



PENERAPAN ALGORITMA KLASIFIKASI C4.5 REVIEW PELANGGAN DALAM BERTRANSAKSI PADA PT. DAPOER MAMIH

Wahana Indra Komala ^{a*}, Andri Yansah ^b, Resti Pebrina ^c, Suhardjono ^d

^a Sistem Informasi; komala.wahana@gmail.com, Universitas Bina Sarana Informatika, Jakarta, Indonesia

^b Sistem Informasi; andri@gmail.com, Universitas Bina Sarana Informatika, Jakarta, Indonesia

^c Sistem Informasi; restipebrina2502@gmail.com, Universitas Bina Sarana Informatika, Jakarta, Indonesia

^d Sistem Informasi; suhardjono@bsi.ac.id, Universitas Bina Sarana Informatika, Jakarta, Indonesia

*Penulis Korespondensi: Wahana Indra Komala

ABSTRACT

Wahana Indra Komala (19242542), Andri Yansah (19242561), Resti Pebrina (19242563), Implementation of C4.5 Classification Algorithm Customer Review in Transactions at PT. Dapoer Mamih

PT. Dapoer Mamih is a culinary company that utilizes the INAPROC e-catalogue platform in its business transactions. The company faces challenges due to the increasing number of customer reviews that have not been systematically analyzed. This study aims to classify customer reviews using the C4.5 algorithm to understand customer perceptions of products and services. The research data were collected from customer reviews during April–June 2025 through the INAPROC platform and direct questionnaires. The applied method includes data preprocessing, sentiment labeling, decision tree construction, and model evaluation using RapidMiner. The results show that the C4.5 algorithm can classify reviews into positive, neutral, and negative categories with satisfactory accuracy. The model also produces classification rules that can serve as a basis for the company to improve service quality. These findings demonstrate that the application of the C4.5 algorithm can be an effective decision support tool for managing customer feedback.

Keywords: Classification; C4.5; Customer Reviews; Decision Trees; Data Mining.

ABSTRAK

Wahana Indra Komala (19242542), Andri Yansah (19242561), Resti Pebrina (19242563), Penerapan Algoritma Klasifikasi C4.5 Review Pelanggan Dalam Bertransaksi Pada PT. Dapoer Mamih

PT. Dapoer Mamih merupakan perusahaan kuliner yang memanfaatkan platform *e-catalogue* INAPROC dalam transaksi bisnisnya. Perusahaan menghadapi tantangan berupa meningkatnya jumlah *review* pelanggan yang belum diolah secara sistematis. Penelitian ini bertujuan mengklasifikasikan *review* pelanggan menggunakan algoritma C4.5 untuk mengetahui persepsi pelanggan terhadap produk dan layanan. Data penelitian dikumpulkan dari *review* pelanggan pada periode April–Juni 2025 melalui platform INAPROC dan kuesioner langsung. Metode yang digunakan meliputi *data preprocessing*, pelabelan sentimen, pembentukan *decision tree*, serta evaluasi model menggunakan RapidMiner. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu mengklasifikasikan *review* ke dalam kategori positif, netral, dan negatif dengan tingkat akurasi yang memadai. Model ini menghasilkan aturan-aturan klasifikasi yang dapat dijadikan dasar bagi perusahaan dalam meningkatkan kualitas layanan. Temuan ini membuktikan bahwa penerapan C4.5 dapat menjadi alat pendukung keputusan bagi manajemen dalam pengelolaan umpan balik pelanggan.

Kata Kunci: Klasifikasi; Algoritma C4.5; Review Pelanggan; Pohon Keputusan; Data Mining

1. PENDAHULUAN

Seiring menyebarnya perkembangan di era industri 4.0, teknologi informasi dan komunikasi semakin berkembang mengikuti zaman. Tingkat mobilitas masyarakat di kegiatan sehari-hari menjadi meningkat dan melonjak. Dengan meningkatnya mobilitas yang tinggi, masyarakat akan membiasakan diri dengan gaya hidup yang serba instan untuk memenuhi kebutuhan hidup.

Semakin banyak pelaku usaha yang memanfaatkan Platform E-commerce untuk menjangkau pasar yang lebih luas, sehingga menciptakan persaingan yang ketat di antara mereka. E-commerce memberikan dampak positif bagi perekonomian, termasuk peningkatan konsumsi domestik, aksesibilitas produk yang lebih luas, dan kemudahan transaksi jual beli secara online.

E-commerce juga berperan penting dalam mendukung Usaha Mikro, Kecil, dan Menengah (UMKM) untuk bersaing di pasar yang lebih luas, termasuk pasar global. Selain faktor teknologi, pertumbuhan E-commerce juga didorong oleh peningkatan jumlah penduduk, khususnya kelas menengah, yang memiliki daya beli lebih tinggi.

Salah satu sektor yang turut memanfaatkan kemajuan ini adalah industri kuliner. Dalam menjalankan usahanya, perusahaan makanan tidak hanya dituntut untuk menyajikan produk yang berkualitas, tetapi juga harus mampu memahami dan merespon pengalaman serta kepuasan pelanggan dengan cepat dan tepat.

PT. Dapoer Mamih merupakan perusahaan yang bergerak di bidang kuliner dan telah melayani berbagai transaksi baik secara langsung maupun melalui Platform online. PT. Dapoer Mamih juga merupakan perusahaan yang memanfaatkan E-government dalam mendukung kegiatan bisnis dan operasional mereka. Seiring bertambahnya jumlah pelanggan dan transaksi, jumlah ulasan atau review pelanggan yang diterima oleh perusahaan juga meningkat secara signifikan. Review tersebut mencerminkan pengalaman, kepuasan, maupun keluhan pelanggan selama berinteraksi dengan produk dan layanan yang diberikan. Namun, hingga saat ini, review-review tersebut belum dimanfaatkan secara maksimal sebagai sumber informasi yang dapat digunakan untuk pengambilan keputusan strategis oleh manajemen.

2. TINJAUAN PUSTAKA

2.1. INAPROC (Indonesia National Procurement)

Indonesia National Procurement atau INAPROC merupakan sistem e-procurement nasional yang dikembangkan pemerintah untuk memfasilitasi proses pengadaan barang dan jasa secara daring. INAPROC terintegrasi dengan Layanan Pengadaan Secara Elektronik (LPSE) yang memungkinkan instansi pemerintah maupun pelaku usaha melakukan transaksi secara transparan, efisien, dan akuntabel [6], [7].

Platform ini menyediakan berbagai fitur, salah satunya adalah e-catalogue, yang berfungsi sebagai sarana transaksi langsung antara penyedia dan pengguna barang/jasa. Melalui e-catalogue, proses pemesanan dapat dilakukan dengan lebih cepat tanpa melalui prosedur tender yang kompleks. Bagi perusahaan penyedia seperti PT. Dapoer Mamih, keberadaan e-catalogue INAPROC mempermudah dalam memasarkan produk kuliner sekaligus membuka peluang memperluas pasar.

Selain itu, sistem INAPROC juga menyediakan ruang bagi pelanggan untuk memberikan review atau umpan balik setelah transaksi. Review ini berperan penting sebagai indikator kualitas layanan sekaligus sumber informasi bagi penyedia jasa dalam memperbaiki kinerja mereka [8]. Namun, jumlah review yang terus bertambah membuat analisis manual menjadi sulit, sehingga diperlukan metode berbasis data mining untuk mengklasifikasikan informasi tersebut secara sistematis.

2.2. Review Pelanggan

Review pelanggan merupakan bentuk umpan balik yang diberikan konsumen setelah melakukan transaksi, baik berupa teks komentar maupun rating numerik. Informasi dalam review dapat menggambarkan tingkat kepuasan, persepsi kualitas produk, serta pengalaman layanan yang diterima pelanggan [8].

Analisis review pelanggan sangat penting karena dapat menjadi dasar evaluasi perusahaan untuk memperbaiki kualitas produk dan layanan. Semakin tinggi jumlah review yang tersedia, semakin besar pula peluang bagi perusahaan untuk memperoleh wawasan strategis mengenai kebutuhan dan ekspektasi konsumen. Namun, data review yang besar membutuhkan teknik analisis otomatis seperti text mining atau data mining agar dapat diolah secara efektif.

2.3. Data Mining dan Klasifikasi

Data mining merupakan proses menemukan pola tersembunyi dari kumpulan data dalam jumlah besar sehingga dapat diubah menjadi informasi yang bermanfaat. Proses ini melibatkan teknik analisis statistik, kecerdasan buatan, dan pembelajaran mesin untuk mengidentifikasi keterhubungan antar data [9].

Salah satu metode populer dalam data mining adalah klasifikasi, yaitu teknik untuk memetakan data ke dalam kategori tertentu. Klasifikasi banyak digunakan pada bidang pemasaran, prediksi perilaku konsumen, serta analisis sentimen pelanggan. Dengan memanfaatkan klasifikasi, perusahaan dapat mengelompokkan review pelanggan menjadi kategori positif, netral, atau negatif secara otomatis.

2.4. Algoritma C4.5 dan *Decision Tree*

Algoritma C4.5 merupakan algoritma klasifikasi berbasis *decision tree* yang dikembangkan oleh Quinlan. Algoritma ini bekerja dengan cara menghitung nilai *entropy* dan *information gain* untuk memilih atribut terbaik yang dijadikan simpul (node) pada pohon keputusan [10].

Keunggulan algoritma C4.5 adalah: (1) Mampu mengolah data kategorikal maupun numerik, (2) Dapat menangani missing value, (3) Menghasilkan aturan klasifikasi yang mudah dipahami dalam bentuk *if-then rules*. Karena keunggulannya, algoritma C4.5 banyak digunakan untuk menganalisis review pelanggan, memprediksi kepuasan, serta mendukung pengambilan keputusan manajerial.

Decision tree adalah metode klasifikasi berbentuk struktur pohon yang menyajikan aturan-aturan keputusan dalam bentuk hierarki [10]. Setiap simpul (node) merepresentasikan atribut, cabang menggambarkan hasil uji atribut, dan daun (leaf) mewakili kelas atau keputusan akhir.

2.5. RapidMiner

RapidMiner adalah perangkat lunak *open-source* yang banyak digunakan untuk melakukan analisis data, *data mining*, dan pembelajaran mesin (*machine learning*). RapidMiner menyediakan antarmuka visual berbasis *drag and drop* sehingga memudahkan peneliti dalam membangun model analisis tanpa harus menulis kode pemrograman secara manual [11].

Dalam penelitian ini, RapidMiner digunakan untuk membangun model klasifikasi menggunakan algoritma C4.5, menghasilkan *decision tree*, serta melakukan evaluasi model dengan data uji.

2.6. Penelitian Terkait

Tabel 1. Referensi Jurnal

Peneliti	Tahun	Objek Penelitian	Metode	Hasil Utama
Oktalia & Zakaria [11]	2024	Review aplikasi JNE (Google Playstore)	C4.5 vs Naïve Bayes	Algoritma C4.5 memberikan akurasi lebih tinggi dibandingkan Naïve Bayes.
Angie, et al. [12]	2025	Review pelanggan di Shopee Marketplace	C4.5	C4.5 berhasil mengklasifikasikan faktor <i>review</i> dan rekomendasi produk dengan baik.
Uape & Elisa [13]	2023	Penilaian layanan transportasi Gojek	C4.5	C4.5 mampu mengelompokkan kualitas layanan driver berdasarkan <i>review</i> .
Siahaan & Muliono [14]	2024	Review kualitas kopi di e-commerce	C4.5	Klasifikasi <i>review</i> kopi akurat, faktor utama adalah rasa dan aroma.
Viska & Elisa [15]	2023	Review pelanggan pada fashion thrift	C4.5	Algoritma C4.5 akurat dalam memetakan kepuasan pelanggan terhadap produk fashion bekas.

3. METODOLOGI PENELITIAN

Penelitian ini dilakukan dengan pendekatan *data mining* menggunakan algoritma klasifikasi C4.5 untuk menganalisis *review* pelanggan pada PT. Dapoer Mamih. Metodologi penelitian terdiri dari beberapa tahapan sistematis sebagai berikut:

3.1. Tahapan Penelitian

Tahapan penelitian disusun berdasarkan proses *Knowledge Discovery in Databases* (KDD): 1) Studi Literatur; 2) Identifikasi Masalah; 3) Pengumpulan Data; 4) Preprocessing Data; 5) Penerapan Algoritma C4.5; 6) Evaluasi Model; 7) Analisis Hasil.

3.2. Instrumen Penelitian

Instrumen penelitian terdiri dari: (1) Dataset 500 review pelanggan (April–Juni 2025); (2) Pedoman Pelabelan Sentimen; (3) Microsoft Excel untuk pengolahan awal; dan RapidMiner Studio untuk penerapan algoritma C4.5.

3.3. Populasi dan Sampel Penelitian

Populasi penelitian adalah seluruh review pelanggan PT. Dapoer Mamih yang ditulis melalui platform INAPROC maupun kuesioner offline. Sampel penelitian berjumlah 500 review pelanggan yang memenuhi kriteria: ditulis pada periode April–Juni 2025, berbahasa Indonesia, dan berisi komentar relevan terhadap produk atau layanan.

3.4. Metode Analisis Data

Analisis data dilakukan melalui: (1) Penyusunan dataset dengan atribut utama (rasa makanan, kualitas pelayanan, ketepatan pengiriman, komentar pelanggan); (2) Pelabelan sentimen (positif, netral, negatif); (3) Pelatihan model klasifikasi menggunakan algoritma C4.5; (4) Evaluasi model dengan confusion matrix untuk memperoleh akurasi, presisi, recall, dan F-measure; (5) Penyajian decision tree dan aturan klasifikasi yang terbentuk.

4. HASIL DAN PEMBAHASAN

4.1. Perancangan Penelitian

Penelitian ini menggunakan 500 data *review* pelanggan PT. Dapoer Mamih yang diperoleh dari dua sumber, yaitu: (1) Transaksi melalui platform INAPROC (e-catalogue) sebanyak 245; (2) Transaksi langsung di booth perusahaan sebanyak 255 data.

Atribut yang digunakan sebagai variabel input meliputi: kategori produk, jenis pelanggan, tempat transaksi, rasa makanan, akurasi produk, kecepatan pengiriman, serta pelayanan. Variabel target adalah tingkat kepuasan pelanggan yang diklasifikasikan menjadi dua kelas, yaitu Puas dan Tidak Puas.

Pra-pemrosesan data dilakukan melalui: (1) Pembersihan data untuk menghapus data kosong atau duplikat; (2) Transformasi data dengan mengubah data kategorikal menjadi numerik; (3) Feature selection untuk memilih atribut paling relevan dalam klasifikasi.

4.2. Klasifikasi Algoritma C4.5 dalam Pembentukan *Decision Tree*

Proses klasifikasi dilakukan dengan algoritma C4.5 melalui tahapan perhitungan *entropy*, *information gain*, dan pembentukan *decision tree*. Dengan rumus entropy sebagai berikut:

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

dengan:

p_i = proporsi jumlah data pada kelas ke- i

n = jumlah kelas

Pada penelitian ini terdapat 500 data *review*. Langkah pertama dalam proses klasifikasi menggunakan algoritma C4.5 adalah menghitung nilai *Entropy* total, yaitu untuk mengetahui tingkat ketidakpastian dalam keseluruhan data.

Tabel 2. Perhitungan Entropi Total (S)

Total Kasus	Sum (Puas)	Sum (Tidak Puas)	Entropi Total
500	464	36	0,373343321

Setelah dilakukan perhitungan, diperoleh nilai Entropy total sebesar 0,373343321. Nilai entropy yang relatif rendah ini menunjukkan bahwa data *Review* pelanggan cenderung homogen, di mana sebagian besar

pelanggan memberikan *Review* dengan kategori "Puas". Perhitungan Entropy ini menjadi dasar dalam proses selanjutnya, yaitu penghitungan *Information Gain* untuk setiap atribut yang akan menentukan cabang-cabang dalam pembentukan *Decision Tree*.

Adapun rumus *Information Gain* sebagai berikut:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times \text{Entropy}(S_i)$$

dengan:

S : Himpunan Kasus

A : Atribut

n : Jumlah Partisi Atribut A

|S_i| : Jumlah Kasus pada Partisi ke-i

|S| : Jumlah Kasus dalam S

Tabel 3. Hasil Perhitungan *Entropy* dan *Gain* setiap Atribut

Node	Atribut	Nilai	Total Kasus	Puas	Tidak Puas	Entropi (Si)	Gain
1	Kategori Produk	Ice Cream Singapura Coklat	60	57	3	0.286396957	
		Ice Cream Singapura Durian	30	21	9	0.881290899	
		Ice Cream Singapura Mix	30	29	1	0.2108423	
		Nasi Box	172	164	8	0.271389588	
		Snack Box	80	71	9	0.507410964	
		Paket Rice Bowl	48	43	5	0.482066148	
		Paket Dimsum	80	79	1	0.096944606	0
2	Pelanggan	Perusahaan	245	227	18	0.37873934	
		Umum	255	237	18	0.368115005	0
3	Tempat Transaksi	E-Katalog	245	227	18	0.37873934	
		Booth	255	237	18	0.368115005	0
4	Rasa Makanan	Sedap	450	450	0	0	
		Kurang Sedap	14	14	0	0	
		Tidak Sedap	36	0	36	0	0.373343321
5	Akurasi Produk	Baik	454	454	0	0	
		Kurang Baik	10	10	0	0	
		Tidak Baik	36	0	36	0	0.373343321
6	Kecepatan Kirim	Baik	454	454	0	0	
		Kurang Baik	10	10	0	0	
		Tidak Baik	36	0	36	0	0.373343321
7	Pelayanan	Baik	464	464	0	0	0.373343321
		Tidak Baik	36	0	36	0	

Dari hasil perhitungan *Information Gain*, atribut Rasa Makanan dan Pelayanan memiliki nilai Gain tertinggi yaitu 0.3733433. Dalam kasus ini nilai Gain yang didapatkan sama, maka di ambil node pertama atribut yang memiliki nilai besar yaitu atribut pelayanan dengan nilai total puas sebesar 464 yang dihasilkan dari nilai pelayanan baik dari *Review* pelanggan untuk menentukan *Decision Tree* (Pohon Keputusan). Langkah selanjutnya adalah membangun cabang dari kategori "Baik" dan "Tidak Baik".



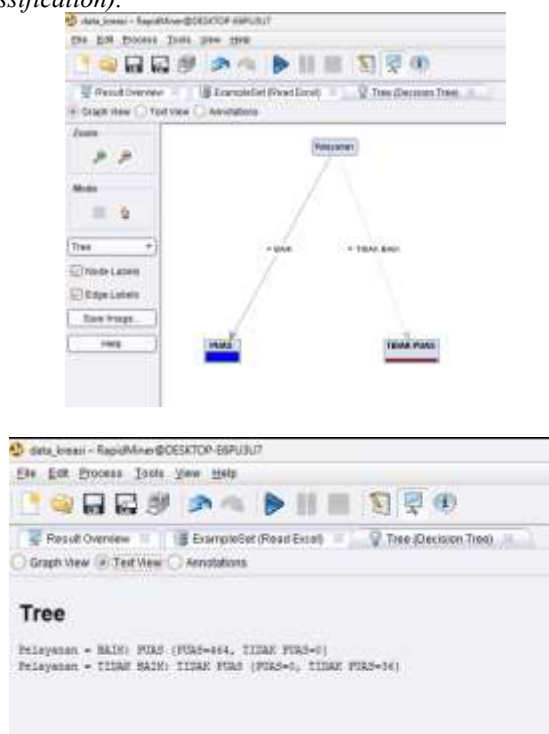
Gambar 1. Hasil *Decision Tree*

Berdasarkan hasil perhitungan *Information Gain* dan analisis data *Review* pelanggan di PT. Dapoer Mamih, atribut yang memiliki pengaruh paling besar dalam proses klasifikasi adalah atribut "Pelayanan", sehingga atribut ini dipilih sebagai akar pohon keputusan (*root node*).

Langkah pertama dalam proses klasifikasi adalah mengevaluasi kondisi pelayanan yang diberikan kepada pelanggan. Jika pelanggan memberikan *Review* bahwa "Pelayanan" dalam kategori "Baik", maka sistem secara otomatis mengklasifikasikan *Review* tersebut sebagai "Puas". Di sisi lain, jika kondisi "Pelayanan" berada dalam kategori "Tidak Baik", maka sistem langsung mengklasifikasikan *Review* tersebut sebagai "Tidak Puas" tanpa perlu mempertimbangkan atribut lainnya.

4.3. Penerapan RapidMiner

Model klasifikasi juga diimplementasikan dengan perangkat lunak RapidMiner menggunakan alur: (1) *Import* dataset dengan operator *Read Excel*; (2) Menentukan variabel target (label) yaitu *Review* pelanggan; (3) Membangun model dengan operator *Decision Tree* (C4.5); (4) Mengevaluasi model menggunakan operator *Performance* (*Classification*).



Gambar 2. Tampilan *Decision Tree* pada RapidMiner

Hasil *decision tree* dari RapidMiner konsisten dengan perhitungan manual, dengan **Pelayanan** sebagai node utama. Visualisasi pohon keputusan menunjukkan bahwa pelanggan dengan pelayanan baik diklasifikasikan sebagai **Puas** (464 data), sedangkan pelayanan tidak baik diklasifikasikan sebagai **Tidak Puas** (36 data).

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma C4.5 efektif digunakan dalam klasifikasi review pelanggan PT. Dapoer Mamih. Model yang dihasilkan menghasilkan aturan klasifikasi yang jelas dan mudah dipahami. Penelitian ini membuktikan bahwa algoritma C4.5 efektif dalam mengklasifikasikan review pelanggan PT. Dapoer Mamih. Model yang dihasilkan memiliki akurasi tinggi dan mampu mengidentifikasi faktor utama kepuasan pelanggan. Atribut utama yang memengaruhi kepuasan pelanggan adalah kualitas pelayanan.

Saran untuk penelitian selanjutnya adalah melakukan perbandingan antara algoritma C4.5 dengan algoritma klasifikasi lain, seperti Naïve Bayes atau Random Forest, guna mengetahui model dengan performa terbaik. Selain itu, penelitian dapat diperluas dengan menggunakan dataset yang lebih besar dan mencakup periode waktu yang lebih panjang agar hasil yang diperoleh lebih general.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada PT. Dapoer Mamih yang telah memberikan izin penelitian serta Universitas Bina Sarana Informatika atas dukungan akademik.

DAFTAR PUSTAKA

- [1] V. Haekal, et al., “Perkembangan E-Commerce di Era Industri 4.0,” Jurnal Teknologi Informasi, vol. 7, no. 2, pp. 45–53, 2021.
- [2] L. B. Rahayu, “Analisis Penerapan E-Commerce dalam Bisnis UMKM,” Jurnal Manajemen Bisnis, vol. 9, no. 1, pp. 12–20, 2021.
- [3] Emantonio, et al., “Pertumbuhan E-Commerce dan Perubahan Gaya Hidup Masyarakat Digital,” Jurnal Ekonomi Digital Indonesia, vol. 5, no. 1, pp. 33–41, 2025.
- [4] D. Juhana, A. T. Komara, and I. Sidharta, “Kontribusi E-Commerce terhadap Pertumbuhan Ekonomi Domestik,” Jurnal Ekonomi dan Bisnis, vol. 10, no. 3, pp. 77–89, 2024.
- [5] Prasetyo, “Dampak E-Commerce terhadap Perekonomian Global,” Jurnal Bisnis Internasional, vol. 8, no. 2, pp. 55–64, 2023.
- [6] Wardhono, “E-Government sebagai Metode Transaksi Publik,” Jurnal Sistem Pemerintahan Digital, vol. 6, no. 1, pp. 21–31, 2023.
- [7] Syaepudin and D. Nurlukman, “Model E-Government dalam Pengadaan Barang dan Jasa,” Prosiding Seminar Nasional Teknologi Informasi, pp. 102–109, 2022.
- [8] H. Halim and A. Tyra, “Analisis Review Pelanggan sebagai Umpan Balik Kualitas Layanan,” Jurnal Manajemen Pemasaran, vol. 9, no. 1, pp. 17–25, 2020.
- [9] Yendrizal, “Klasifikasi Data Mining: Perbandingan ID3, C4.5, dan CART,” Jurnal Ilmu Komputer, vol. 4, no. 2, pp. 66–75, 2022.
- [10] Jollyta, et al., “Implementasi Algoritma C4.5 dalam Klasifikasi Data,” Jurnal Teknologi dan Sistem Informasi, vol. 11, no. 2, pp. 115–124, 2023.
- [11] Oktalia and Zakaria, “Perbandingan C4.5 dan Naïve Bayes dalam Analisis Review JNE di Google Playstore,” Jurnal Sistem Informasi, vol. 13, no. 1, pp. 44–53, 2024.
- [12] N. Angie, et al., “Analisis Faktor Review dan Rekomendasi pada Marketplace Shopee dengan Algoritma C4.5,” Jurnal Sistem Informasi, vol. 12, no. 3, pp. 88–97, 2025.
- [13] Uape and Elisa, “Decision Tree pada Penilaian Layanan Driver Gojek dengan C4.5,” Jurnal Informatika dan Sistem Cerdas, vol. 9, no. 1, pp. 15–23, 2023.
- [14] Siahaan and R. Muliono, “Klasifikasi Kualitas Kopi Berdasarkan Review Pelanggan dengan C4.5,” Jurnal Teknologi Pangan Digital, vol. 7, no. 2, pp. 50–59, 2024.
- [15] Viska and Elisa, “Penerapan Algoritma C4.5 dalam Analisis Kepuasan Pelanggan Fashion Thrift,” Jurnal Sistem Informasi, vol. 8, no. 4, pp. 91–99, 2023.

NOMENKLATUR

S = Himpunan data (dataset keseluruhan)

A = Atribut (variabel) yang digunakan dalam klasifikasi

n = Jumlah kelas dalam data

p_i = Probabilitas proporsi kelas ke-i dalam dataset

S_i = Subset data ke-i berdasarkan atribut A

$|S_i|$ = Jumlah data pada subset ke-i

$|S|$ = Jumlah data total dalam himpunan S

Entropy(S) = Nilai ketidakpastian atau impurity dalam himpunan data S

Entropy(S_i) = Nilai entropy pada subset data ke-i

Gain(S,A) = Nilai information gain atribut A terhadap dataset S

Node = Simpul pada pohon keputusan (decision tree)

Leaf = Daun atau simpul akhir pada pohon keputusan (hasil klasifikasi)

Review = Umpan balik atau komentar pelanggan setelah transaksi

Klasifikasi = Proses pengelompokan data ke dalam kelas tertentu (Puas/Tidak Puas)

Decision Tree = Struktur pohon keputusan yang digunakan dalam klasifikasi

RapidMiner = Perangkat lunak data mining yang digunakan untuk membangun model klasifikasi