75,47%	75,61%	80,05%	77,77%
90%	10%	74,75%	75,25%	78,71%	76,94%



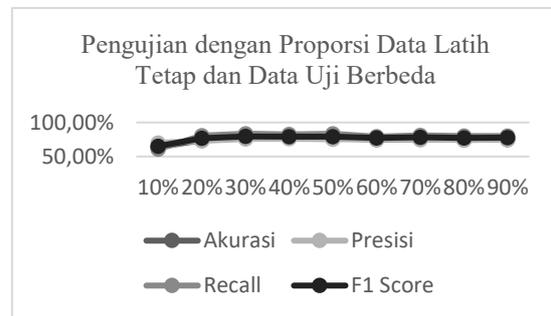
Gambar 6. Pengujian dengan Proporsi Data Berjumlah 100%

Tabel 3 dan Gambar 6 menunjukkan bahwa

akurasi, presisi, recall, dan F1-score meningkat seiring bertambahnya proporsi data uji. Proporsi terbaik ditemukan pada 70% data latih dan 30% data uji, ditunjukkan oleh nilai F1-score tertinggi. Nilai F1-score yang tinggi mencerminkan keseimbangan precision dan recall, yang penting dalam klasifikasi dengan distribusi data tidak seimbang<sup>[12]</sup>. Pengujian lanjutan mempertahankan 70% data latih dan memvariasikan proporsi data uji untuk menganalisis pengaruhnya terhadap performa model. Hasilnya ditampilkan pada Tabel 4 dan Gambar 7.

Tabel 4. Pengujian dengan Proporsi Data Latih Tetap Data Uji Berbeda

Data Uji	Akurasi	Presisi	Recall	F1 Score
10%	63,71%	70,00%	60,87%	65,12%
20%	73,78%	73,28%	80,67%	76,80%
30%	77,85%	75,74%	83,54%	79,45%
40%	77,29%	76,21%	82,63%	79,29%
50%	76,81%	75,21%	83,49%	79,13%
60%	74,73%	75,48%	79,12%	77,26%
70%	75,65%	75,17%	81,13%	78,04%
80%	74,53%	74,27%	80,28%	77,16%
90%	74,78%	75,08%	80,26%	77,58%



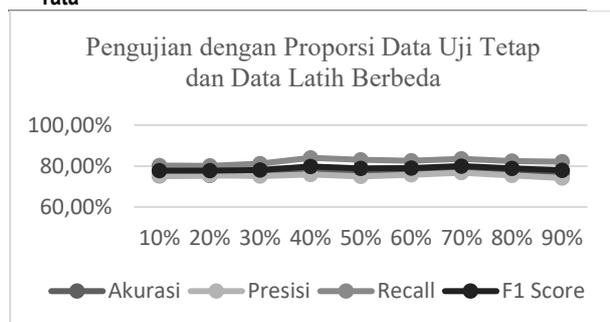
Gambar 7. Pengujian dengan Proporsi Data Latih Tetap Data Uji Berbeda

Tabel 4 dan Gambar 7 menunjukkan peningkatan signifikan performa model dari 10% ke 20% data

uji, dengan akurasi naik dari 63,71% menjadi 73,78% dan recall dari 60,87% ke 80,67%. Pada proporsi 30%–50%, metrik stabil di kisaran akurasi 76%–78% dan recall mencapai 83,49% pada 50% data uji. Setelah 50%, peningkatan performa cenderung melambat. Secara keseluruhan, model menunjukkan keseimbangan antara akurasi dan recall, dengan recall tertinggi 79,11%. Hasil terbaik terdapat pada proporsi 30% data uji, dengan F1-score tertinggi sebesar 79,45%. Selanjutnya, pengujian dilakukan dengan variasi proporsi data latih, sementara data uji tetap 30%, untuk menganalisis pengaruh ukuran data latih terhadap performa model. Hasilnya ditampilkan pada Tabel 9 dan Gambar 8.

Tabel 5. Pengujian dengan Proporsi Data Uji Tetap Data Latih Berbeda

Data Latih	Akurasi	Presisi	Recall	F1-Score
10%	75,30%	75,35%	80,25%	77,72%
20%	75,34%	75,59%	80,06%	77,76%
30%	75,65%	75,17%	81,13%	78,04%
40%	77,46%	75,90%	84,00%	79,74%
50%	76,68%	75,00%	83,08%	78,83%
60%	77,04%	75,66%	82,66%	79,01%
70%	77,85%	76,74%	83,54%	80,00%
80%	76,75%	75,48%	82,39%	78,78%
90%	75,81%	74,13%	82,17%	77,94%
Rata-rata	76,43%	75,45%	82,14%	78,65%



Gambar 8. Pengujian dengan Proporsi Data Uji Tetap Data Latih Berbeda

Tabel 5 dan Gambar 8 menunjukkan bahwa dengan data uji tetap 30% dan variasi data latih 10%–90%, performa model relatif stabil. Akurasi berkisar 75,30%–77,94%, presisi 75,17%–76,74%, recall 80,06%–84,00%, dan F1-score 77,72%–80,00%. Rata-rata keseluruhan menunjukkan performa yang baik dengan akurasi 76,43% dan F1-score 78,65%. Hasil terbaik dicapai pada 70% data latih dengan F1-score tertinggi (80,00%), menunjukkan bahwa proporsi 70:30 adalah konfigurasi optimal untuk akurasi dan efektivitas model.

#### 4.6. Analisis Hasil

Berdasarkan hasil pengujian, algoritma Decision Tree C4.5 efektif dalam menganalisis kepuasan pelanggan dan menentukan faktor yang memengaruhinya. Model awal menghasilkan akurasi 71,54%, presisi 79,11%, recall 72,95%, dan F1-score 75,91%, dengan harga sebagai atribut paling berpengaruh. Produk mahal dan sedang cenderung memuaskan, sedangkan produk murah lebih dipengaruhi oleh jumlah ulasan—semakin sedikit ulasan, kepuasan cenderung rendah. Proporsi data latih 70% dan data uji 30% memberikan hasil terbaik dengan akurasi 77,85%, presisi 76,74%, recall 83,54%, dan F1-score 80,00%, menegaskan pentingnya proporsi data dalam meningkatkan performa model. Temuan ini sejalan dengan penelitian

Hasibuan & Mahdiana (2023) dengan akurasi 75,52% dalam prediksi kelulusan [19], Han et al. (2019) dalam survei kepuasan pelanggan dengan akurasi 78,26% [21], serta Hamdan et al. (2023) yang mencatat akurasi 71,44% untuk prediksi loyalitas pelanggan hotel [22], menunjukkan bahwa Decision Tree memberikan performa yang cukup baik dan dapat diandalkan dalam konteks prediksi kepuasan pelanggan

## 5. Kesimpulan

Penelitian ini menyimpulkan bahwa algoritma Decision Tree C4.5 efektif memprediksi kepuasan pelanggan berdasarkan ulasan produk di Lazada, dengan akurasi 77,85%. Model dibangun melalui tahapan pengumpulan data, preprocessing, dan pembentukan pohon keputusan dengan kategori kepuasan sebagai target. Evaluasi menunjukkan kinerja klasifikasi yang baik. Untuk pengembangan, disarankan menggunakan algoritma lain seperti Random Forest atau SVM, menerapkan pendekatan multikategori, memanfaatkan data real-time, memperluas sumber data, dan mengintegrasikan model ke dalam sistem rekomendasi produk.

## 6. Daftar Pustaka

- [1] Nursania Dasopang, "Jiemas E - Commerce Bisnis Dan Internet," *J. Ilm. Ekon. Manaj. dan Syariah JIEMAS*, vol. 2, pp. 129–135, 2023, [Online]. Available: <https://jiemas.stai-dq.org/index.php/home>
- [2] D. S. Nur and A. Octavia, "Pengaruh Electronic Word of Mouth Terhadap Keputusan Pembelian Dengan Kepercayaan Konsumen Sebagai Mediasi Pada Marketplace Shopee Di Kota Jambi," *J. Manaj. Terap. dan Keuang.*, vol. 11, no. 2, pp. 387–399, 2022, doi: 10.22437/jmk.v11i2.17960.
- [3] F. Baisyir and D. A. Prasetyo, "Analisis Perbedaan Kepuasan Konsumen Pada Pengguna E-Commerce Lazada Di Wilayah Jakarta Timur," *ProBank*, vol. 7, no. 2, pp. 185–196, 2023, doi: 10.36587/probank.v7i2.1341.
- [4] E. Kurniah, Reni; Putra, Dadang Yunika Surya; Diana, "Penerapan Data Mining Decision Tree Algoritma C4.5 Untuk Mengetahui Tingkat Kepuasan Mahasiswa Terhadap Layanan Akademik Dan Kemahasiswaan(Studi Kasus Universitas.Prof.Dr. Hazairin,SH)," *Infotek J. Inform. dan Teknol.*, vol. 5, no. 2, pp. 316–326, 2022, doi: <https://doi.org/10.29408/jit.v5i2.5910>.
- [5] 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.
- [6] E. D. Sikumbang, F. Ariani, T. Handayani, and K. Ramanda, "Penerapan Algoritma C4.5 Untuk Menentukan Tingkat Kepuasan Pelanggan Kartu Telkomsel Prabayar," *J. Sains Komput. Inform.*, vol. 6, no. September, pp. 811–820, 2022.
- [7] B. V. Haekal, I. Ernawati, and N. Chamidah, "Klasifikasi Kepuasan Pengguna Layanan Aplikasi Shopee Menggunakan Metode Decision Tree C4.5," *Inform. J. Ilmu Komput.*, vol. 17, no. 3, p. 188, 2021, doi: 10.52958/iftk.v17i3.3648.
- [8] S. M. Putri and S. A. Arnomo, "Penerapan Algoritma C4.5 Untuk Prediksi Kualitas Pelayanan Terhadap Kepuasan Konsumen (Studi Kasus: Hinet Batam)," *J. Inf. Syst. Res.*, vol. 1, no. 2, pp. 70–76, 2020, [Online].

- Available: <http://ejournal.seminar-id.com/index.php/josh/article/view/69>
- [9] C. R. Aditya Nugroho and T. Kristiana, "Penerapan Algoritma C4.5 Untuk Kepuasan Pelanggan Toko Online Parfume Chantik," *J. Algoritma*, vol. 3, no. 1, pp. 10–21, 2022, doi: 10.35957/algoritma.v3i1.3169.
- [10] E. S. Budi, A. R. Kadafi, Y. Kharismawan, R. Fadillah, and D. S. Putri, "Analisa Kepuasan Pelanggan Terhadap Layanan Aplikasi E-Commerce," *RESOLUSI Rekayasa Tek. Inform. dan Inf.*, vol. 4, no. 6, pp. 530–542, 2024.
- [11] 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.
- [12] K. Soewardy and G. M. G. Bororing, "Analisis Tingkat Kepuasan Pengguna Aplikasi Marketplace Tokopedia Menggunakan Algoritma C4.5," *J. Inform. dan Bisnis*, vol. 11, no. 1, pp. 38–47, 2022, doi: 10.46806/jib.v11i1.883.
- [13] R. A. Safitri, Y. Tajul Arifin, and M. Dhea Kirana, "Penerapan Algoritma C4.5 Untuk Klasifikasi Kepuasan Layanan Aplikasi Depok Single Window," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, pp. 4094–4102, 2024, doi: 10.36040/jati.v8i3.9865.
- [14] J. Juliana and K. Keni, "Prediksi Kualitas Pelayanan Dan Kepercayaan Terhadap Loyalitas Pelanggan: Kepuasan Pelanggan Sebagai Variabel Mediasi," *J. Manajerial Dan Kewirausahaan*, vol. 2, no. 1, p. 110, 2020, doi: 10.24912/jmk.v2i1.7451.
- [15] M. Hamonangan Nasution and M. Iqbal Sofyan, "Penerapan Metode Decision Tree Untuk Memprediksi Prestasi Siswa Kelas XII Dilihat dari Nilai Akhir Semester di SMK Negeri 1 Selong Tahun Pelajaran 2017/2018," *Infotek J. Inform. dan Teknol.*, vol. 3, no. 1, pp. 58–65, 2020, doi: 10.29408/jit.v3i1.1825.
- [16] A. B. Yamantri and A. Ahmad, "Penerapan Algoritma C4 . 5 Untuk Prediksi Faktor Risiko Obesitas Pada Penduduk Dewasa," *J. Komput. Antart.*, vol. 2, no. c, pp. 118–125, 2024.
- [17] K. Wardiono and I. Firmansyah, "Hak Atas Informasiterhadap Perlindungan Hukumbagi Konsumen Dalam Jual Belionline Melalui Lazada," *Justisi*, vol. 10, no. 1, pp. 1–19, 2023, doi: 10.33506/jurnaljustisi.v10i1.2255.
- [18] N. F. Andriani and A. H. Setiawan, "Analisis Preferensi Konsumen Terhadap Penggunaan Produk Skincare Korea Selatan Dan Lokal," *Diponegoro J. Econ.*, vol. 9, no. 4, pp. 1–8, 2020, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/jme/article/view/29050>
- [19] T. H. Hasibuan and D. Mahdiana, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta," *Skanika*, vol. 6, no. 1, pp. 61–74, 2023, doi: 10.36080/skanika.v6i1.2976.
- [20] T. Tundo and S. 'Uyun, "Konsep Decision Tree Reptree untuk Melakukan Optimasi Rule dalam Fuzzy Inference System Tsukamoto," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 3, pp. 513–522, 2022, doi: 10.25126/jtiik.2022922601.
- [21] J. Han, M. Fang, S. Ye, C. Chen, Q. Wan, and X. Qian, "Using decision tree to predict response rates of consumer satisfaction, attitude, and loyalty surveys," *Sustain.*, vol. 11, no. 8, 2019, doi: 10.3390/su11082306.
- [22] I. Z. P. Hamdan, M. Othman, Y. M. M. Hassim, S. Marjudi, and M. M. Yusof, "Customer Loyalty Prediction for Hotel Industry Using Machine Learning Approach," *Int. J. Informatics Vis.*, vol. 7, no. 3, pp. 695–703, 2023, doi: 10.30630/joiv.7.3.1335.