

Implementation of Data Mining Using the C4.5 Algorithm to Predict Customer Satisfaction

Muhamad Ryansyah¹⁾, Priatno²⁾, Muhammad Ridwan Effendi^{3)*}

¹⁾Informatika, Fakultas Teknologi Informatika, Universitas Bina Sarana Informatika

²⁾Teknik Industri Universitas Bina Sarana Informatika

³⁾Sistem Informasi, Fakultas Komputer, Universitas Mohammd Husni Thamrin

***)Correspondence author:** Muhammad Ridwan Effendi, ridwan.oke79bgt@gmail.com, Indonesia

DOI : <https://doi.org/10.37012/jtik.v12i1.3364>

Abstract

The C4.5 algorithm is one of many data mining methods that can generate rules for analyzing customer satisfaction. This process produces approximately four rules with influential attributes, namely Security, Timeliness, and Customer Service. The rapid changes in information technology have had a major impact on many aspects of life, including the business sector. One of the main transformations is the shift from traditional sales systems to digital ones, known as e-commerce. Tight competition in sales and maintaining customer loyalty makes customer satisfaction a crucial aspect to maintain good customer relationships. This research is expected to improve customer satisfaction and address issues related to it. In this context, the application of data mining is considered an appropriate tool for processing data in research to predict customer satisfaction using the C4.5 algorithm, as well as to test and validate information regarding customer satisfaction. The results of the research that has been carried out show that comparative testing on the classification model with the C4.5 algorithm produces a high level of accuracy in predicting data, namely with an accuracy value of 98.13%. For the recall test value, a value of 100.00% was obtained, while the precision value was recorded at 91.00%, and the ROC/AUC value reached 0.695%.

Keywords: Classification, Customer Satisfaction, C4.5 Algorithm (Decision Tree).

Abstrak

Algoritma C4.5 adalah merupakan salah satu dari banyak metode dalam penambangan data yang dapat menghasilkan aturan untuk menganalisis mutu kepuasan pelanggan. Dari proses ini, menghasilkan sekitar terdapat 4 aturan yang terbentuk dengan atribut yang berpengaruh, yaitu Keamanan, Ketepatan Waktu, dan dalam hal Pelayanan kepada pelanggan. Perubahan teknologi informasi yang terjadi dengan cepat telah memberikan pengaruh besar terhadap banyak aspek kehidupan, termasuk sektor bisnis. Salah satu transformasi utama adalah beralihnya sistem penjualan dari cara tradisional menuju digital, yang disebut dengan e-commerce. Ketatnya persaingan dalam penjualan dan menjaga loyalitas pelanggan membuat kepuasan pelanggan menjadi hal yang penting untuk diperhatikan agar hubungan kerjasama dengan pelanggan tetap terjalin dengan baik. Penelitian ini diharapkan dapat meningkatkan kepuasan pelanggan dan mengatasi masalah yang berkaitan dengan kepuasan tersebut. Dalam konteks ini, penerapan data mining dianggap sebagai alat yang tepat untuk mengolah data dalam penelitian guna memprediksi kepuasan pelanggan dengan menggunakan algoritma C4.5, serta untuk menguji dan memvalidasi informasi mengenai kepuasan pelanggan. Hasil dari penelitian yang telah dilaksanakan menunjukkan bahwa pengujian komparatif pada model klasifikasi dengan algoritma C4.5 menghasilkan tingkat akurasi yang tinggi dalam memprediksi data, yakni dengan nilai akurasi sebesar 98,13%. Untuk nilai uji recall, diperoleh nilai 100,00%, sedangkan nilai precision tercatat sebesar 91,00%, serta nilai ROC/AUC mencapai 0,695%.

Kata Kunci : Klasifikasi, Kepuasan Pelanggan, Algoritma C4.5 (Decision Tree).

PENDAHULUAN

Perubahan teknologi informasi yang terjadi dengan cepat telah memberikan pengaruh besar terhadap banyak aspek kehidupan, termasuk sektor bisnis. Salah satu transformasi utama adalah beralihnya sistem penjualan dari cara tradisional menuju digital, yang disebut dengan e-commerce. (Patria Adiguna, Fajar Sasongko, Nurhayati, Firmansyah, 2025). Persaingan yang sengit dalam dunia penjualan dan banyaknya kemiripan produk bisa menjadi rintangan yang signifikan, sehingga perusahaan harus memiliki pendekatan pemasaran yang tepat dan kreatif untuk mencapai target pasar dengan efisien dan menciptakan citra yang baik serta positif di mata konsumen. (Suryanti et al., 2025). Era digital telah merombak cara-cara perusahaan dalam membangun koneksi dengan konsumen, sehingga metode konvensional digantikan oleh sistem yang otomatis dan terintegrasi.

Konsumen kini diarahkan untuk berkomunikasi melalui kanal digital yang memanfaatkan teknologi informasi dan Manajemen Hubungan Pelanggan (CRM), sehingga perilaku data dapat dicatat secara real-time. (Syarif, 2025) Kondisi pertumbuhan usaha yang sangat cepat dan persaingan yang kian rumit, menuntut perusahaan untuk dapat menjaga keberlangsungan hidupnya baik dalam waktu dekat maupun di masa depan untuk menghadapi kompetisi global. (Mtw, 2020). Di Indonesia, platform e-commerce mengalami pertumbuhan yang sangat cepat berkat kemudahan dan biaya yang terjangkau. Munculnya e-commerce memperkenalkan sebuah fenomena atau gaya hidup terbaru kepada masyarakat, di mana ketika konsumen ingin berbelanja, mereka lebih memilih untuk menghabiskan waktu meneliti produk secara daring daripada pergi langsung ke toko di pasar tradisional. (Nur'aeni et al., 2024).

Pentingnya interaksi dalam menyampaikan informasi kepada target pelanggan adalah agar dapat menarik minat konsumen untuk membeli barang yang disediakan. Di Indonesia, diperkirakan akan terjadi peningkatan jumlah lansia pada tahun 2050, sehingga komunikasi dengan kelompok lansia ini perlu dilakukan mulai sekarang. (Nana Noviada Kwartawat, Deasy Virka Sari, 2023). Kepuasan pengguna adalah penanda paling utama dari

kesuksesan suatu institusi atau bisnis. Dalam zaman teknologi komunikasi yang semakin maju, perasaan puas atau tidak puas pelanggan dapat dengan mudah terbentuk dan berdampak pada pasar. Dunia maya menawarkan kesempatan bagi pelanggan untuk menyebarkan informasi negatif atau positif dengan cepat ke seluruh penjuru dunia. (Apriasty & Simbolon, 2022). Kepuasan dari konsumen timbul ketika ada suatu perbandingan antara apa yang mereka harapkan dan apa yang mereka alami. Jika hasil yang didapat sesuai atau lebih baik dari yang diantisipasi, maka konsumen akan merasakan kepuasan, tetapi jika tidak, kekecewaan dapat muncul. Bagi para pelaku usaha, kemampuan untuk memenuhi atau bahkan melebihi harapan pelanggan merupakan aspek penting dalam mempertahankan kelangsungan usaha. Dua faktor yang sangat berpengaruh dalam hal ini adalah lokasi usaha dan kualitas layanan. Posisi yang strategis sering kali menjadi pertimbangan utama bagi pelanggan dalam memilih tempat untuk berbelanja. (Agnesta et al., 2025).

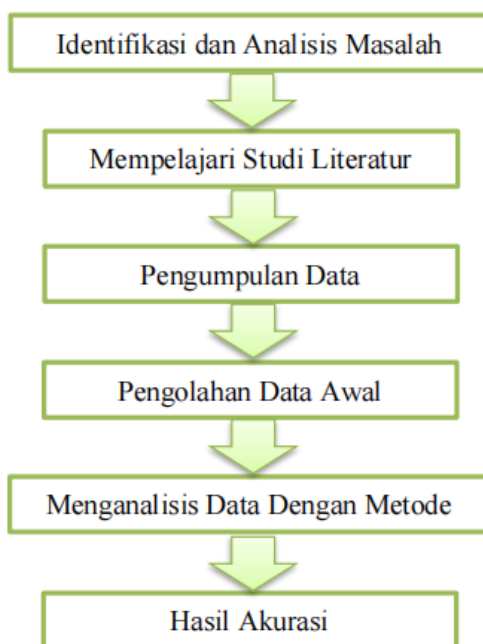
Pelanggan dianggap sebagai raja yang perlu diperlakukan secara baik. Namun, hal ini tidak berarti bahwa segala sesuatu diserahkan kepada pelanggan secara mutlak. Perusahaan perlu memenuhi kebutuhan pelanggan dengan pendekatan yang saling menguntungkan. Kepuasan pelanggan merupakan faktor penting dalam menjaga keberlangsungan usaha. (Fahleti & Badak, 2025). Analisis data bukan hanya sekedar pengolahan informasi, tetapi juga melibatkan penentuan pola, ramalan untuk masa mendatang, serta memahami perilaku berdasarkan informasi yang tersedia. Dalam praktiknya, analisis data mencakup beragam metode dan proses yang meliputi pengumpulan informasi, pembersihan data, analisis statistik, dan penerapan algoritma kecerdasan buatan atau pembelajaran mesin untuk memperoleh wawasan yang lebih mendalam dan berarti. Sebagai ilustrasi, di sektor bisnis, analisis data dapat dimanfaatkan untuk memahami tingkah laku pelanggan, meramalkan kebutuhan produk, dan juga untuk meningkatkan efisiensi dalam operasi. (Rismaninda Putri Dwi Prasetya et al., 2024).

Metode Algoritma C4.5 adalah lanjutan dari algoritma ID3. Dengan adanya pengembangan ini, algoritma C4.5 tetap mengikuti prinsip dasar yang serupa dengan yang ada pada ID3. Namun, dalam algoritma C4.5, langkah pertama adalah menghitung nilai entropy, kemudian seleksi atribut dilakukan berdasarkan nilai gain ratio yang paling

tinggi.(Sahadan et al., 2023). Algoritma C4.5 adalah merupakan salah satu dari banyak metode dalam penambangan data yang dapat menghasilkan aturan untuk menganalisis mutu kepuasan pelanggan. Dari proses ini, menghasilkan sekitat terdapat 4 aturan yang terbentuk dengan atribut yang berpengaruh, yaitu Keamanan, Ketepatan Waktu, dan dalam hal Pelayanan kepada pelanggan. Aturan pertama menyatakan bahwa jika Keamanan dianggap tidak aman, maka pelanggan pada akhirnya akan merasa Tidak Puas. Aturan kedua menyatakan jika Keamanan dinyatakan aman, tetapi Ketepatan Waktu dan Pelayanan kurang baik, maka pelanggan tetap akan merasa Tidak Puas.(Nurul Azwanti et al., 2023).

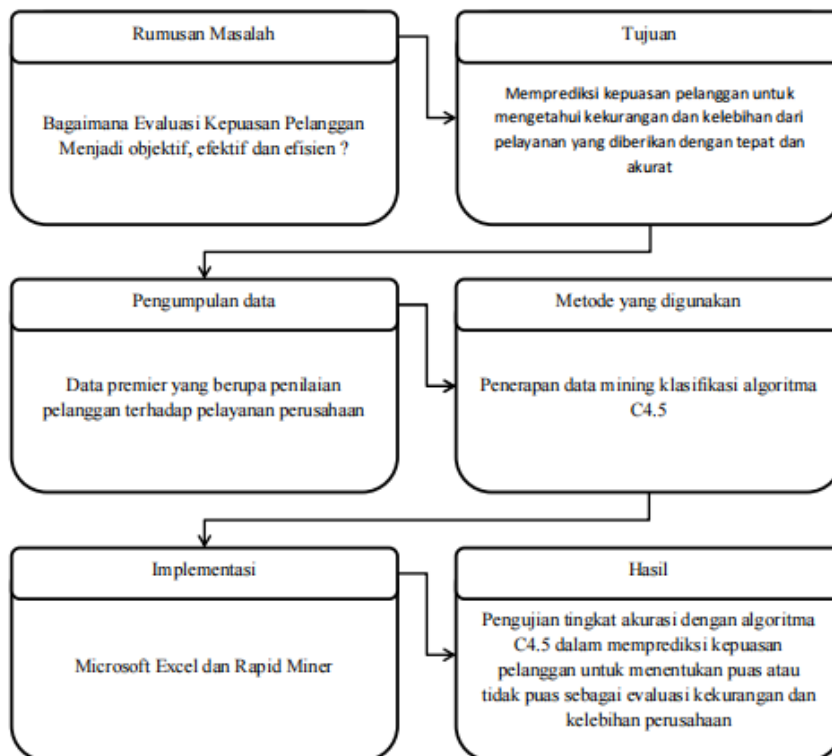
METODE PENELITIAN

Kerangka penelitian adalah struktur terencana yang diterapkan dalam sebuah studi untuk menghasilkan pola dan susunan penyusunan yang teratur. Penelitian yang bertujuan untuk memperkirakan tingkat kepuasan pelanggan adalah penelitian yang bersifat eksperimen dengan penerapan metode dan teori atau pola yaitu dengan algoritma C4.5 yang kemudian diuji menggunakan perangkat Rapid Miner Studio.



Gambar 1. Kerangka penelitian

Untuk meningkatkan validitas model, penelitian ini menerapkan teknik k-fold cross validation (k=10) guna mengurangi bias pada pembagian data training dan testing. Selain itu, evaluasi model dilakukan menggunakan Confusion Matrix yang menghasilkan metrik performa berupa accuracy, precision, recall, dan F1-score.



Gambar 2. Rumusan Masalah

Berdasarkan hasil pengujian menggunakan confusion matrix, diperoleh nilai: Accuracy: 98,13% , Precision: 91,00% , Recall: 100% , F1-Score: 95,28% . Hasil ini menunjukkan bahwa model memiliki performa yang sangat baik dalam mengklasifikasikan tingkat kepuasan pelanggan. Model C4.5 yang dihasilkan divisualisasikan dalam bentuk pohon keputusan (decision tree) untuk memudahkan interpretasi aturan klasifikasi. Berdasarkan pohon keputusan, atribut yang paling dominan dalam menentukan kepuasan pelanggan adalah: K8 (Jaminan Keamanan) , K6 (Jasa Pengiriman) dan K9 (Respon Pelayanan). Visualisasi ini menunjukkan bahwa faktor keamanan dan pelayanan menjadi penentu utama kepuasan pelanggan. Untuk meningkatkan kualitas penelitian, dilakukan

perbandingan model dengan beberapa algoritma klasifikasi lain, yaitu: Random Forest dan Support Vector Machine (SVM). Pada Susunan Penelitian ini dibutuhkan adanya sebuah kerangka pemikiran yang melandasi sebagai pedoman agar tahapan penelitian dapat lebih konsisten. Pada studi ini, informasi yang dipakai berasal dari data yang didapatkan secara langsung dalam format dokumen excel. Data yang digunakan dalam penelitian ini diperoleh dari 335 responden pelanggan yang berasal dari platform e-commerce yang beroperasi di Indonesia. Responden dalam penelitian ini merupakan pengguna aktif layanan e-commerce yang pernah melakukan transaksi pembelian minimal satu kali dalam kurun waktu 6 bulan terakhir.

Secara geografis, responden berasal dari berbagai wilayah di Indonesia dengan dominasi dari wilayah Jabodetabek dan Jawa Barat, yang merupakan pusat aktivitas perdagangan digital. Dari sisi karakteristik pekerjaan, responden terdiri dari pelajar/mahasiswa, karyawan swasta, wiraswasta, serta profesi lainnya yang menggunakan layanan e-commerce dalam aktivitas sehari-hari. Sektor bisnis yang diteliti berfokus pada platform e-commerce retail (B2C) yang menyediakan berbagai produk kebutuhan sehari-hari, seperti fashion, elektronik, dan kebutuhan rumah tangga. Pengumpulan data dilakukan melalui dua metode, yaitu: (1) Wawancara dengan pihak manajemen/administrator platform untuk memperoleh gambaran yang detail jelas dan terinci mengenai atau dalam hal sistem layanan, dan (2) Penyebaran kuesioner kepada pelanggan menggunakan skala Likert 1–5 untuk mengukur tingkat kepuasan terhadap 10 atribut layanan.. Dengan penjelasan ini, dataset yang digunakan dalam penelitian memiliki karakteristik yang jelas, terukur, dan dapat dipertanggungjawabkan secara ilmiah.

Data yang diperoleh berasal dari seluruh publikasi penelitian yang membahas tentang CRISP-DM. Pembahasan dalam publikasi ini bisa meliputi teori, tinjauan pustaka, hasil penelitian, dan lain-lain. Referensi yang digunakan diambil dari artikel jurnal, prosiding, buku, skripsi, dan sumber lainnya yang relevan untuk menjawab pertanyaan penelitian. (Sahadan et al., 2023). Dalam kajian ini, metode yang diterapkan ialah Cross Industry Standard Model for Data Mining (CRISP-DM) yang merupakan pendekatan data science yang dihasilkan oleh IBM.(Abdusyukur, 2023).



Gambar 3. Data

Pembersihan data dan pengurangan data sangat penting untuk mendapatkan representasi dalam bentuk volume, seluruh atribut yang telah melalui proses pembersihan data.

Tabel 1. Pembersihan data dan pengurangan data

| Atribut | Proses | Penjelasan |
|---------------------|---------------------------|-------------------------------|
| Nama | Data Cleaning | Nama Responden |
| Jenis Kelamin | Data Cleaning | Jenis Kelamin Responden |
| Usia | Data Cleaning | Usia Responden |
| Pekerjaan | Data Cleaning | Pekerjaan Responden |
| Pendapatan | Data Cleaning | Pendapatan Responden |
| Berapa Kali Belanja | Data Cleaning | Berapa Kali Responden Belanja |
| Kriteria 1 | Digunakan Sebagai Atribut | Kemudahan Penggunaan Website |
| Kriteria 2 | Digunakan Sebagai Atribut | Metode Pembayaran |
| Kriteria 3 | Digunakan Sebagai Atribut | Harga Produk |
| Kriteria 4 | Digunakan Sebagai Atribut | Kualitas Produk |
| Kriteria 5 | Digunakan Sebagai Atribut | Varian Ragam Produk |
| Kriteria 6 | Digunakan Sebagai Atribut | Jasa Pengiriman |
| Kriteria 7 | Digunakan Sebagai Atribut | Ketersediaan Jasa Pengiriman |
| Kriteria 8 | Digunakan Sebagai Atribut | Jaminan Keamanan |
| Kriteria 9 | Digunakan Sebagai Atribut | Respon Pelayanan |
| Kriteria 10 | Digunakan Sebagai Atribut | Informasi Promosi |
| Remark | Digunakan Sebagai Label | Kesimpulan Puas / Tidak Puas |

Data yang telah selesai melalui tahap pembersihan selanjutnya adalah pengurangan data yang menyederhanakan fitur dengan cara mengurangi data. Fitur yang telah dikurangi tersebut menjadi Kriteria 1, Kriteria 2, Kriteria 3, Kriteria 4, Kriteria 5, Kriteria 6, Kriteria 7, Kriteria 8, Kriteria 9, Kriteria 10 dan atribut keputusan sebagai fitur terakhir. Berikut adalah tabel yang telah mengalami pengurangan data dengan cara mengurangi atribut melalui proses pembersihan data.

Tabel 2. Pengurangan Data Dengan Cara Mengurangi Atribut Melalui Proses Pembersihan

| Data | | |
|------------|--------|--|
| Atribut | Proses | Keterangan |
| Kriteria 1 | K1 | Skala Penilaian 1 = Sangat Tidak Puas 2 = Tidak Puas 3 = Kurang Puas 4 = Puas 5 = Sangat Puas |
| Kriteria 2 | K2 | Skala Penilaian 1 = Sangat Tidak Puas 2 = Tidak Puas 3 = Kurang Puas 4 = Puas 5 = Sangat Puas |
| Kriteria 3 | K3 | Skala Penilaian 1 = Sangat Tidak Puas 2 = Tidak Puas 3 = Kurang Puas 4 = Puas 5 = Sangat Puas |
| Kriteria 4 | K4 | Skala Penilaian 1 = Sangat Tidak Puas 2 = Tidak Puas 3 = Kurang Puas 4 = Puas 5 = Sangat Puas |

Berdasarkan informasi yang telah dikumpulkan seperti tabel di atas, setelah melalui tahap pembersihan data dan penyederhanaan data, langkah selanjutnya adalah agar proses klasifikasi data menjadi lebih mudah, kita olah dan siapkan menjadi dataset dengan memisahkan menjadi data pelatihan dan data pengujian. Umumnya, data pelatihan bertujuan untuk melatih algoritma dalam menemukan model yang sesuai dan sering digunakan sebagai alat untuk merepresentasikan data baru yang sebelumnya tidak ada. Sementara itu, data

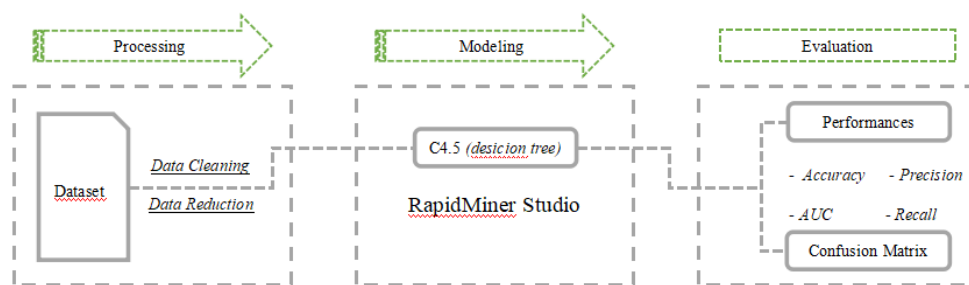
pengujian berfungsi untuk mengevaluasi dan mengukur kinerja model tersebut. Pemisahan ini dilakukan agar model klasifikasi dapat memiliki kemampuan generalisasi yang baik dalam melakukan klasifikasi data.

Tabel 3. Dataset Penilaian Kepuasan Pelanggan

| | K1 | K2 | K3 | K4 | K5 | K6 | K7 | K8 | K9 | K10 | Hasil |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|---------------|
| 1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | Puas |
| 2 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | Puas |
| 3 | 4 | 5 | 4 | 5 | 5 | 4 | 3 | 4 | 4 | 4 | Puas |
| 4 | 3 | 3 | 3 | 4 | 5 | 4 | 4 | 5 | 5 | 4 | Puas |
| 5 | 5 | 5 | 3 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | Puas |
| 6 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | Puas |
| 7 | 3 | 5 | 3 | 4 | 5 | 3 | 3 | 4 | 3 | 4 | Puas |
| 8 | 2 | 4 | 2 | 4 | 4 | 4 | 3 | 3 | 3 | 4 | Tidak Puas |
| 9 | 3 | 5 | 4 | 5 | 5 | 4 | 4 | 4 | 5 | 5 | Puas |
| 10 | 4 | 5 | 4 | 5 | 5 | 4 | 4 | 5 | 5 | 5 | Puas |
| 11 | 3 | 5 | 4 | 4 | 5 | 5 | 4 | 4 | 5 | 5 | Puas |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 328 | 5 | 5 | 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | Puas |
| 329 | 3 | 3 | 2 | 3 | 4 | 3 | 2 | 3 | 3 | 3 | Tidak Puas |

Hasil pengolahan data digunakan sebanyak 10 atribut yaitu K1, K2, K3, K4, K5, K6, K7, K8, K9, K10 dan rekomendasi sebagai label. Sedangkan nomor yang terdapat pada kolom pertama pada tabel hanya sebagai tanda jumlah record. selanjutnya akan dilakukan pemodelan untuk mengklasifikasi dengan menggunakan algoritma sebagai komparasi, yaitu C4.5 (decision tree) yang berperan untuk membuat model atau rule yang dipresentasikan dalam bentuk gambar yang merupakan pohon keputusan, yang kemudian data akan diuji menggunakan aplikasi Rapid Miner Studio.

Desain pemodelan ditunjukkan pada gambar 4 dibawah ini:



Gambar 4. Model Yang Diusulkan

Berdasarkan penggunaan model seperti diatas, maka akan dilakukan pengujian terhadap model yang bertujuan dapat memperoleh model yang paling akurat, Sehingga tingkat akurasi dari komparasi yang dilakukan dapat langsung dilihat.

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan 335 sampel dataset, yang kemudian dibagi menjadi dua subset: data training dan data testing. Pada Tabel IV.1 ditampilkan sampel data yang diambil dari dataset penilaian kepuasan pelanggan, dengan 231 data digunakan sebagai data training untuk diklasifikasikan menggunakan algoritma C4.5.

Tabel 4. Sampel Data Training Penilaian Kepuasan Pelanggan

| No. | K1 | K2 | K3 | K4 | K5 | K6 | K7 | K8 | K9 | K10 | Hasil |
|-----|----|----|----|----|----|----|----|----|----|-----|-------|
| 1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | Puas |
| 2 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | Puas |
| 3 | 4 | 5 | 4 | 5 | 5 | 4 | 3 | 4 | 4 | 4 | Puas |
| 4 | 3 | 3 | 3 | 4 | 5 | 4 | 4 | 5 | 5 | 4 | Puas |
| 5 | 5 | 5 | 3 | 5 | 5 | 4 | 5 | 5 | 5 | 5 | Puas |
| 6 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | Puas |
| 7 | 5 | 5 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 5 | Puas |
| 8 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | Puas |
| 9 | 4 | 4 | 4 | 5 | 5 | 4 | 3 | 4 | 4 | 4 | Puas |
| 10 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | Puas |
| 11 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | Puas |
| .. | .. | .. | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| .. | .. | .. | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| 230 | 5 | 3 | 3 | 5 | 4 | 5 | 4 | 3 | 4 | 2 | Puas |
| 231 | 5 | 5 | 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | Puas |

Langkah pertama adalah menghitung jumlah total kasus dari data, yakni membagi antara kasus puas dan tidak puas. Selanjutnya, hitung nilai entropy untuk seluruh kasus yang dikelompokkan berdasarkan atribut K1 hingga K10. Setelah mendapatkan nilai entropy untuk tiap atribut, lanjutkan dengan menghitung nilai gain untuk setiap atribut tersebut, lalu tentukan atribut dengan nilai gain tertinggi dari hasil perhitungan.

Setelah memperoleh jumlah kasus dari setiap atribut seperti table diatas selanjutnya langkah untuk mencari perhitungan nilai entropy dengan pencarian akar (node).

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

$$\begin{aligned} \text{Entropy (Total)} &= \left(-\frac{218}{231} * \log_2 \left(\frac{218}{231}\right)\right) + \left(-\frac{13}{231} * \log_2 \left(\frac{13}{231}\right)\right) \\ &= (-0,944 * \log_2(0,944)) + (-0,056 * \log_2(0,056)) \\ &= (-0,944 * (-0,084)) + (-0,056 * (-4,151)) \\ &= 0,079 + 0,234 \\ &= 0,312. \end{aligned}$$

Hasil perbandingan performa algoritma ditunjukkan pada tabel berikut:

Tabel 4. Perbandingan Performa Algoritma

| Algoritma | Accuracy | Precision | Recall |
|---------------|----------|-----------|--------|
| C4.5 | 98,13% | 91% | 100% |
| Random Forest | 97,45% | 90% | 98% |
| SVM | 96,80% | 89% | 97% |

KESIMPULAN DAN REKOMENDASI

Penelitian ini bertujuan untuk memprediksi tingkat kepuasan pelanggan pada platform e-commerce dengan memanfaatkan teknik data mining menggunakan algoritma C4.5. Berdasarkan hasil pengujian model, diperoleh tingkat akurasi sebesar 98,13% dengan nilai recall 100% dan precision 91%, yang menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik dalam mengidentifikasi pelanggan puas dan tidak puas. Namun demikian, capaian akurasi yang tinggi tersebut perlu ditinjau secara kritis. Tanpa adanya validasi model yang komprehensif seperti k-fold cross validation, terdapat potensi terjadinya

overfitting, di mana model terlalu menyesuaikan diri dengan data pelatihan sehingga kurang optimal dalam melakukan generalisasi terhadap data baru. Selain itu, distribusi data yang kemungkinan tidak seimbang antara kelas “puas” dan “tidak puas” juga dapat mempengaruhi tingginya nilai akurasi yang diperoleh. Oleh karena itu, interpretasi terhadap hasil performa model perlu dilakukan secara hati-hati dan tidak hanya bergantung pada satu metrik evaluasi.

Dari hasil analisis pohon keputusan, diketahui bahwa atribut seperti jaminan keamanan, kualitas pelayanan, dan ketepatan pengiriman menjadi faktor dominan yang mempengaruhi tingkat kepuasan pelanggan. Temuan ini memberikan kontribusi praktis bagi pelaku bisnis e-commerce dalam merancang strategi peningkatan layanan yang lebih tepat sasaran. Meskipun demikian, penelitian ini memiliki beberapa keterbatasan, antara lain keterbatasan pada jumlah dan cakupan responden yang belum sepenuhnya merepresentasikan seluruh populasi pengguna e-commerce di Indonesia, serta penggunaan satu algoritma utama tanpa eksplorasi mendalam terhadap teknik optimasi model.

Oleh karena itu, penelitian selanjutnya disarankan untuk mengembangkan pendekatan yang lebih komprehensif, seperti penerapan cross validation, penggunaan perbandingan multi-algoritma (misalnya Random Forest dan Support Vector Machine), serta integrasi metode feature selection atau optimasi parameter guna meningkatkan performa dan robustness model. Selain itu, perlu dilakukan pengujian pada dataset yang lebih besar dan beragam agar hasil penelitian memiliki tingkat generalisasi yang lebih tinggi. Secara keseluruhan, penelitian ini memberikan kontribusi awal dalam penerapan algoritma klasifikasi untuk analisis kepuasan pelanggan, namun masih memerlukan pengembangan lebih lanjut agar dapat memberikan hasil yang lebih robust, generalizable, dan bernilai ilmiah tinggi dalam bidang data mining

REFERENSI

Abdusyukur, F. (2023). KOMPUTA : Jurnal Ilmiah Komputer dan Informatika Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Pencemaran Nama Baik Komputa : Jurnal Ilmiah Komputer dan Informatika. Komputa : Jurnal Ilmiah Komputer Dan Informatika, 12(1), 73–82.

- Agnesta, K., Alfiani, R. Y., Erlita, R., & Hanafi, M. I. (2025). Pengaruh Lokasi , Kualitas Pelayanan Terhadap Kepuasan Pelanggan Pada Warung Sembako Harby. *Jurnal Mirai Management*, 10(1), 559–574. <https://journal.stieamkop.ac.id/index.php/mirai/article/download/9271/5849>
- Apriasty, I., & Simbolon, M. E. (2022). Faktor-faktor yang mempengaruhi kepuasan pelanggan: kualitas produk, kualitas pelayanan dan harga (literature strategic marketing management). *Jurnal Ilmu Multidisiplin*, 1(1), 135–145. <https://doi.org/10.38035/jim.v1i1>
- Fahleti, W. H., & Badak, M. (2025). Analisis Tingkat Kepuasan Pelanggan Terhadap Kualitas Pelayanan Pada Toko Plastik Rizka Di Tenggarong. *Jurnal Ekonomi Manajemen Dan Bisnis*, 5(1), 1–10. <https://journal.stietenggarong.ac.id/index.php/ICM/article/download/101/61>
- Mtw, E. (2020). Strategi Pemasaran Untuk Meningkatkan Volume Penjualan Sepeda Motor Pada CV Trio Motor Muara Teweh. *Jurnal Ilmu Ekonomi (Manajemen Perusahaan) Dan Bisnis*, 4(01). <https://doi.org/10.51512/jimb.v4i01.46>
- Nana Noviada Kwartawat, Deasy Virka Sari, S. T. W. (2023). ANALISIS KOMUNIKASI PEMASARAN BAGI PASAR LANJUT USIA. *Gema Ekonomi (Jurnal Fakultas Ekonomi)*, 12(2), 2621–0444. <https://journal.unigres.ac.id/index.php/GemaEkonomi/index>
- Nur'aeni, N. N., Ainulyaqin, M., & Edy, S. (2024). Dampak Fenomena E-Commerce Pada Tingkat Penjualan Di Pasar Tradisional Ditinjau Dari Psikologi Dan Ekonomi Islam. *Jurnal Ilmiah Ekonomi Islam*, 10(1), 270. <https://doi.org/10.29040/jiei.v10i1.12146>
- Nurul Azwanti, Sri Tria Siska, & Liga Mayola. (2023). Kepuasan Pelanggan Terhadap Jasa Pengiriman Barang Menggunakan Metode Algoritma C4.5. *Ensiklopedia of Journal*, 5(2), 1–8. <http://jurnal.ensiklopediaku.org>
- Patria Adiguna, Fajar Sasongko, Nurhayati, Firmansyah, M. F. (2025). OPTIMALISASI PENGGUNAAN AFFILIATE MARKETING TERHADAP BUSINESS TO BUSINESS (B2B). *Jurnal Pengabdian Kepada Masyarakat*, 3, 1–9. <https://jurnal.unisa.ac.id/index.php/jpkmunisa/article/download/771/584>
-
- <https://journal.thamrin.ac.id/index.php/jtik/article/view/3346/2806>

-
- Rismaninda Putri Dwi Prasetya, Azizah, R. N., Halwa, J. B. W., Nugroho, R. H., & Kusumasari, I. R. (2024). Implementasi Penggunaan Data Analytics untuk Mengoptimalkan Pengambilan Keputusan Bisnis di Era Digital. *Jurnal Bisnis Dan Komunikasi Digital*, 2(2), 12. <https://doi.org/10.47134/jbkd.v2i2.3459>
- Sahadan, M. A., Rompas, P., & Munaiseche, C. (2023). Analisis Kepuasan Pelanggan Menggunakan Metode Algoritma C4.5 Berdasarkan E-Survey Kejaksaaan Negeri Minahasa. *Journal of Science and Social Research*, 4(2), 227. <https://doi.org/10.54314/jssr.v4i2.647>
- Suryanti, E., Zulfa, F., Lesmana, H., Cahya, B., & Utami, P. (2025). MANABIS (Jurnal Manajemen dan Bisnis) Pemanfaatan Tehnologi Digital dalam Strategi Pemasaran untuk Meningkatkan Penjualan pada UMKM. *Media Eletronik*, 4(3), 256–265. <https://doi.org/10.54259/manabis.v4i3.5411>
- Syarif, Y. (2025). Penerapan Digital Customer Relationship Management Terhadap Pengalaman Dan Loyalitas Pelanggan. *Jurnal Ekonomi, Manajemen Dan Akuntansi*, 4(2024), 232–241. <https://jurnal.anfa.co.id/index.php/mufakat/article/download/2890/2599/8869>