

Klasifikasi Minat Beli Konsumen pada Platform E-Commerce Menggunakan Algoritma C4.5

Rafli Kalisiano Kalkausar^{1,*}, Aura Unga Safitri², Muhammad Dzaki Fadhli Sugianto³

¹Teknik dan Informatika, Sistem Informasi, Universitas Bina Sarana Informatika, Pontianak, Indonesia

Email: ^{1,*}raflikalisiano3@gmail.com, ²Agnuarua13@gmail.com, ³dzakifadhli1255@gmail.com,

^{*)} Email Penulis Utama

Abstrak– Penelitian ini menerapkan teknik data mining berbasis klasifikasi untuk mengidentifikasi pola minat beli konsumen pada platform e-commerce. Data penelitian diperoleh melalui kuesioner terstruktur kepada pengguna aktif e-commerce dengan skala Likert 1–5, yang mengukur empat variabel utama, yaitu harga, keragaman produk, kemudahan penggunaan, dan kepercayaan. Sebelum pemodelan, seluruh instrumen kuesioner diuji validitas menggunakan korelasi Pearson dan reliabilitas menggunakan Cronbach's Alpha, dengan hasil seluruh butir pernyataan dinyatakan valid dan reliabel. Data yang terkumpul diseleksi, dikodekan menjadi skor numerik, dan dilabeli ke dalam tiga kategori minat beli yaitu Tinggi, Sedang, dan Rendah berdasarkan skor komposit rata-rata keempat variabel. Pemodelan klasifikasi dilakukan menggunakan algoritma C4.5 dan Naive Bayes melalui perangkat lunak RapidMiner, dengan evaluasi menggunakan split validation (70:30) dan 10-fold cross validation. Hasil evaluasi menunjukkan bahwa C4.5 menghasilkan akurasi 83,87% pada split validation dan 91,51% pada cross validation, sementara Naive Bayes menghasilkan akurasi 96,77% dan 95,36% pada masing-masing metode evaluasi. Keunggulan Naive Bayes atas C4.5 turut dipengaruhi oleh kondisi ketidakseimbangan kelas pada dataset, di mana kelas Rendah hanya diwakili oleh 5 data dari total 106 responden. Meskipun demikian, C4.5 tetap unggul dalam interpretabilitas karena struktur pohon keputusan yang dihasilkan memungkinkan penelusuran aturan klasifikasi secara eksplisit dalam bentuk aturan if-then yang dapat langsung diterapkan dalam pengambilan keputusan strategis.

Kata Kunci: Algoritma C4.5, Decision Tree, Naive Bayes, Minat Beli, E-commerce, Data Mining.

Abstract– This study applies a classification-based data mining approach to identify consumer purchase intention patterns on e-commerce platforms. The dataset was collected through a structured questionnaire administered to active e-commerce users using a 1–5 Likert scale, covering four key variables: price, product variety, ease of use, and trust. Prior to modeling, all questionnaire instruments were tested for validity using Pearson correlation and for reliability using Cronbach's Alpha, with all items confirmed as valid and reliable. The collected data were screened, encoded into numerical scores, and labeled into three purchase intention categories namely High, Medium, and Low based on the composite average score of the four variables. Classification modeling was performed using the C4.5 and Naive Bayes algorithms implemented in RapidMiner, with evaluation conducted using split validation (70:30) and 10-fold cross validation. Evaluation results show that C4.5 achieved an accuracy of 83.87% on split validation and 91.51% on cross validation, while Naive Bayes achieved 96.77% and 95.36% respectively. The superiority of Naive Bayes over C4.5 was partly influenced by class imbalance in the dataset, where the Low class was represented by only 5 out of 106 respondents. Nevertheless, C4.5 remains advantageous in terms of interpretability, as the resulting decision tree structure enables explicit tracing of classification rules in the form of if-then rules that can be directly applied in strategic decision-making.

Keywords: C4.5 Algorithm, Decision Tree, Naive Bayes, Purchase Intention, E-commerce, Data Mining.

1. PENDAHULUAN

Persaingan bisnis di industri *e-commerce* semakin ketat seiring pesatnya perkembangan teknologi digital di Indonesia [1]. Transformasi digital mendorong pergeseran pola konsumsi masyarakat dari transaksi konvensional ke transaksi berbasis platform *online*, yang menawarkan kecepatan, kemudahan, serta kenyamanan dalam proses jual beli. Berbagai platform *e-commerce* berlomba-lomba menghadirkan inovasi fitur, program promosi, dan layanan yang berorientasi pada pengalaman pengguna untuk menarik dan mempertahankan konsumen. Namun, di tengah pertumbuhan tersebut, fenomena tingginya variasi minat beli konsumen menunjukkan bahwa tidak semua konsumen memberikan respons yang sama terhadap penawaran yang disediakan. Sebagian konsumen menunjukkan minat beli yang tinggi dan cenderung loyal, sementara sebagian lainnya bersikap ragu, sensitif terhadap perubahan harga, atau mudah berpindah ke platform pesaing.

Kondisi ini memunculkan tantangan yang signifikan bagi pelaku usaha, terutama dalam memahami perilaku dan preferensi konsumen secara lebih akurat dan mendalam. Minat beli konsumen tidak hanya ditentukan oleh satu faktor tunggal, tetapi merupakan hasil interaksi kompleks dari berbagai aspek, seperti persepsi harga, kualitas dan keragaman produk, kemudahan penggunaan platform, keamanan transaksi, hingga tingkat kepercayaan konsumen terhadap platform yang digunakan. Dalam konteks persaingan yang ketat, pelaku usaha yang tidak mampu membaca dan mengolah data perilaku konsumen secara sistematis berisiko mengalami penurunan efektivitas strategi pemasaran, salah sasaran dalam penentuan segmen, serta inefisiensi dalam alokasi sumber daya promosi [2].

Pengumpulan data perilaku dan preferensi konsumen melalui kuesioner terstruktur menjadi salah satu pendekatan yang lazim digunakan untuk memperoleh gambaran persepsi konsumen secara terukur. Melalui instrumen kuesioner, peneliti dapat mengkuantifikasi sikap konsumen terhadap berbagai aspek platform *e-commerce*, mulai dari sensitivitas harga, preferensi keragaman produk, kemudahan penggunaan antarmuka, hingga kepercayaan terhadap keamanan transaksi. Data yang terkumpul secara sistematis kemudian dapat dianalisis lebih lanjut untuk mengungkap pola minat beli konsumen yang dapat dijadikan dasar pengambilan keputusan strategis.

Dalam ranah analisis data, *data mining* menyediakan sekumpulan teknik untuk menggali informasi tersembunyi dari kumpulan data yang kompleks[3]. Di antara berbagai algoritma *data mining*, algoritma C4.5 dikenal sebagai salah satu metode yang efektif untuk membangun model klasifikasi berbasis pohon keputusan (*decision tree*) yang menghasilkan aturan keputusan eksplisit dan mudah diinterpretasikan[5]. Selain C4.5, algoritma Naive Bayes juga dikenal sebagai metode klasifikasi yang efisien dan kerap digunakan sebagai pembandingan dalam penelitian *data mining* untuk mengevaluasi efektivitas relatif antar algoritma pada karakteristik dataset tertentu.

Penelitian terdahulu yang menerapkan algoritma C4.5 pada ranah *e-commerce* maupun analisis perilaku pengguna umumnya bersumber dari data transaksional atau data penggunaan yang terekam secara objektif. Penelitian mengenai tren produk *e-commerce* Indonesia, misalnya, menerapkan C4.5 pada data transaksi dengan atribut target kategori produk dan gender konsumen, menghasilkan akurasi 78,97% untuk klasifikasi kategori produk dan 99,67% untuk klasifikasi gender[4]. Pada konteks lain, penelitian mengenai pola penggunaan perangkat seluler menerapkan C4.5 pada dataset penggunaan aplikasi dan durasi pemakaian perangkat, dengan akurasi model sebesar 41,71% dalam mengklasifikasikan perilaku pengguna[5]. Kedua penelitian tersebut mengandalkan data perilaku atau transaksional yang terekam secara pasif, sehingga belum menyentuh aspek persepsi dan sikap konsumen yang bersifat subjektif terhadap suatu platform.

Penelitian ini berupaya mengisi kesenjangan tersebut dengan menerapkan algoritma C4.5 dan Naive Bayes pada data persepsi konsumen yang dihimpun melalui kuesioner terstruktur, mencakup empat dimensi yang secara teoretis relevan dengan minat beli pada platform *e-commerce*, yaitu harga, keragaman produk, kemudahan penggunaan, dan kepercayaan. Dengan demikian, penelitian ini menawarkan perspektif yang berbeda, yaitu memodelkan minat beli sebagai konstruk psikologis yang diukur secara terstruktur, bukan sekadar pola transaksi atau penggunaan yang terekam secara pasif.

Tujuan utama penelitian ini adalah mengklasifikasikan minat beli konsumen pada platform *e-commerce* ke dalam tiga kategori menggunakan algoritma C4.5, sekaligus membandingkan performanya dengan Naive Bayes untuk mengetahui algoritma yang lebih efektif pada karakteristik dataset yang digunakan. Penelitian ini diharapkan dapat memberikan kontribusi berupa kerangka analisis berbasis data yang transparan dan mudah diinterpretasikan, sehingga pelaku usaha *e-commerce* dapat mengidentifikasi faktor-faktor dominan yang memengaruhi minat beli konsumen dan menjadikannya dasar dalam pengembangan strategi pemasaran yang lebih terarah.

2. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

2.1 Studi Literatur

Pada tahap ini dilakukan studi literatur terhadap penelitian-penelitian terdahulu yang membahas minat beli konsumen pada *e-commerce*, faktor-faktor yang memengaruhinya seperti harga, keragaman, kemudahan penggunaan, dan kepercayaan, serta penerapan algoritma C4.5 dalam klasifikasi data kuesioner. Studi literatur digunakan sebagai dasar dalam penentuan variabel penelitian, penyusunan instrumen kuesioner, serta pemilihan metode analisis yang digunakan[6].

2.1.1 Klasifikasi

Klasifikasi merupakan salah satu teknik utama dalam data mining yang bertujuan untuk memetakan suatu objek data ke dalam kelas tertentu berdasarkan sekumpulan atribut atau variabel yang dimilikinya[5]. Dalam konteks penelitian ini, klasifikasi digunakan untuk mengelompokkan responden ke dalam kategori tingkat minat beli (misalnya rendah dan tinggi) berdasarkan jawaban kuesioner yang merepresentasikan faktor-faktor seperti harga, keragaman produk, kemudahan penggunaan, dan kepercayaan. Secara umum, hasil proses klasifikasi berupa pengelompokan objek ke dalam kelas-kelas tertentu, sehingga peneliti dapat mengidentifikasi pola hubungan

antara atribut yang dimiliki objek dengan kelas yang dihasilkan dan memperoleh pemahaman yang lebih sistematis mengenai karakteristik data yang dianalisis.

2.1.2 Minat Beli Konsumen

Minat beli konsumen didefinisikan sebagai kecenderungan atau keinginan individu untuk melakukan pembelian terhadap suatu produk atau layanan pada masa mendatang, yang tercermin dari sikap positif, perhatian, serta kesediaan untuk mempertimbangkan transaksi. Dalam lingkungan e-commerce, minat beli dipengaruhi oleh persepsi konsumen terhadap manfaat, kemudahan, kenyamanan, keamanan, serta nilai yang dirasakan dari pengalaman berbelanja online. Oleh karena itu, pengukuran minat beli melalui instrumen kuesioner yang terstruktur menjadi dasar penting untuk menganalisis dan mengklasifikasikan perilaku konsumen pada platform e-commerce.

2.1.3 E-Commerce

E-commerce merupakan aktivitas transaksi jual beli barang atau jasa yang dilakukan melalui media elektronik, terutama internet, yang memungkinkan interaksi antara penjual dan pembeli tanpa harus bertemu secara fisik[7]. Perkembangan teknologi informasi dan penetrasi internet telah mendorong pertumbuhan pesat platform e-commerce, sehingga konsumen dapat dengan mudah membandingkan harga, variasi produk, kualitas layanan, dan promosi dari berbagai penyedia dalam waktu yang relatif singkat. Dalam konteks perilaku konsumen, e-commerce menyediakan lingkungan berbelanja yang ditandai oleh kemudahan akses, fitur pencarian dan filter produk, sistem pembayaran digital, serta mekanisme ulasan dan rating yang mempengaruhi persepsi dan kepercayaan konsumen. Kondisi ini menjadikan platform e-commerce sebagai sumber data yang kaya untuk dianalisis, termasuk dalam mengidentifikasi pola minat beli konsumen melalui teknik klasifikasi.

2.1.4 Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang digunakan untuk membangun model pohon keputusan (*decision tree*) berdasarkan data yang bersifat kategorikal maupun numerik. Algoritma ini bekerja dengan menghitung entropi dan gain ratio untuk menentukan atribut terbaik sebagai node pemecah (*splitting attribute*), sehingga pohon yang dihasilkan dapat memisahkan kelas dengan seefisien mungkin. C4.5 memiliki beberapa kelebihan, antara lain mampu menangani nilai atribut kontinu melalui proses pembentukan threshold, mengelola data dengan nilai yang hilang (*missing value*), serta menghasilkan aturan keputusan (*if-then rules*) yang mudah dipahami oleh pengguna non-teknis[8]. Pada penelitian ini, C4.5 digunakan untuk mengklasifikasikan tingkat minat beli konsumen pada platform e-commerce berdasarkan kombinasi faktor seperti harga, keragaman produk, kemudahan penggunaan, dan kepercayaan, sehingga dihasilkan pohon keputusan dan aturan yang dapat dijadikan dasar rekomendasi strategi pemasaran.

2.2 Pengumpulan Data

2.2.1 Sumber Data

Data dalam penelitian ini dikumpulkan menggunakan instrumen berupa kuesioner terstruktur yang disusun berdasarkan hasil studi literatur terkait faktor-faktor yang memengaruhi minat beli konsumen pada e-commerce. Kuesioner diberikan kepada responden yang merupakan pengguna aktif platform e-commerce, dengan tujuan memperoleh informasi mengenai persepsi mereka terhadap variabel-variabel penelitian seperti harga, keragaman produk, kemudahan penggunaan, dan kepercayaan. Setiap butir pernyataan pada kuesioner diukur menggunakan skala Likert dengan rentang 1 hingga 5, di mana skor 1 menunjukkan tingkat ketidaksetujuan atau penilaian yang paling rendah, dan skor 5 menunjukkan tingkat persetujuan atau penilaian yang paling tinggi terhadap pernyataan yang diberikan. Penggunaan skala Likert 1–5 dipilih untuk memudahkan responden dalam memberikan jawaban sekaligus memungkinkan peneliti mengkuantifikasi sikap dan persepsi konsumen secara lebih terukur dalam proses analisis data[9].

2.2.2 Variabel dan Indikator Penelitian

Pada penelitian ini, setiap variabel dioperasionalkan ke dalam beberapa indikator yang diukur menggunakan butir pernyataan dalam kuesioner. Seluruh butir pernyataan menggunakan skala Likert lima poin dengan pilihan jawaban ‘sangat tidak setuju’ sampai ‘sangat setuju’ yang kemudian diberi skor 1 sampai 5 untuk keperluan analisis data. Rincian variabel, indikator, dan contoh butir pernyataan kuesioner disajikan pada Tabel 1.

2.2.3 Proses Pengumpulan Data

Pengumpulan data dilakukan dengan kuesioner online yang dibuat dalam format formulir digital melalui Google Form. Kuesioner tersebut berisi butir pernyataan yang mengukur faktor-faktor yang diduga memengaruhi minat beli konsumen pada platform e-commerce, seperti harga, keragaman produk, kemudahan penggunaan, dan kepercayaan, dengan skala penilaian yang konsisten di setiap item.

Formulir kemudian dibagikan secara daring melalui berbagai kanal media sosial, seperti grup percakapan dan platform jejaring sosial yang relevan dengan target responden, sehingga memungkinkan penyebaran kuesioner secara lebih luas dan efisien. Responden diminta untuk mengisi kuesioner secara sukarela sesuai pengalaman mereka dalam menggunakan platform e-commerce, dan jawaban yang masuk secara otomatis terekam dalam basis data Google Form untuk selanjutnya diunduh dan diolah pada tahap analisis

Tabel 1. Variabel Penelitian, Indikator, dan Butir Pernyataan Kuesioner

Variabel	Indikator	Butir Pernyataan
Harga Produk (X1)	Keterjangkauan harga	Harga produk di platform e-commerce terjangkau bagi saya.
	Ketersediaan promo/diskon	Saya mudah menemukan promo atau diskon di platform e-commerce.
	Kejelasan informasi harga	Informasi harga yang ditampilkan di platform jelas dan transparan.
Keragaman Produk (X2)	Banyaknya kategori produk	Tersedia banyak pilihan kategori produk di platform e-commerce.
	Ketersediaan produk yang dicari	Produk yang saya cari hampir selalu tersedia di platform.
	Variasi/varian produk	Platform menyediakan berbagai varian produk sesuai kebutuhan saya.
Kemudahan Penggunaan Platform (X3)	Kejelasan tampilan	Tampilan aplikasi/platform e-commerce mudah dipahami.
	Kemudahan proses pencarian & pembelian	Proses pencarian dan pembelian produk di platform sangat sederhana.
	Kestabilan/responsivitas sistem	Platform responsif dan transaksi berjalan lancar tanpa kendala teknis.
Kepercayaan terhadap Platform (X4)	Keamanan transaksi	Saya merasa aman melakukan transaksi di platform e-commerce.
	Keandalan ulasan dan rating	Ulasan dan rating produk yang ditampilkan dapat dipercaya.
	Reputasi platform	Platform memiliki reputasi baik dan banyak digunakan oleh masyarakat.

2.3 Uji Instrumen Penelitian

Sebelum data kuesioner digunakan dalam proses pemodelan, dilakukan pengujian kualitas instrumen melalui uji validitas dan uji reliabilitas. Pengujian ini bertujuan untuk memastikan bahwa setiap butir pernyataan dalam kuesioner benar-benar mengukur konstruk yang dimaksud dan menghasilkan hasil yang konsisten[10]. Uji validitas dilakukan menggunakan teknik korelasi Pearson (Pearson Product Moment Correlation), di mana setiap item dikorelasikan dengan skor total variabelnya masing-masing. Item dinyatakan valid apabila nilai r hitung lebih besar dari r tabel pada taraf signifikansi 5%.[11]. Uji reliabilitas dilakukan menggunakan metode Cronbach's Alpha, dengan kriteria instrumen dinyatakan reliabel apabila nilai $\alpha \geq 0,6$ [12]. Kedua pengujian dilakukan menggunakan perangkat lunak SPSS, dengan pengujian per variabel secara terpisah untuk menjaga akurasi interpretasi hasil.

2.4 Pengolahan Data

Pada Pengolahan data dilakukan untuk menyiapkan data hasil kuesioner agar siap digunakan dalam proses klasifikasi minat beli konsumen menggunakan algoritma C4.5. Tahapan ini meliputi seleksi data, pengkodean, perhitungan nilai variabel, penentuan label kelas, serta penyusunan dataset pemodelan sehingga data bersifat konsisten dan sesuai dengan kebutuhan analisis.

2.4.1 Seleksi dan Pemeriksaan Data

Data kuesioner yang diperoleh melalui Google Form terlebih dahulu diekspor ke dalam format spreadsheet dan diperiksa kelengkapannya. Hanya respons yang mengisi seluruh butir pernyataan dan tidak mengandung nilai kosong (missing value) yang disertakan dalam pengolahan, sedangkan data yang tidak lengkap dikeluarkan untuk menjaga kualitas dan mengurangi bias.

2.4.2 Pengkodean dan Penyeragaman Data

Seluruh jawaban pada skala Likert lima poin dikonversi menjadi skor numerik 1 sampai 5 sesuai tingkat persetujuan. Skor dari setiap butir kemudian dikelompokkan sesuai variabel penelitian, yaitu harga produk, keragaman produk, kemudahan penggunaan platform, dan kepercayaan terhadap platform, sehingga dihasilkan data numerik yang seragam dan siap diproses lebih lanjut.

Tabel 2. Skala Penilaian Responden

Skor Numerik	Kategori Jawaban
5	Sangat setuju
4	Setuju
3	Netral
2	Tidak setuju
1	Sangat tidak setuju

2.4.3 Perhitungan Nilai Variabel dan Penentuan Label Minat Beli

Untuk setiap variabel, skor dari tiga butir pernyataan dijumlahkan dan dibagi tiga sehingga diperoleh nilai rata-rata yang merepresentasikan persepsi responden terhadap variabel tersebut secara ringkas dan proporsional. Nilai rata-rata keempat variabel kemudian dijumlahkan dan dibagi empat untuk memperoleh skor komposit minat beli. Skor komposit ini digunakan sebagai dasar pelabelan kelas dan label tersebut digunakan sebagai atribut kelas pada proses klasifikasi C4.5.

Tabel 3. Kategori Minat Beli Konsumen Berdasarkan Skor Rata-rata Variabel

Skor Rata-rata	Kategori Minat Beli
> 4,00	Tinggi
3,01 – 4,00	Sedang
≤ 3,00	Rendah

2.4.4 Penyusunan Dataset untuk Pemodelan

Data yang telah melalui proses seleksi, pengkodean, perhitungan nilai rata-rata, dan pelabelan kelas disusun menjadi satu dataset terstruktur dengan baris sebagai responden dan kolom sebagai variabel prediktor serta kelas minat beli. Dataset ini kemudian diimpor ke dalam perangkat lunak RapidMiner untuk digunakan pada tahap pemodelan klasifikasi minat beli konsumen menggunakan algoritma C4.5.

2.5 Analisis Data

Tahap analisis data pada penelitian ini dilakukan setelah seluruh data melewati proses pengolahan secara sistematis, sehingga data berada dalam kondisi siap dianalisis. Pada tahap ini, algoritma C4.5 diterapkan dengan

bantuan perangkat lunak RapidMiner untuk membangun model klasifikasi yang menggambarkan hubungan antara variabel-variabel penelitian dengan kelas minat beli konsumen pada platform e-commerce.

Proses analisis data tersebut menjadi landasan krusial sebelum dilakukan pembahasan yang lebih mendalam. Dengan persiapan dan pengelolaan data yang teliti, dataset yang digunakan diharapkan menghasilkan informasi yang akurat dan relevan dengan tujuan penelitian.. Tingkat ketelitian pada tahap ini akan berpengaruh langsung terhadap mutu serta kredibilitas hasil penelitian, yang selanjutnya menjadi dasar dalam penarikan kesimpulan.

Tabel 4. Sampel Data yang Telah Diolah

No	Harga	Keragaman	Kemudahan	Kepercayaan	Rata-Rata	Minat Beli
1	4,33	5,00	4,67	4,67	4,67	Tinggi
2	4,00	5,00	4,67	3,67	4,33	Tinggi
3	5,00	5,00	5,00	5,00	5,00	Tinggi
4	4,33	4,67	4,33	4,33	4,42	Tinggi
5	4,00	3,33	3,67	3,33	3,58	Sedang
6	3,67	4,00	4,00	4,00	3,92	Sedang
7	4,00	5,00	5,00	4,67	4,67	Tinggi
8	3,67	4,00	3,67	3,00	3,58	Sedang
9	3,00	3,00	3,00	3,00	3,00	Rendah
10	5,00	5,00	5,00	5,00	5,00	Tinggi
11	4,00	4,67	4,67	5,00	4,58	Tinggi
12	3,67	4,00	4,00	3,67	3,83	Sedang
13	3,00	3,00	3,00	3,00	3,00	Rendah
14	5,00	5,00	5,00	4,33	4,83	Tinggi
15	5,00	5,00	4,67	3,67	4,58	Tinggi
16	4,33	4,33	4,00	3,67	4,08	Tinggi
17	4,33	5,00	5,00	4,33	4,67	Tinggi
18	4,33	5,00	4,67	3,00	4,25	Tinggi
19	3,67	4,67	3,33	2,33	3,50	Sedang
20	4,00	3,33	4,00	4,00	3,83	Sedang

2.5.1 Algoritma C4.5

Algoritma C4.5 merupakan metode pembentukan pohon keputusan yang bekerja secara rekursif dengan memilih atribut terbaik sebagai pemisah data pada setiap node, berdasarkan kemampuan atribut tersebut dalam memisahkan kelas secara optimal [13]. Proses ini dimulai dari node akar dan berlanjut ke cabang-cabang berikutnya hingga seluruh data terklasifikasi ke dalam kelas yang homogen atau memenuhi kondisi berhenti yang telah ditetapkan.

Perhitungan nilai gain pada algoritma C4.5 merupakan langkah yang sangat penting dalam pembentukan pohon keputusan. Nilai ini digunakan untuk menilai atribut mana yang paling informatif dalam memisahkan data ke dalam kelompok-kelompok kelas minat beli yang lebih homogen, sehingga semakin besar nilai gain suatu atribut, semakin baik kemampuannya dalam membedakan kategori minat beli konsumen. Dalam penelitian ini, atribut dengan nilai gain tertinggi menjadi indikator utama dalam membedakan konsumen berdasarkan kategori minat beli Tinggi, Sedang, dan Rendah, serta menjadi dasar pembentukan struktur pohon keputusan yang optimal.

Sebelum menghitung gain, terlebih dahulu dihitung nilai entropy yang merepresentasikan tingkat ketidakpastian atau ketidakhomogenitas dalam suatu himpunan data. Semakin tinggi nilai entropy, semakin beragam distribusi kelas yang terdapat dalam data tersebut, sehingga semakin sulit dilakukan pemisahan kelas secara akurat. Sebaliknya, nilai entropy yang mendekati nol menandakan bahwa data hampir seluruhnya berada dalam satu kelas yang sama. Secara matematis, entropy dihitung menggunakan rumus:

$$Entropy(S) = - \sum_{k=1}^c p_k \log_2(p_k) \quad (1)$$

Keterangan:

S = himpunan kasus yang sedang dianalisis

c = jumlah kelas pada himpunan S

p_k = proporsi atau probabilitas kasus kelas ke-k terhadap keseluruhan kasus pada S

$\log_2(p_k)$ = logaritma basis 2 dari proporsi kelas ke-k

Setelah nilai entropy diperoleh, langkah berikutnya adalah menghitung nilai gain untuk setiap atribut yang tersedia. Gain mengukur seberapa besar penurunan ketidakpastian yang diperoleh ketika himpunan data S dibagi berdasarkan atribut V tertentu. Atribut dengan nilai gain tertinggi dipilih sebagai node pemisah karena dianggap paling efektif dalam membedakan kelas pada tahap tersebut. Secara matematis, gain dihitung menggunakan rumus:

$$Gain(S, V) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i) \quad (2)$$

Keterangan:

S = himpunan kasus keseluruhan sebelum pembagian

V = atribut yang sedang dievaluasi sebagai kandidat pemisah

n = jumlah partisi atau nilai yang dimiliki atribut V

$|S_i|$ = jumlah kasus pada partisi ke-i setelah pembagian berdasarkan atribut V

$|S|$ = jumlah total kasus pada himpunan S

$Entropy(S_i)$ = nilai entropy pada partisi ke-i

Nilai gain yang diperoleh kemudian dibandingkan antar atribut, dan atribut dengan gain tertinggi ditetapkan sebagai node pemisah pada posisi tersebut dalam pohon keputusan. Proses ini diulangi secara rekursif pada setiap cabang yang terbentuk hingga pohon keputusan selesai dibangun dan seluruh data terklasifikasikan ke dalam kelas minat beli yang sesuai, yaitu Tinggi, Sedang, atau Rendah.

2.5.2 RapidMiner

Pada penelitian ini, perangkat lunak RapidMiner digunakan sebagai alat bantu untuk mengimplementasikan algoritma C4.5 dalam proses klasifikasi tingkat minat beli konsumen pada platform *e-commerce* [14]. Dataset yang telah melalui tahap pengolahan data diimpor ke dalam RapidMiner dalam format lembar kerja Excel, kemudian ditetapkan peran masing-masing kolom pada tahap import data, di mana variabel Minat Beli didefinisikan sebagai label, sedangkan variabel Harga Produk, Keragaman Produk, Kemudahan Penggunaan, dan Kepercayaan terhadap Platform ditetapkan sebagai atribut prediktor. Kolom Rata-Rata tidak diikutsertakan sebagai atribut prediktor untuk menghindari kebocoran data (*data leakage*), mengingat kolom tersebut merupakan dasar pembentukan label kelas Minat Beli.

Evaluasi model dilakukan menggunakan dua pendekatan untuk memastikan hasil yang konsisten dan tidak bergantung pada satu pembagian data tertentu. Pertama, split validation dengan rasio 70:30 menggunakan stratified sampling, di mana 70% data digunakan untuk melatih model dan 30% data digunakan untuk menguji model. Alur proses split validation disusun menggunakan operator Retrieve Dataset → Split Data → Decision Tree → Apply Model → Performance. Kedua, 10-fold cross validation dengan stratified sampling, di mana seluruh data dibagi menjadi 10 lipatan secara bergantian sehingga setiap data pernah digunakan sebagai data latih maupun data uji. Alur proses cross validation disusun menggunakan operator Retrieve Dataset → Cross Validation, dengan Decision Tree di sisi Training serta Apply Model dan Performance di sisi Testing.

Operator Decision Tree dikonfigurasi menggunakan kriteria gain ratio sesuai prinsip algoritma C4.5. Sebagai pembanding, algoritma Naive Bayes juga diimplementasikan dengan konfigurasi evaluasi yang sama untuk mengukur efektivitas relatif kedua algoritma dalam mengklasifikasikan minat beli konsumen ke dalam tiga kategori, yaitu Tinggi, Sedang, dan Rendah[15].

3. HASIL DAN PEMBAHASAN

3.1 Hasil Uji Instrumen Penelitian

3.1.1 Hasil Uji Validitas

Uji validitas dilakukan terhadap seluruh butir pernyataan kuesioner menggunakan teknik korelasi Pearson (Pearson Product Moment Correlation) dengan bantuan perangkat lunak SPSS. Setiap item dikorelasikan dengan skor total variabelnya masing-masing, dan item dinyatakan valid apabila nilai r hitung lebih besar dari r tabel pada taraf signifikansi 5%. Dengan jumlah responden sebanyak 106, diperoleh nilai derajat kebebasan ($df = n - 2$) sebesar 104, sehingga nilai r tabel yang digunakan adalah 0,191[10].

Tabel 5. Hasil Uji Validitas Instrumen Penelitian

Variabel	Item	r hitung	r tabel	Keterangan
Harga Produk (X1)	X1.1	0,793	0,191	Valid
	X1.2	0,785	0,191	Valid
	X1.3	0,785	0,191	Valid
Keragaman Produk (X2)	X2.1	0,855	0,191	Valid
	X2.2	0,923	0,191	Valid
	X2.3	0,883	0,191	Valid
Kemudahan Penggunaan (X3)	X3.1	0,849	0,191	Valid
	X3.2	0,904	0,191	Valid
	X3.3	0,841	0,191	Valid
Kepercayaan Platform (X4)	X4.1	0,852	0,191	Valid
	X4.2	0,854	0,191	Valid
	X4.3	0,858	0,191	Valid

Berdasarkan Tabel 5, seluruh 12 butir pernyataan dari keempat variabel penelitian menunjukkan nilai r hitung yang lebih besar dari r tabel (0,191) dengan taraf signifikansi 5%. Nilai r hitung tertinggi diperoleh oleh item X2.2 sebesar 0,923, sedangkan nilai terendah diperoleh oleh item X1.2 dan X1.3 masing-masing sebesar 0,785 — keduanya tetap jauh melampaui nilai r tabel. Dengan demikian, seluruh butir pernyataan dinyatakan valid dan layak digunakan sebagai instrumen pengumpulan data dalam penelitian ini.

3.1.2 Hasil Uji Reliabilitas

Uji reliabilitas dilakukan untuk mengukur tingkat konsistensi internal instrumen kuesioner menggunakan metode Cronbach's Alpha dengan bantuan perangkat lunak SPSS. Pengujian dilakukan secara terpisah untuk setiap variabel, dan instrumen dinyatakan reliabel apabila nilai Cronbach's Alpha $> 0,6$ [12].

Berdasarkan Tabel 6, seluruh variabel penelitian menunjukkan nilai Cronbach's Alpha yang memenuhi syarat reliabilitas. Nilai tertinggi diperoleh variabel Keragaman Produk (X2) sebesar 0,860, yang menunjukkan konsistensi internal yang sangat baik. Sementara itu, variabel Harga Produk (X1) memperoleh nilai terendah sebesar 0,690, namun tetap berada di atas ambang batas minimum 0,6 sehingga masih dinyatakan reliabel. Dengan demikian, seluruh instrumen penelitian dinyatakan reliabel dan dapat menghasilkan data yang konsisten apabila digunakan secara berulang dalam kondisi yang serupa.

Tabel 6. Hasil Uji Reliabilitas Instrumen Penelitian

Variabel	Cronbach's Alpha	Syarat	Keterangan
Harga Produk (X1)	0,690	> 0,6	Reliabel
Keragaman Produk (X2)	0,860	> 0,6	Reliabel
Kemudahan Penggunaan (X3)	0,830	> 0,6	Reliabel
Kepercayaan Platform (X4)	0,803	> 0,6	Reliabel

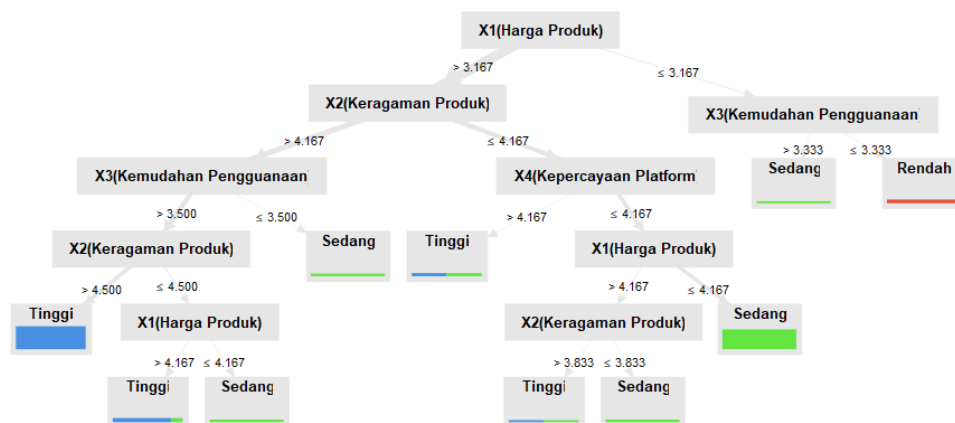
3.2 Hasil Pemodelan dan Evaluasi

3.2.1 Pohon Keputusan



Gambar 2. Hasil Pohon Keputusan Split Validation

Penerapan algoritma C4.5 dengan metode split validation menghasilkan struktur pohon keputusan yang menempatkan X1 (Harga Produk) sebagai node akar dengan ambang 3,167. Konsumen yang menilai harga produk di bawah atau sama dengan 3,167 langsung diklasifikasikan sebagai minat beli Rendah. Sementara itu, konsumen yang menilai di atas 3,167 dilanjutkan ke pemisahan berikutnya berdasarkan X2 (Keragaman Produk) dengan ambang 4,167. Pohon ini melibatkan keempat variabel penelitian sebagai node pemisah dan menghasilkan enam aturan klasifikasi yang dapat ditelusuri secara eksplisit dari struktur pohon.



Gambar 3. Hasil Pohon Keputusan Cross Validation

Penerapan algoritma C4.5 dengan metode 10-fold cross validation menghasilkan pohon keputusan yang lebih kompleks dibandingkan split validation. X1 (Harga Produk) tetap menjadi node akar dengan ambang yang sama yaitu 3,167, menunjukkan konsistensi dominasi variabel harga di kedua metode evaluasi. Pada cabang kanan ($X1 \leq 3,167$), X3 (Kemudahan Penggunaan) menjadi pemisah berikutnya dengan ambang 3,333 yang membedakan

kelas Sedang dan Rendah. Pada cabang kiri ($X1 > 3,167$), $X2$ (Keragaman Produk) menjadi pemisah kedua, diikuti oleh $X3$, $X4$, $X1$, dan $X2$ secara bergantian di cabang-cabang berikutnya. Seluruh empat variabel penelitian muncul sebagai node pemisah dalam pohon ini, menghasilkan sepuluh aturan klasifikasi yang lebih rinci dibandingkan split validation.

3.2.2 Evaluasi Model C4.5

Evaluasi performa model C4.5 dilakukan menggunakan dua metode, yaitu split validation dengan rasio 70:30 dan 10-fold cross validation. Hasil evaluasi disajikan dalam bentuk confusion matrix yang mencakup nilai akurasi, precision, dan recall per kelas.

a. Split Validation

Evaluasi menggunakan split validation menghasilkan akurasi sebesar 83,87%. Hasil confusion matrix disajikan pada Tabel 7.

Tabel 7. Confusion Matrix C4.5 Split Validation

	True Tinggi	True Sedang	True Rendah	Class Precision
Pred. Tinggi	13	2	0	86,67%
Pred. Sedang	2	12	0	85,71%
Pred. Rendah	0	1	1	50,00%
Class Recall	86,67%	80,00%	100,00%	

Berdasarkan Tabel 7, dari total 32 data testing, model berhasil mengklasifikasikan dengan benar sebanyak 13 data kelas Tinggi, 12 data kelas Sedang, dan 1 data kelas Rendah. Terdapat 5 kesalahan klasifikasi, yaitu 2 data kelas Tinggi salah diprediksi sebagai Sedang, 2 data kelas Sedang salah diprediksi sebagai Tinggi, dan 1 data kelas Sedang salah diprediksi sebagai Rendah. Kelas Tinggi dan Sedang menunjukkan performa yang seimbang dengan precision masing-masing 86,67% dan 85,71%. Kelas Rendah memperoleh recall 100% yang berarti seluruh data Rendah berhasil diidentifikasi, namun precision hanya 50% karena satu data kelas Sedang ikut terprediksi sebagai Rendah.

b. Cross Validation

Evaluasi menggunakan 10-fold cross validation menghasilkan akurasi sebesar 91,51% dengan standar deviasi $\pm 9,28\%$. Hasil confusion matrix disajikan pada Tabel 8.

Tabel 8. Confusion Matrix C4.5 — Cross Validation

	True Tinggi	True Sedang	True Rendah	Class Precision
Pred. Tinggi	45	2	0	95,74%
Pred. Sedang	5	47	0	90,38%
Pred. Rendah	0	2	5	71,43%
Class Recall	90,00%	92,16%	100,00%	

Berdasarkan Tabel 8, dari total 106 data, model berhasil mengklasifikasikan dengan benar sebanyak 45 data kelas Tinggi, 47 data kelas Sedang, dan 5 data kelas Rendah. Terdapat 9 kesalahan klasifikasi, yaitu 2 data kelas Tinggi salah diprediksi sebagai Sedang, 5 data kelas Sedang salah diprediksi sebagai Tinggi, dan 2 data kelas Rendah salah diprediksi sebagai Sedang. Kelas Tinggi memperoleh precision tertinggi sebesar 95,74%, sedangkan kelas Rendah memperoleh precision terendah sebesar 71,43% namun tetap mencapai recall 100%, yang berarti seluruh 5 data kelas Rendah berhasil diidentifikasi dengan benar oleh model. Akurasi cross validation yang lebih tinggi dibandingkan split validation menunjukkan bahwa penggunaan seluruh data secara bergantian menghasilkan estimasi performa yang lebih stabil.

3.2.3 Evaluasi Model Naive Bayes

Evaluasi performa model Naive Bayes dilakukan menggunakan metode evaluasi yang sama dengan C4.5, yaitu split validation dengan rasio 70:30 dan 10-fold cross validation dengan stratified sampling.

a. Split Validation

Evaluasi menggunakan split validation menghasilkan akurasi sebesar 96,77%. Hasil confusion matrix disajikan pada Tabel 9.

Tabel 9. Confusion Matrix Naive Bayes Split Validation

	True Tinggi	True Sedang	True Rendah	Class Precision
Pred. Tinggi	14	0	0	100,00%
Pred. Sedang	1	15	0	93,75%
Pred. Rendah	0	0	1	100,00%
Class Recall	93,33%	100,00%	100,00%	

Berdasarkan Tabel 9, dari total 31 data testing, model berhasil mengklasifikasikan dengan benar sebanyak 14 data kelas Tinggi, 15 data kelas Sedang, dan 1 data kelas Rendah. Hanya terdapat 1 kesalahan klasifikasi, yaitu 1 data kelas Tinggi salah diprediksi sebagai Sedang. Kelas Tinggi dan Rendah memperoleh precision sempurna sebesar 100%, sedangkan kelas Sedang memperoleh precision 93,75%. Kelas Sedang dan Rendah mencapai recall 100%, menunjukkan bahwa seluruh data pada kedua kelas tersebut berhasil diidentifikasi dengan benar.

a. Cross Validation

Evaluasi menggunakan 10-fold cross validation menghasilkan akurasi sebesar 95,36% dengan standar deviasi $\pm 4,89\%$. Hasil confusion matrix disajikan pada Tabel 10.

Tabel 10. Confusion Matrix Naive Bayes Cross Validation

	True Tinggi	True Sedang	True Rendah	Class Precision
Pred. Tinggi	45	0	0	100,00%
Pred. Sedang	5	51	0	91,07%
Pred. Rendah	0	0	5	100,00%
Class Recall	90,00%	100,00%	100,00%	

Berdasarkan Tabel 10, dari total 106 data, model berhasil mengklasifikasikan dengan benar sebanyak 45 data kelas Tinggi, 51 data kelas Sedang, dan 5 data kelas Rendah. Hanya terdapat 5 kesalahan klasifikasi, seluruhnya berasal dari data kelas Tinggi yang salah diprediksi sebagai Sedang. Kelas Tinggi dan Rendah memperoleh precision sempurna sebesar 100%, sedangkan kelas Sedang memperoleh precision 91,07%. Kelas Sedang dan Rendah kembali mencapai recall 100%, mengonfirmasi konsistensi model Naive Bayes dalam mengidentifikasi kedua kelas tersebut secara akurat di kedua metode evaluasi.

3.2.4 Perbandingan Performa C4.5 dan Naive Bayes

Perbandingan performa kedua algoritma dilakukan berdasarkan nilai akurasi, precision, dan recall per kelas yang diperoleh dari masing-masing metode evaluasi. Hasil perbandingan disajikan pada Tabel 11.

Tabel 11. Perbandingan Performa C4.5 dan Naive Bayes

No	Algoritma	Metode Evaluasi	Akurasi	Kelas	Precision	Recall
1	C4.5	Split Validation	83,87%	Tinggi	86,67%	86,67%
				Sedang	85,71%	80,00%
				Rendah	50,00%	100,00%
2	C4.5	Cross Validation	91,51%	Tinggi	95,74%	90,00%
				Sedang	90,38%	92,16%
				Rendah	71,43%	100,00%
3	Naive Bayes	Split Validation	96,77%	Tinggi	100,00%	93,33%
				Sedang	93,75%	100,00%
				Rendah	100,00%	100,00%
4	Naive Bayes	Cross Validation	95,36%	Tinggi	100,00%	90,00%
				Sedang	91,07%	100,00%
				Rendah	100,00%	100,00%

Berdasarkan Tabel 11, Naive Bayes secara konsisten menghasilkan akurasi yang lebih tinggi dibandingkan C4.5 pada kedua metode evaluasi. Perbedaan yang paling mencolok terlihat pada precision kelas Rendah — C4.5 hanya mencapai 50,00% pada split validation dan 71,43% pada cross validation, sementara Naive Bayes memperoleh precision sempurna 100,00% pada keduanya. Nilai recall kelas Rendah pada kedua algoritma sama-sama mencapai 100,00%, yang berarti kedua model berhasil mengidentifikasi seluruh data kelas Rendah dengan benar. Pada kelas Tinggi dan Sedang, performa kedua algoritma relatif sebanding dengan selisih yang tidak terlalu signifikan.

3.3 Analisis Ketidakseimbangan Kelas dan Implikasinya terhadap Performa Model

Distribusi kelas pada dataset penelitian ini menunjukkan ketidakseimbangan yang cukup signifikan, di mana kelas Tinggi dan Sedang masing-masing diwakili oleh 50 dan 51 data, sementara kelas Rendah hanya terdiri dari 5 data atau sekitar 4,7% dari total 106 responden. Kondisi ketidakseimbangan kelas (class imbalance) seperti ini merupakan karakteristik yang umum dijumpai pada data survei persepsi konsumen, terutama ketika sebagian besar responden memiliki kecenderungan penilaian yang positif terhadap platform e-commerce yang mereka gunakan.

Pengaruh ketidakseimbangan ini terlihat jelas pada performa model C4.5, khususnya pada metrik precision kelas Rendah. Pada evaluasi menggunakan split validation, precision kelas Rendah hanya mencapai 50,00%, yang berarti separuh dari prediksi kelas Rendah yang dihasilkan model ternyata merupakan kesalahan klasifikasi. Meskipun nilai ini meningkat menjadi 71,43% pada evaluasi cross validation, angka tersebut tetap jauh lebih rendah dibandingkan precision kelas Tinggi dan Sedang yang berkisar antara 85–95%. Kondisi ini terjadi karena algoritma C4.5 dalam membangun struktur pohon keputusan cenderung mengoptimalkan pemisahan berdasarkan kelas yang memiliki jumlah data lebih besar, sehingga kemampuan model dalam mengenali pola kelas minoritas menjadi terbatas.

Sebaliknya, model Naive Bayes menunjukkan performa yang lebih robust terhadap ketidakseimbangan kelas ini, dengan precision kelas Rendah mencapai 100,00% pada kedua metode evaluasi. Hal ini dapat dikaitkan dengan cara kerja Naive Bayes yang menghitung probabilitas setiap kelas secara independen berdasarkan distribusi atribut per kelas, sehingga estimasi probabilitas kelas minoritas tidak terdistorsi oleh dominasi kelas mayoritas sebagaimana yang terjadi pada pendekatan berbasis pemisahan rekursif seperti decision tree. Perbedaan performa ini menegaskan bahwa keunggulan Naive Bayes atas C4.5 pada dataset penelitian ini tidak semata-mata mencerminkan superioritas algoritma secara umum, melainkan lebih dipengaruhi oleh karakteristik distribusi data yang tidak seimbang.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model klasifikasi minat beli konsumen pada platform e-commerce menggunakan algoritma C4.5 di RapidMiner, dengan dataset yang diperoleh dari kuesioner terstruktur yang mengukur empat variabel, yaitu harga produk, keragaman produk, kemudahan penggunaan, dan kepercayaan terhadap platform. Sebelum pemodelan, seluruh instrumen kuesioner telah diverifikasi melalui uji validitas dan reliabilitas menggunakan SPSS, dengan hasil seluruh butir pernyataan dinyatakan valid dan reliabel. Label kelas minat beli ditetapkan dalam tiga kategori yaitu Tinggi, Sedang, dan Rendah berdasarkan skor komposit rata-rata keempat variabel, sementara kolom Rata-Rata tidak diikutsertakan sebagai atribut prediktor untuk menghindari kebocoran data.

Evaluasi model C4.5 dilakukan menggunakan dua metode, yaitu split validation (70:30) dan 10-fold cross validation, menghasilkan akurasi masing-masing sebesar 83,87% dan 91,51%. Sebagai pembanding, model Naive Bayes menghasilkan akurasi yang lebih tinggi, yaitu 96,77% pada split validation dan 95,36% pada cross validation. Keunggulan Naive Bayes atas C4.5 pada penelitian ini tidak semata-mata mencerminkan superioritas algoritma secara umum, melainkan turut dipengaruhi oleh kondisi ketidakseimbangan kelas pada dataset, di mana kelas Rendah hanya diwakili oleh 5 data dari total 106 responden. Meskipun demikian, C4.5 tetap memberikan nilai lebih dalam hal interpretabilitas, karena struktur pohon keputusan yang dihasilkan memungkinkan penelusuran aturan klasifikasi secara eksplisit melalui aturan if-then yang dapat langsung diterapkan dalam pengambilan keputusan strategis.

Untuk penelitian selanjutnya, disarankan agar pengumpulan data dilakukan dengan jumlah responden yang lebih besar dan distribusi kelas yang lebih seimbang, atau menerapkan teknik penanganan ketidakseimbangan kelas seperti oversampling SMOTE (Synthetic Minority Oversampling Technique) guna meningkatkan kemampuan model dalam mengenali kelas minoritas secara lebih akurat. Selain itu, validasi model pada dataset independen dari konteks yang berbeda perlu dilakukan untuk mengonfirmasi generalisasi model di luar lingkup penelitian ini.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada seluruh pihak yang telah memberikan dukungan dalam penyusunan dan penyelesaian penelitian ini. Apresiasi disampaikan kepada seluruh responden yang telah meluangkan waktu untuk mengisi kuesioner, sehingga data penelitian dapat terkumpul secara lengkap. Penulis turut berterima kasih kepada rekan-rekan yang telah membantu dalam diskusi dan penyempurnaan naskah, serta kepada keluarga yang memberikan dukungan moral selama proses penelitian. Semoga hasil penelitian ini dapat memberikan kontribusi yang bermanfaat bagi pengembangan kajian *e-commerce* dan penerapan metode klasifikasi *data mining* dalam analisis perilaku konsumen.

DAFTAR PUSTAKA

- [1] M. Yuli and S. Aisah, "Perkembangan dan Tren E-Commerce di Indonesia," *J. Manaj. dan Bisnis Ekon.*, vol. 3, no. 4, pp. 131–140, 2025.
- [2] N. A. P. Haryandini, W. Prihartono, and Fathurrohman, "ANALISIS SEGMENTASI PELANGGAN UNTUK MENYUSUN STRATEGI PROMOSI MENGGUNAKAN K-MEANS: STUDI KASUS DI PT XYZ," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 13, no. 1, 2025.
- [3] A. M. Pratama, Siswanto, and R. Zulfiandry, "Implementasi Data Mining Untuk Analisis Prilaku Konsumen Dalam Meningkatkan Penjualan Di Shopee," *J. Media Infotama*, vol. 21, no. 2, pp. 703–709, 2025.
- [4] Z. Fatah and I. A. Fikri, "ANALISIS TREND PRODUK E-COMMERCE INDONESIA MENGGUNAKAN ALGORITMA C4.5 DECISION TREE," *Merdeka Indones. J. Int.*, vol. 5, no. 2, pp. 246–255, 2025.
- [5] A. Ernawati and S. Wahyuni, "Analisis Data Mining Pola Penggunaan Seluler dan Klasifikasi Perilaku Pengguna di Berbagai Perangkat Menggunakan Metode C4.5," *Bull. Inf. Technol.*, vol. 5, no. 4, pp. 162–168, 2024, doi: 10.47065/bit.v5i2.1689.
- [6] M. W. A. Prastya, M. Tahir, A. A. Ningrum, and A. P. Zaibintoro, "Analisis Ancaman Pishing melalui Aplikasi WhatsApp : Review Metode Studi Literatur," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 7, no. 3, pp. 190–197, 2024.
- [7] S. N. Faizah, H. Sasana, and J. A. Prakoso, "Analisis Perubahan Konsumsi E-commerce Di Indonesia Selama Pandemi Covid-19," *J. Ekon. Pembangunan(JEP)*, vol. 10, no. 2, pp. 74–83, 2021.
- [8] W. Indra Kamala, A. Yansah, R. Pebrina, and Suhardhono, "PENERAPAN ALGORITMA KLASIFIKASI C4.5 REVIEW PELANGGAN DALAM BERTRANSAKSI PADAPT. DAPOER MAMIH," *J. Ris. Sist. Informasi(JISSI)*, vol. 3, no. 1, pp. 1–8, 2026.

- [9] P. P. Pramono and I. G. Sanica, "USER EXPERIENCE INFLUENCE ON PURCHASE INTENTION IN TOKOPEDIA," *J. Univers. Stud.*, vol. 5, no. 2, pp. 2248–2264, 2025.
- [10] Y. Utami, P. M. Rasmanna, and Khairunnisa, "Uji Validitas dan Uji Reliabilitas Instrument Penilaian Kinerja Dosen," *SAINTEK (Jurnal Sains dan Teknol.*, vol. 4, no. 2, pp. 21–24, 2023.
- [11] Fitriyani, M. Hasibuan, R. Sinaga, and W. W. Hulu, "Uji Validitas Dan Reliabilitas Instrumen Persepsi Mahasiswa Terhadap Penugasan Artikel Dalam Pembelajaran Di Jurusan Matematika UNIMED," *Menulis J. Penelit. Nusantara*, vol. 1, no. 12, pp. 714–719, 2025.
- [12] F. D. P. Anggraini, Aprianti, V. A. V. Setyawati, and A. A. Hartanto, "Pembelajaran Statistika Menggunakan Software SPSS untuk Uji Validitas dan Reliabilitas," *J. BASICEDU*, vol. 6, no. 4, pp. 6491–6504, 2022.
- [13] A. Sugarda, S. Jalaluddin, A. P. Windarto, and W. Robiansyah, "Penerapan Metode Data Mining C4 . 5 dalam Penentuan Kelayakan Rehabilitas Rumah Warga," *J. Comput. Informatics Res.*, vol. 1, no. 3, pp. 56–64, 2022, doi: 10.47065/comforch.v1i3.321.
- [14] N. Artisya and S. P. Saragih, "PREDICTION TINGKAT POTENSI PEMBELI PADA HARI SPECIAL DI TOKO MITRA SHOPEE," *J. Comasie*, vol. 6, no. 4, 2022.
- [15] F. Harahap, N. E. Saragih, E. T. Siregar, and H. Sariangsah, "Penerapan Data Mining Dengan Algoritma Naive Bayes Classifier Dalam Memprediksi Pembelian Cat," *J. Ilm. Inform.*, vol. 9, no. 1, 2021.