

Data Mining Berbasis Algoritma C4.5 untuk Klasifikasi Kepuasan Pelanggan

C4.5 Algorithm–Based Data Mining for Customer Satisfaction Classification

^{1,2,3} Program Studi Sistem Informasi, Fakultas Teknik, Universitas Nahdlatul Ulama
Nusa Tenggara Barat

e-mail: ¹rukaiyaimalia@gmail.com, ²soenardhi.75@gmail.com, ³riana.ununtb@gmail.com

Abstrak

Perkembangan era digital menempatkan Internet Service Provider (ISP) sebagai komponen strategis dalam mendukung aktivitas bisnis, pendidikan, dan kehidupan masyarakat. Kondisi ini menyebabkan industri jasa internet berada dalam tingkat persaingan yang tinggi, sehingga kualitas layanan dan kepuasan pelanggan menjadi aspek utama dalam menjaga keberlanjutan serta meningkatkan daya saing perusahaan. Penelitian ini bertujuan untuk mengklasifikasikan tingkat kepuasan pelanggan pada salah satu penyedia jasa internet menggunakan metode data mining berbasis algoritma C4.5, sekaligus mengidentifikasi atribut yang paling berpengaruh terhadap kepuasan pelanggan. Enam atribut yang dianalisis meliputi kecepatan internet, stabilitas koneksi, respons pelayanan, harga, variasi paket, dan kemudahan pembayaran. Proses analisis dilakukan menggunakan perangkat lunak WEKA dengan pembagian data sebesar 70% sebagai data latih (105 data) dan 30% sebagai data uji (45 data). Hasil penelitian menunjukkan bahwa atribut variasi paket dan stabilitas koneksi merupakan faktor paling dominan dalam menentukan kepuasan pelanggan. Model klasifikasi yang dihasilkan mencapai tingkat akurasi sebesar 88,89%, yang mengindikasikan kinerja model sangat baik.

Kata Kunci: *Data mining, Algoritma C4.5, Klasifikasi, Kepuasan, Atribut.*

Abstrack

In the digital era, Internet Service Providers (ISPs) play a strategic role in supporting various activities, including business, education, and daily communication. This condition has intensified competition within the internet service industry, making service quality and customer satisfaction critical factors for maintaining sustainability and enhancing competitiveness. This study aims to classify customer satisfaction levels of an internet service provider using a data mining approach based on the C4.5 algorithm, as well as to identify the most influential attributes affecting customer satisfaction. Six attributes are analyzed, namely internet speed, connection stability, service responsiveness, pricing, package variety, and payment convenience. The analysis is conducted using WEKA software, with the dataset divided into 70% training data (105 instances) and 30% testing data (45 instances). The results indicate that package variety and connection stability are the most dominant factors in determining customer satisfaction. The proposed classification model achieves an accuracy of 88.89%, demonstrating excellent performance and confirming the effectiveness of the C4.5 algorithm for customer satisfaction classification in the ISP domain.

Keyword: *Data Mining, C4.5 mining, Classification, Satisfaction, Attributes.*

PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah mendorong peningkatan kebutuhan masyarakat terhadap layanan internet yang cepat, stabil, dan handal. Di era digital ini, penyedia jasa internet atau Internet Service Provider (ISP) memegang peranan penting dalam menunjang berbagai aktivitas, baik di sektor bisnis, pendidikan, maupun kehidupan sehari-hari. Hal ini menjadikan industri jasa internet sebagai salah satu sektor yang sangat kompetitif, di mana kualitas layanan dan kepuasan pelanggan menjadi faktor kunci dalam mempertahankan eksistensi dan meningkatkan daya saing perusahaan [1]. Dalam konteks ini, memahami dan mengelola kepuasan pelanggan menjadi tantangan strategis bagi perusahaan. Kepuasan pelanggan tidak hanya mencerminkan keberhasilan perusahaan dalam memberikan layanan, tetapi juga menjadi indikator potensial terhadap loyalitas dan retensi pelanggan. Oleh karena itu, perusahaan perlu melakukan analisis yang mendalam terhadap data pelanggan untuk

mengidentifikasi pola-pola tertentu yang dapat digunakan sebagai dasar pengambilan keputusan yang lebih tepat sasaran [2].

Salah satu perusahaan penyedia jasa internet (ISP) yang berlokasi di Mataram, Nusa Tenggara Barat, dan telah memiliki jumlah pelanggan yang cukup banyak. Meskipun demikian, hingga saat observasi dilakukan pada tanggal 05 April 2025, perusahaan tempat lokasi penelitian ini dilakukan belum pernah melakukan evaluasi secara sistematis terhadap tingkat kepuasan pelanggan terhadap layanan yang diberikan. Padahal, evaluasi semacam ini sangat penting sebagai dasar pengambilan keputusan dalam upaya peningkatan kualitas layanan di masa mendatang.

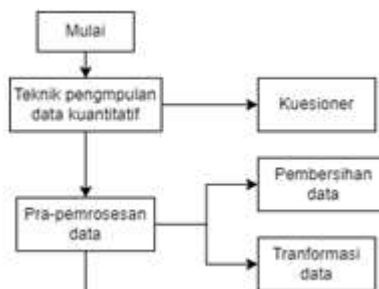
Kondisi tersebut menunjukkan adanya kesenjangan antara pentingnya evaluasi kepuasan pelanggan dengan praktik aktual yang diterapkan. Ketiadaan sistem evaluasi membuat perusahaan berpotensi kehilangan wawasan penting terkait persepsi pelanggan terhadap kualitas layanan yang diberikan [3]. Hal ini juga dapat mengakibatkan ketidaksesuaian antara harapan pelanggan dengan performa layanan yang sesungguhnya, yang dalam jangka panjang dapat mempengaruhi reputasi dan keberlanjutan usaha.

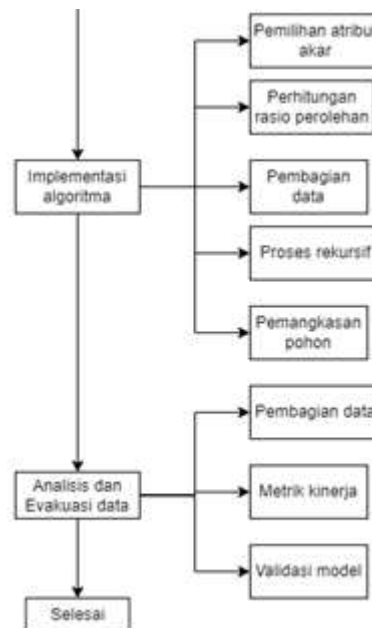
Pendekatan berbasis data semakin relevan dalam pengambilan keputusan bisnis, salah satunya melalui data mining, yang menemukan pola signifikan dari kumpulan data besar. Dalam evaluasi kepuasan pelanggan, data mining membantu mengidentifikasi faktor-faktor yang memengaruhi kepuasan dan memprediksi kecenderungan pelanggan [4]. Dengan teknik ini, perusahaan dapat memahami preferensi pelanggan dan merancang strategi yang lebih efektif untuk meningkatkan kepuasan dan loyalitas. Selain itu, analisis data mining mendukung pengambilan keputusan yang lebih tepat, mengurangi risiko, dan meningkatkan efisiensi operasional, sehingga berkontribusi pada pertumbuhan dan keberlanjutan bisnis.

Algoritma C4.5 merupakan salah satu metode klasifikasi dalam data mining yang efektif untuk mengolah data kategorikal maupun numerik [5]. Algoritma ini bekerja dengan membentuk pohon keputusan (decision tree) yang mampu mengelompokkan data ke dalam kategori-kategori tertentu berdasarkan atribut yang paling berpengaruh [6]- [8]. Dengan menggunakan algoritma ini, data pelanggan dapat dianalisis untuk diklasifikasikan ke dalam kelompok puas dan tidak puas, sehingga memudahkan perusahaan dalam memahami pola kepuasan pelanggan secara lebih terstruktur [9]. Penerapan algoritma C4.5 untuk klasifikasi kepuasan pelanggan menjadi relevan dan signifikan karena dapat menjawab kesenjangan yang ada. Metode ini tidak hanya memberikan hasil klasifikasi, tetapi juga menyajikan pohon keputusan yang mudah diinterpretasikan sebagai dasar rekomendasi tindakan perbaikan. Selain itu, hasil klasifikasi dapat digunakan untuk mengidentifikasi segmen pelanggan yang membutuhkan perhatian khusus atau peningkatan layanan [10]- [12]. Dengan demikian, penelitian ini bertujuan untuk menerapkan metode data mining dengan algoritma C4.5 guna mengklasifikasikan tingkat kepuasan pelanggan sebagai langkah awal dalam membangun sistem evaluasi kepuasan pelanggan yang berbasis data [13]. Diharapkan hasil dari penelitian ini tidak hanya memberikan kontribusi pada pengembangan internal perusahaan, tetapi juga menjadi referensi bagi penyedia layanan sejenis dalam mengadopsi teknologi analitik untuk mendukung pengambilan keputusan strategis.

METODE PENELITIAN

Secara umum, tahapan penelitian ini seperti pada gambar 1 berikut:





Gambar 1. Tahapan Penelitian

Penelitian ini dimulai dengan pengambilan data dengan menggunakan metode kuesioner yang dibuat dan didistribusikan secara online menggunakan *platform Google Forms* kepada para pelanggan. Data yang diperoleh merupakan data terstruktur dalam format table, kemudian data masuk pada tahap pra-pemrosesan, implementasi algoritma, analisis dan evaluasi data.

1. Tahap pra-Pemrosesan

Ada dua tahapan pada tahap pra-pemrosesan, pertama pembersihan data yang melibatkan penanganan nilai-nilai yang hilang dari kuesioner, identifikasi dan penanganan data yang tidak konsisten atau mengandung noise, serta deteksi dan penanganan outlier. Kedua tahap transformasi data dimana data dikonversi dari variabel kategorik ke format numerik [14]- [16].

2. Implementasi Algoritma

Algoritma C4.5 adalah algoritma pembelajaran mesin yang digunakan untuk menghasilkan pohon keputusan yang dapat digunakan untuk klasifikasi [17]. Algoritma ini bekerja dengan membangun pohon keputusan dari sekumpulan data yang telah diklasifikasikan sebelumnya. Pohon keputusan terdiri dari node-node yang mewakili atribut, cabang-cabang yang mewakili nilai-nilai atribut-atribut, dan daun-daun yang mewakili kelas target. Proses pembangunan pohon keputusan pada algoritma C4.5 melibatkan beberapa langkah utama, yaitu:

a. Pemilihan Atribut Akar

Algoritma dimulai dengan memilih atribut yang paling baik memisahkan data menjadi kelas-kelas yang berbeda sebagai node akar. Pemilihan ini didasarkan pada konsep entropi informasi dan perolehan informasi. Entropi mengukur tingkat ketidakmurnian atau keheterogenan dalam kumpulan data, sedangkan perolehan informasi mengukur pengurangan entropi setelah mempratisi data berdasarkan suatu atribut. Adapun atribut yang digunakan untuk penelitian ini meliputi kecepatan internet, stabilitas koneksi, respon pelayanan, harga, variasi paket, dan kemudahan pembayaran. C4.5 bekerja dengan cara menghitung nilai entropy dan gain untuk menentukan atribut mana yang paling informatif dalam membagi data, berikut ini adalah rumus yang digunakan untuk menghitung entropi dan gain [18]- [20].

$$H(S) = -\sum_{i=1}^c p_i \log_2 p_i \quad (1)$$

dengan $H(S)$ adalah entropi dari himpunan S ; c adalah jumlah nilai yang terdapat pada variabel target; dan p_i adalah proporsi dari kelas ke- i pada himpunan S . Sedangkan Rumus Information Gain untuk atribut (A) dapat dihitung dengan rumus berikut:

$$IG(S, A) = H(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} H(S_v) \quad (2)$$

dengan $IG(S, A)$ adalah gain informasi dari atribut A pada himpunan S ; $H(S)$ adalah entropi dari himpunan S ; $v \in \text{Values}(A)$ adalah nilai-nilai yang mungkin dari atribut A ; S_v adalah subset dari S yang memiliki nilai v untuk atribut A ; $|S|$ adalah jumlah total himpunan S ; dan $|S_v|$ adalah jumlah subset dari S_v .

b. Perhitungan Rasio Perolehan

Untuk mengatasi bias terhadap atribut dengan banyak nilai, algoritma C4.5 menggunakan rasio perolehan sebagai kriteria pembagian. Rasio perolehan dinormalisasikan berdasarkan informasi pembagian dari suatu atribut [21].

c. Pembagian Data

Setelah atribut terbaik dipilih sebagai node, data dibagi menjadi subnet-subnet berdasarkan nilai atribut tersebut. Setiap subnet kemudian direpresentasikan oleh cabang yang keluar dari node tersebut.

d. Proses Rekursif

Proses pemilihan atribut terbaik dan pembagian data diulangi secara rekursif untuk setiap subnet sehingga semua instans dalam subnet termasuk ke dalam kelas yang sama atau tidak ada lagi atribut yang memberikan perolehan informasi yang signifikan [22].

e. Pemangkasan Pohon

Setelah pohon keputusan terbentuk, proses pemangkasan dapat dilakukan untuk menghindari overfitting, yaitu kondisi di mana model terlalu cocok dengan data pelatihan dan kurang baik dalam memprediksi data baru. Pemangkasan melibatkan penghapusan cabang yang tidak signifikan atau yang dapat mengurangi akurasi pada data yang belum pernah dilihat sebelumnya [23].

3. Analisis dan Evaluasi Data

Kinerja model klasifikasi C4.5 dievaluasi menggunakan sejumlah metrik kinerja yang lazim diterapkan dalam studi klasifikasi, meliputi akurasi, presisi, recall, F1-score, dan confusion matrix. Akurasi merepresentasikan proporsi instans yang berhasil diklasifikasikan secara tepat oleh model. Presisi menunjukkan proporsi instans yang diprediksi sebagai kelas positif dan benar-benar termasuk dalam kelas tersebut. Recall menggambarkan kemampuan model dalam mengidentifikasi instans positif secara benar. F1-score merupakan nilai rata-rata harmonik antara presisi dan recall yang digunakan untuk menilai keseimbangan kinerja keduanya. Sementara itu, confusion matrix menyajikan informasi terperinci mengenai kinerja model melalui distribusi jumlah instans yang diklasifikasikan secara benar maupun keliru pada setiap kelas [24].

Tabel 1. Metrik Kinerja Klasifikasi

Metrik	Rumus	Deskripsi
Akurasi	$(TP + TN) / (TP + TN + FP + FN)$	Proporsi total prediksi yang benar.
Presisi	$TP / (TP + FP)$	Proporsi instans yang diprediksi sebagai positif yang sebenarnya positif.
Recall	$TP / (TP + FN)$	Proporsi instans positif yang sebenarnya yang diprediksi sebagai positif.
F1-score	$2 * (\text{Presisi} * \text{Recall}) / (\text{Presisi} + \text{Recall})$	Rata-rata harmonik dari presisi dan recall.

TP	True Positive: Instans positif yang diprediksi sebagai positif.
TN	True Negative: Instans negatif yang diprediksi sebagai negatif.
FP	False Positive: Instans negatif yang diprediksi sebagai positif (kesalahan tipe I).
FN	False Negative: Instans positif yang diprediksi sebagai negatif (kesalahan tipe II).

HASIL DAN PEMBAHASAN

Pada bagian ini akan diberikan hasil dan pembahasan yang sesuai dengan alur metodologi penelitian. Pada bagian ini juga akan diberikan pembahasan berupa tabel dan lainnya.

3.1 Tahap Prapemrosesan

Tahap prapemrosesan dibagi menjadi dua yaitu: proses pembersihan data dan transformasi data. Proses pembersihan data pada berkas Excel dilakukan melalui sejumlah tahapan sistematis guna memastikan data siap untuk dianalisis lebih lanjut. Tahapan awal meliputi identifikasi dan pemilihan atribut yang relevan, yaitu variabel-variabel yang secara teoritis dan empiris berpengaruh terhadap tingkat kepuasan pelanggan. Atribut tersebut antara lain kecepatan internet, stabilitas koneksi, respons pelayanan, harga, variasi paket layanan, serta kemudahan dalam proses pembayaran. Selain itu, ditetapkan pula variabel target berupa tingkat kepuasan pelanggan yang diklasifikasikan ke dalam tiga kategori, yaitu “tidak puas”, “netral”, dan “puas”. Pengelompokan ini bertujuan untuk menyederhanakan hasil kuesioner menjadi label kelas yang dapat digunakan secara efektif dalam proses pemodelan menggunakan algoritma C4.5. Tahapan ini merupakan bagian krusial dalam pra-pemrosesan data karena menentukan struktur dan karakteristik data yang akan dipelajari oleh model klasifikasi.

Selanjutnya, dilakukan proses transformasi data dengan menyesuaikan format atribut agar selaras dengan kebutuhan algoritma C4.5 yang diimplementasikan menggunakan perangkat lunak Weka. Transformasi ini mencakup konversi nilai atribut ke dalam bentuk nominal maupun numerik sesuai dengan tipe data yang dapat diproses oleh algoritma. Data yang telah melalui tahap pembersihan dan penyesuaian kemudian dikonversi ke dalam format *.arff* (Attribute-Relation File Format), yaitu format standar yang digunakan oleh Weka untuk keperluan pemodelan. Proses konversi tersebut melibatkan penyesuaian struktur data, termasuk penamaan atribut, deklarasi tipe data, serta pendefinisian nilai atribut nominal, sehingga seluruh dataset dapat dikenali dan diproses secara optimal oleh Weka.

3.2 Implementasi Algoritma

Tabel 2 berikut merupakan hasil implementasi algoritma C4.5 dengan menggunakan rumus (1) dan (2) serta tabel 1 untuk menentukan nilai entropi dan gain

Tabel 2. Nilai Entropi dan Gain Algoritma C4.5

No	Node	Jumlah (S)	TP	N	P	Entropy	Gain
1	Total	150	18	30	102	1,2098	
2	Kecepatan Internet						0,1634
	<=4	1132	18	30	65	1,3890	
	>4	37	0	0	37	0	
3	Stabilitas Koneksi						0,3907
	<=3	74	18	29	27	1,5569	
	>3	76	0	1	75	0,1010	
4	Respon Pelayanan						0,4727
	<=3	62	17	29	16	1,5289	
	>3	88	1	1	86	0,1789	
5	Harga						0,199
	<=4	107	18	30	59	1,454	
	>4	42	0	0	43	0,2762	
6	Varian paket						0,4136
	<= 2	23	18	1	4	0,9104	
	> 2	127	0	29	98	0,775	
7	Proses Pembayaran						0,6156
	<=3	60	18	30	12	1,4855	
	>3	90	0	0	90	0	

3.3 Analisis dan Evaluasi Data

Pada tahap ini dilakukan analisis terhadap hasil klasifikasi tingkat kepuasan pelanggan menggunakan algoritma C4.5 yang diimplementasikan melalui perangkat lunak Weka. Dataset yang digunakan terdiri atas 150 data responden dengan tujuh atribut, yaitu kecepatan_internet, stabilitas_koneksi, respon_pelayanan, harga, variasi_paket, proses_pembayaran, serta kepuasan sebagai atribut target. Data dibagi menggunakan metode split 70:30, dengan 70% sebagai data pelatihan dan 30% sebagai data pengujian, yaitu masing-masing sebanyak 105 data latih dan 45 data uji. Hasil klasifikasi menghasilkan sebuah decision tree (pohon keputusan) dengan atribut variasi_paket sebagai akar (root node). Hal ini menunjukkan bahwa atribut tersebut memiliki nilai Gain tertinggi dalam membedakan tingkat kepuasan pelanggan. Berdasarkan struktur pohon keputusan, jalur klasifikasi pertama dimulai dari kondisi variasi_paket ≤ 3 , yang selanjutnya bercabang ke atribut kecepatan_internet, kemudian ke atribut harga dan proses_pembayaran. Sementara itu, pada kondisi variasi_paket > 3 , pohon keputusan mengarah ke atribut stabilitas_koneksi, harga, dan respon_pelayanan.

a. Akurasi Model

$$\text{Akurasi} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100\%$$

(3)

$$\text{Akurasi} = \frac{(\text{jumlah data benar})}{(\text{Total data uji})} \times 100\%$$

$$\text{Akurasi} = \frac{(40)}{(45)} \times 100\% = 88,89\%$$

Tabel 3. Evaluasi Hasil Klasifikasi

No	Keterangan	Nilai
1	Total Data Uji	45
2	Data Diklasifikasikan Benar	40
3	Data Diklasifikasikan Salah	5
4	Akurasi	88,89%

Berdasarkan Tabel 3, secara keseluruhan pohon keputusan yang terbentuk terdiri atas 9 daun dan 17 simpul, serta merupakan hasil pemangkasan (pruned tree) yang bertujuan untuk menghindari terjadinya overfitting. Model yang dihasilkan mampu melakukan prediksi dengan

tingkat akurasi sebesar 88,89%, di mana 40 dari 45 data uji berhasil diklasifikasikan dengan benar dan hanya 5 data yang mengalami salah klasifikasi.

Selanjutnya, berdasarkan hasil confusion matrix pada Tabel 4, dari total 45 data uji, kelas Puas memiliki 28 data, dengan 25 data terklasifikasi benar, 1 data salah diklasifikasikan sebagai Tidak Puas, dan 2 data salah sebagai Netral. Kelas Tidak Puas terdiri atas 5 data, dengan 4 data terklasifikasi benar dan 1 data salah diklasifikasikan sebagai Puas, tanpa kesalahan ke kelas Netral. Sementara itu, kelas Netral memiliki 12 data, dengan 11 data terklasifikasi benar dan 1 data salah diklasifikasikan sebagai Puas, tanpa kesalahan ke kelas Tidak Puas. Dengan demikian, confusion matrix menunjukkan bahwa model mampu mengenali pola data dengan sangat baik, menghasilkan total 40 prediksi benar dari 45 data uji dan tingkat akurasi keseluruhan sebesar 88,89%.

Tabel 4. Hasil Confusion Matrix

No	Aktial / Prediksi	Puas	Tidak Puas	Netral
1	Puas	25	1	2
2	Tidak Puas	1	4	0
3	Netral	1	0	11

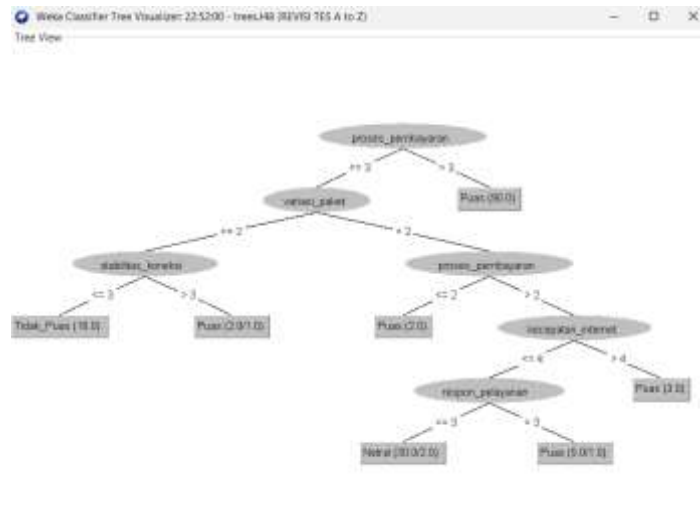
Sedangkan untuk matrik evaluasi, seperti pada tabel 5 di bawah ini. Precision tertinggi dicapai pada kelas Puas sebesar 0,9259, menandakan bahwa seluruh prediksi pada kelas tersebut benar tanpa kesalahan (*false positive*).

Tabel 5. Matriks Evaluasi

No	Kelas	Precision	Recall
1	Puas	0,9259	0,8929
2	Tidak Puas	0,800	0,800
3	Netral	0,8462	0,9167

Hasil penerapan algoritma C4.5 menunjukkan model klasifikasi mampu membedakan tingkat kepuasan pelanggan berdasarkan atribut-atribut yang telah dipilih, seperti kecepatan internet, stabilitas koneksi, respon pelayanan, harga, variasi paket, dan kemudahan pembayaran. Pohon keputusan yang dihasilkan memperlihatkan struktur hirarki yang logis, di mana atribut dengan rasio perolehan informasi tertinggi ditempatkan di node atas [26]. Seperti, atribut proses pembayaran dan variasi paket sering muncul di level awal pohon, yang menunjukkan bahwa kedua atribut ini memiliki pengaruh dominan terhadap kepuasan pelanggan. Hal ini sesuai dengan ekspektasi pelanggan terhadap kualitas layanan ISP.

Pohon keputusan yang dihasilkan oleh algoritma C4.5 (J48) pada Gambar 1 menunjukkan struktur klasifikasi berdasarkan atribut yang paling informatif dalam menentukan tingkat kepuasan pelanggan. Atribut variasi_paket menempati posisi sebagai root node, yang menunjukkan bahwa atribut tersebut memiliki nilai gain tertinggi dan menjadi dasar utama dalam proses pembagian data.



Gambar 2. Hasil Pohon Keputusan Pada WEKA

Pohon keputusan dibaca secara hierarkis, dimulai dari node akar (root node) hingga node daun (leaf node) dengan mengikuti jalur percabangan yang ditentukan oleh kondisi pada setiap atribut. Setiap percabangan merepresentasikan proses pengujian suatu atribut, sedangkan keputusan klasifikasi ditetapkan ketika node daun tercapai.

Sebagai ilustrasi, apabila proses pembayaran ≤ 3 dan variasi paket ≤ 2 , maka dilakukan pengujian terhadap stabilitas koneksi. Jika stabilitas koneksi ≤ 3 , sistem memprediksi kelas Tidak Puas dengan 18 data terklasifikasi benar. Sebaliknya, jika stabilitas koneksi > 3 , sistem memprediksi kelas Puas dengan 2 data terklasifikasi benar dan 1 data salah. Selanjutnya, jika proses pembayaran ≤ 3 dan variasi paket > 2 , maka dilakukan pengujian lanjutan terhadap proses pembayaran. Apabila nilainya ≤ 2 , sistem memprediksi kelas Puas dengan 2 data benar. Jika > 2 , pengujian dilanjutkan ke atribut kecepatan internet. Pada kondisi kecepatan internet ≤ 4 , pengujian dilanjutkan ke respon pelayanan, di mana nilai ≤ 3 diprediksi sebagai Netral (30 data benar dan 2 data salah), sedangkan nilai > 3 diprediksi sebagai Puas (5 data benar dan 1 data salah). Jika kecepatan internet > 4 , maka sistem langsung memprediksi kelas Puas dengan 3 data terklasifikasi benar. Sementara itu, apabila proses pembayaran > 3 , sistem secara langsung memprediksi kelas Puas dengan total 90 data. Notasi pada setiap node daun, seperti (5.0/1.0), menunjukkan jumlah data yang masuk ke dalam suatu kelas, dengan angka sebelum garis miring merepresentasikan jumlah data yang diklasifikasikan dengan benar (true positive), sedangkan angka setelah garis miring menunjukkan jumlah data yang salah klasifikasi (false positive).

Secara keseluruhan, struktur pohon keputusan ini memberikan gambaran yang sistematis mengenai pola hubungan antaratribut yang memengaruhi tingkat kepuasan pelanggan. Pohon keputusan juga berperan sebagai alat bantu pengambilan keputusan berbasis data melalui representasi yang logis dan mudah diinterpretasikan. Berdasarkan hasil klasifikasi menggunakan algoritma C4.5 yang diimplementasikan melalui perangkat lunak Weka, model terbukti mampu mengklasifikasikan tingkat kepuasan pelanggan secara efektif dengan tingkat akurasi sebesar 88,89%. Atribut proses pembayaran teridentifikasi sebagai faktor paling dominan dalam memengaruhi kepuasan pelanggan, diikuti oleh variasi paket, stabilitas koneksi, dan kecepatan internet. Hal ini diperkuat oleh nilai Kappa Statistic sebesar 0,7932, yang menunjukkan tingkat kesepakatan klasifikasi yang kuat. Dengan demikian, algoritma C4.5 dinilai andal dalam mengungkap pola-pola penting dalam data dan layak dijadikan dasar pengambilan keputusan strategis dalam upaya peningkatan kualitas layanan. [27]–[29]

KESIMPULAN

Berdasarkan hasil penelitian mengenai penerapan metode data mining menggunakan algoritma C4.5 untuk klasifikasi kepuasan pelanggan, dapat disimpulkan bahwa model klasifikasi berhasil dibangun dengan tingkat akurasi yang tinggi, yaitu sebesar 88,89%, serta mampu membedakan secara efektif pelanggan yang puas, netral, dan tidak puas. Atribut yang paling berpengaruh dalam proses klasifikasi adalah proses pembayaran, diikuti oleh variasi paket, stabilitas koneksi, dan kecepatan internet, yang tercermin dari posisinya pada level awal pohon keputusan. Hasil evaluasi menggunakan confusion matrix serta metrik precision, recall, dan F1-score menunjukkan performa model yang sangat baik dengan tingkat kesalahan klasifikasi yang rendah. Selain menghasilkan klasifikasi yang akurat, algoritma C4.5 juga menghasilkan model yang mudah diinterpretasikan, sehingga mendukung evaluasi kepuasan pelanggan secara sistematis dan pengambilan keputusan berbasis data untuk peningkatan kualitas layanan.

DAFTAR PUSTAKA

- [1] L. Maulina, "Revitalisasi Industri Perhotelan Dengan Inovasi Teknologi: Meningkatkan Keunggulan Bersaing Dan Pengalaman Pelanggan," *J. Ilm. Manajemen, Ekon. Akunt.*, vol. 7, no. 1, pp. 504–519, 2023.
- [2] D. Ashari, M. S. Ladaina, and T. Hartini, "Peran Big Data Dalam Pengambilan Keputusan Strategis Perusahaan," *EKOMAN J. Ekon. Bisnis dan Manaj.*, vol. 2, no. 3, pp. 401–422, 2024.
- [3] M. H. Hutapea, N. A. B. Rahmani, and K. Tambunan, "Pengaruh Kualitas Pelayanan, Persepsi Harga Dan Cita Rasa Terhadap Kepuasan Konsumen Pada Durian Kocok 77 Tempuling," *J. Ilm. Ekon. Dan Manaj.*, vol. 2, no. 5, pp. 44–60, 2024.
- [4] D. Penjualan, M. Algoritma, and D. Tree, "PENGUNAAN DATA MINIG UNTUK MENGIDENTIFIKASI PELANGGAN BERESIKO TINGGI," vol. 2, no. 1, pp. 46–51, 2025.
- [5] U. Suriani, "Penerapan data mining untuk memprediksi tingkat kelulusan mahasiswa menggunakan algoritma decision tree C4. 5," *J. Comput. Inf. Syst. Ampera*, vol. 4, no. 2, pp. 55–65, 2023.
- [6] S. Ramadhani, "Klasifikasi Berita Menggunakan Algoritma C4. 5," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 2, 2022.
- [7] E. Jannah, V. Sihombing, and M. Masrizal, "Penerapan Data Mining Klasifikasi Kepuasan Pelanggan Transportasi Online Menggunakan Algoritma C4. 5," *MEANS (Media Inf. Anal. dan Sist.*, pp. 1–7, 2023.
- [8] D. Telaumbanua and I. Kurniawati, "Penerapan Algoritma C4.5 Untuk Klasifikasi Kepuasan Pelanggan Pada Jasa Layanan Pengiriman," *JoMMiT J. Multi Media dan IT*, vol. 6, no. 1, 2022, doi: 10.46961/jommit.v6i1.524.
- [9] M. A. C, "Analisa Kepuasan Pelanggan Terhadap Layanan Aplikasi E-Commerce," vol. 4, no. 6, pp. 530–542, 2024.
- [10] C. Pada, T. Pakaian, and R. I. A. Busana, "KLASIFIKASI HASIL PENJUALAN PAKAIAN MENGGUNAKAN ALGORITMA," vol. 9, no. 1, pp. 1291–1299, 2025.
- [11] R. Pratama, M. I. Herdiansyah, D. Syamsuar, and A. Syazili, "Prediksi Customer Retention Perusahaan Asuransi Menggunakan Machine Learning," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 96–104, 2023, doi: 10.32736/sisfokom.v12i1.1507.
- [12] M. A. C, "Cv Karya Abshor Pematang Siantar," vol. 4, pp. 72–79, 2024.

- [13] A. Surahmat and M. Sutrisno, "Analisis Kepuasan Pelanggan Dalam Industri Teknologi Menggunakan Algoritma C4.5," *Anal. Kepuasan Pelangg. Dalam Ind. Teknol. Menggunakan Algoritma C4.5*, vol. 13, no. 2, pp. 75–79, 2023.
- [14] G. Urva *et al.*, *PENERAPAN DATA MINING DI BERBAGAI BIDANG: Konsep, Metode, dan Studi Kasus*. PT. Sonpedia Publishing Indonesia, 2023.
- [15] I. M. Karo Karo and H. Hendriyana, "Klasifikasi Penderita Diabetes menggunakan Algoritma Machine Learning dan Z-Score," *J. Teknol. Terpadu*, vol. 8, no. 2, pp. 94–99, 2022, doi: 10.54914/jtt.v8i2.564.
- [16] A. Novareza, R. Herdiana, and I. Iin, "Analisis data mining pada produksi ikan air tawar di Kecamatan Panjalu menggunakan metode C4. 5," *JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 1, pp. 252–258, 2024.*
- [17] F. Pirmansyah and T. Wahyudi, "Implementasi Data Mining Menggunakan Algoritma C4. 5 Untuk Prediksi Evaluasi Anggota Satuan Pengamanan Studi Kasus Pt. Yimm Pulogadung," *J. Indones. Manaj. Inform. Dan Komun., vol. 4, no. 3, pp. 1566–1580, 2023.*
- [18] M. Nur, "Data Mining Untuk Memprediksi Kelulusan Mahasiswa Jurusan Teknik Informatika UIN Syarif Hidayatullah Jakarta Menggunakan Metode Klasifikasi C4. 5." *Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah ...*, 2022.
- [19] A. Irdianto and S. Supatman, "Pemanfaatan Kecerdasan Buatan dalam Penghitungan Ekuitas Harga Obat di Gudang Farmasi Dinas Kesehatan Kabupaten Purworejo," *J. MEDIA INFOTAMA*, vol. 20, no. 2, pp. 439–447, 2024.
- [20] S. Shalsabilla, P. Rachmawati, K. Vidya Prakusa, and S. Rihastuti, "Penerapan Data Mining dengan Metode Decision Tree untuk Prediksi Cuaca di Kota Seattle menggunakan Aplikasi Weka," *Semin. Nas. Amikom Surakarta*, no. November, pp. 93–100, 2023.
- [21] M. Waruwu, S. N. Puat, P. R. Utami, E. Yanti, and M. Rusydiana, "Metode Penelitian Kuantitatif: Konsep, Jenis, Tahapan dan Kelebihan," *J. Ilm. Profesi Pendidik., vol. 10, no. 1, pp. 917–932, 2025.*
- [22] R. H. Pambudi, B. D. Setiawan, and Indriati, "Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal," *Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 2, no. 7, pp. 2637–2643, 2020.*
- [23] R. A. W. Sujana and I. M. A. Agastya, "Penerapan Algoritma Machine Learning Untuk Sistem Prediksi Penyakit Osteoporosis," *J. Appl. Informatics Comput., vol. 8, no. 2, pp. 304–315, 2024.*
- [24] C. M. Sidiq, A. Faqih, and G. Dwilestari, "ALGORITMA DECISION TREE C4. 5 DIGUNAKAN UNTUK MENGLASIFIKASIKAN DATA STROKE," *JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 2, pp. 1869–1874, 2024.*
- [25] W. Nugraha and M. Syarif, "Evaluasi Performa Algoritma Klasifikasi dalam Prediksi Gagal Jantung: Studi Kasus Dataset Heart Failure Predictio.,," *Techno. com*, vol. 23, no. 4, 2024.
- [26] M. Waterboom, "Klasifikasi Data Mining Pada Tingkat Kepuasan Pengunjung," vol. 7, pp. 270–281, 2024.
- [27] Y. S. Nasution, "Analisis Penerapan Algoritma C4. 5 Dalam Mengukur Tingkat Kepuasan Pelanggan," *SkripsiKu-2022*, vol. 1, no. 1, 2022.
- [28] 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.
- [29] N. Azwanti and E. Elisa, "Analisa Kepuasan Konsumen Menggunakan Algoritma C4.5," *Pros. Semin. Nas. Ilmu Sos. dan Teknol., no. 3, pp. 126–131, 2020.*