**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9702
Hal 909-918
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

# Analisis Komparatif Information Gain Dan Gain Ratio Pada Algoritma C4.5 Untuk Klasifikasi Produk ATK Terlaris Segmen Business-to-School

**Ana Billah*, Dicky Nofriansyah, Ahmad Calam**

Fakultas Informatika dan Komputer, Sistem Informasi, STMIK Triguna Dharma, Medan, Indonesia
Email: [1]anabila2201@email.com, [2]nofriansyahdicky@email.com, [3]calamahmad223@email.com
Email Penulis Korespondensi: anabila2201@email.com
Submitted **15-04-2026;** Accepted **02-06-2026;** Published **30-06-2026**

**Abstrak**

Ketidakpastian dalam manajemen stok, seperti risiko overstock dan stock-out, menjadi tantangan utama bagi distributor Alat Tulis Kantor (ATK) dalam menghadapi pola permintaan musiman yang fluktuatif pada segmen Businesst-to-School (B2S). Penelitian ini bertujuan untuk menganalisis perbandingan kriteria pemilihan atribut dalam algoritma C4.5, yaitu Information Gain dan Gain Ratio, dalam klasifikasi produk ATK terlaris. Dataset yang digunakan terdiri dari 647 data transaksi penjualan periode Januari hingga Desember 2024. Kebaruan penelitian ini terletak pada analisis komparatif kedua kriteria tersebut pada dataset penjualan dengan karakteristik musiman spesifik, yang belum banyak dibahas pada penelitian terdahulu yang umumnya hanya berfokus pada penerapan algoritma tunggal. Metodologi penelitian mengikuti tahapan Knowledge Discovery in Database (KDD) secara sistematis. Hasil penelitian memperlihatkan Information Gain menghasilkan nilai akurasi yang sedikit lebih tinggi, yaitu 78,98%, sedangkan Gain Ratio (73,26 %) menghasilkan model dengan struktur pohon keputusan yang lebih sederhana, stabil dan mudah diinterpretasikan. Atribut Jenis Pengadaan teridentifikasi sebagai faktor paling dominan dalam menentukan tingkat kelarisan produk. Sebagai kesimpulan utama, penelitian ini menetapkan bahwa Gain Ratio merupakan metode yang lebih optimal untuk pengambilan keputusan bisnis strategis karena melalui normalisasi Split Information, metode ini berhasil mengurangi bias terhadap atribut bervariasi banyak dan menghasilkan struktur pohon keputusan yang lebih ringkas serta terhindar dari overfitting dibandingkan Information Gain.

**Kata Kunci:** Data Mining; Algoritma C4.5; Klasifikasi; ATK; Gain Ratio; Information Gain

**Abstract**

Uncertainty in stock management, such as the risk of overstock and stock-out, is a major challenge for Office Stationery (ATK) distributors in facing fluctuating seasonal demand patterns in the Business-to-School (B2S) segment. This study aims to analyze the comparative attribute selection criteria in the C4.5 algorithm, namely Information Gain and Gain Ratio, in the classification of best-selling ATK products. The dataset used consists of 647 sales transaction data from January to December 2024. The novelty of this study lies in the comparative analysis of the two criteria in a sales dataset with specific seasonal characteristics, which has not been widely discussed in previous studies that generally only focus on the application of a single algorithm. The research methodology follows the Knowledge Discovery in Database (KDD) stages systematically. The results show that Information Gain produces a slightly higher accuracy value, namely 78.98%, while Gain Ratio (77.89%) produces a model with a simpler, more stable, and easier to interpret decision tree structure. The Procurement Type attribute is identified as the most dominant factor in determining the level of product sales. As a main conclusion, this study establishes that Gain Ratio is a more optimal method for strategic business decision making because through Split Information normalization, this method successfully reduces bias towards highly variable attributes and produces a more concise decision tree structure and avoids overfitting compared to Information Gain.

**Keywords**: Data Mining; C4.5 Algorithm; Classification; ATK; Gain Ratio; Information Gain

## 1. PENDAHULUAN

Di era digital, pengelolaan data transaksi harian sangat krusial bagi keberlangsungan industri distribusi Alat Tulis Kantor (ATK). Data yang terus meningkat sering kali hanya menjadi arsip administratif, padahal mengandung informasi berharga yang dapat diolah menggunakan teknik data mining untuk pengambilan keputusan strategis[1]. Pada segmen *Business-to-School* (B2S), pola penjualan memiliki karakteristik musiman yang dipengaruhi periode akademik, sehingga pelaku usaha sering menghadapi masalah manajemen stok seperti *overstock* atau *stock-out* yang memicu kerugian finansial.

Permasalahan ini menyebabkan perusahaan kesulitan dalam memprediksi produk terlaris, yang umumnya disebabkan oleh ketergantungan pada intuisi dibandingkan analisis variabel spesifik. Oleh karena itu, diperlukan transformasi data mentah menjadi pengetahuan melalui klasifikasi guna menghasilkan strategi inventaris yang akurat[2][3].

Algoritma C4.5 memiliki kemampuan untuk membuat struktur pohon keputusan yang jelas dan mudah dipahami yang mencakup komponen utama yang berperan dalam menentukan tingkat kelarisan suatu produk. Selain itu, teknik ini telah banyak digunakan dalam penelitian sebelumnya dan menunjukkan hasil yang baik. Misalnya, C4.5 digunakan untuk memasukkan penjualan pakan ternak ke dalam kategori. Studi memperlihatkan fitur kategori pakan memiliki nilai gain tertinggi, dan model yang dibuat dapat dengan mudah membedakan produk laris dan tidak laris[4]. Algoritma C4.5 juga digunakan untuk mengklasifikasikan komponen motor dengan akurasi sebesar 69,86%; ini membantu mengatasi masalah stok kosong saat permintaan tinggi[5]. Studi lain menggunakan C4.5 untuk menganalisis penjualan rokok dan menemukan akurasi 96,73% yang luar biasa[6]. Ini memperlihatkan algoritma ini sangat baik untuk data ritel dengan atribut yang berbeda[3]. Studi sebelumnya juga menggunakan C4.5 untuk memprediksi penjualan spanduk, dan temuan mereka memperlihatkan metode pohon keputusan dapat membantu mengatasi masalah kurang laku dan penentuan barang laku di toko percetakan"[7]. Studi sebelumnya juga menggunakan C4.5 untuk memprediksi penjualan produk herbal, dan temuan enunjukkan bahwa algoritma C4.5 berhasil membangun model prediksi penjualan yang akurat untuk mengidentifikasi

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9702
Hal 909-918
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

pola pembelian produk herbal[8]. Namun, mayoritas studi tersebut hanya berkonsentrasi pada penerapan algoritma tunggal tanpa mengevaluasi pengaruh kriteria internal pemilihan atribut terhadap kualitas model yang dihasilkan.

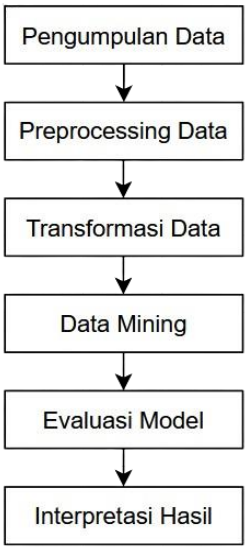
Selain Algoritma C4.5, beberapa metode klasifikasi lain seperti *Naïve Bayes*, *K-Nearest Neighbor (KNN)* dan *Support Vector Machine (SVM)* juga banyak digunakan dalam pengolahan data penjualan. Namun, metode-metode tersebut cenderung berfokus pada minimalisasi eror dan peningkatan akurasi teknis, namun kurang menekankan aspek interpretabilitas model bagi pengguna awam. Dalam konteks distribusi ATK, Algoritma C4.5 tetap sangat relevan karena mampu menghasilkan aturan *If-Then* yang mudah diimplementasikan langsung oleh manajemen dalam operasional bisnis.

Berdasarkan keterbatasan studi terdahulu, penelitian ini menghadirkan kebaruan (novelty) berupa analisis komparatif antara kriteria *Information Gain* dan *Gain Ratio* pada algoritma C4.5 yang diterapkan khusus pada dataset penjualan ATK dengan pola musiman segmen B2S. Penelitian ini mengisi celah (GAP) ilmiah dengan mengevaluasi bagaimana perbedaan kedua metode tersebut mempengaruhi performa model, kompleksitas pohon keputusan, serta tingkat interpretabilitas hasil[9]. Kontribusi utama dari penelitian ini adalah memberikan perspektif yang komprehensif bagi perusahaan dalam memilih metode klasifikasi yang optimal, yang tidak hanya unggul dalam akurasi tetapi juga memberikan aturan keputusan yang stabil dan aplikatif untuk kebutuhan pengambilan keputusan bisnis strategis.

# 2. METODOLOGI PENELITIAN

## 2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan Knowledge Discovery in Database (KDD) yang disesuaikan dengan proses data mining berbasis klasifikasi penjualan[10]. Tahapan penelitian yang telah disusun dapat dilihat pada gambar 1 berikut:




**Gambar 1.** Tahapan Penelitian

Gambar 1 merupakan tahapan penelitian yang dimulai dari pengumpulan data, preprocessing data, transformasi data,data mining, evaluasi model dan interpretasi hail. Adapun penjelasan tahapan sebagai berikut:

a. Pengumpulan Data

Tahap awal penelitian ini diawali dengan pengumpulan data. Data dikumpulkan melalui studi dokumentasi arsip transaksi penjualan perusahaan ATK di Kota Medan periode Januari hingga Desember 2024. Dataset mencakup 647 rekaman transaksi dengan atribut tanggal,nama barang, satuan, jumlah terjual, harga dan total harga. Observasi dan wawancara dilakukan untuk memahami karakteristik data pada segmen *Business-to-School* (B2S).

b. Preprocessing Data

Setelah data terkumpul tahap selanjutnya adalah melakukan preprocessing data, dimulai dengan seleksi data, dimana atribut satuan, harga dan total harga dihapus untuk memfokuskan analisis pada pola volume penjualan serta mengurangi dimensi data yang tidak relevan. Selain itu, dilakukan pembersihan data untuk menangani inkonsistensi, menghapus duplikasi transaksi, dan memastikan tidak terdapat nilai kosong (*missing values*) pada atribut yang tersisa[11]

c. Transformasi Data

Pada tahap ini peneliti mengubah data kedalam format yang sesuai untuk proses analisis. Proses ini meliputi, normalisasi, pengelompokan, serta pengubahan data numerik menjadi kategorikal[12]. Selain itu, transformasi data dilakukan untuk memastikan bahwa algoritma C4.5 dapat memproses semua atribut secara optimal[13].

1. Konversi Format: Atribut Tanggal dikonversi menjadi format Bula*n* (01 Januari 2024 menjadi Januari) untuk menangkap pola penjualan bulanan.

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9702
Hal 909-918
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

2. Pelabelan (*Labeling*): Penetapan atribut target dilakukan dengan metode kategorisasi (*binning*) pada atribut Jumlah Terjual. Menggunakan nilai rata-rata (*average*) sebesar 7 sebagai ambang batas, produk dengan penjualan > 7 dikategorikan sebagai "Laris", sementara 7 dikategorikan sebagai "Tidak Laris". Penggunaan rata-rata dipilih karena mampu memberikan gambaran performa penjualan secara agregat dari seluruh populasi produk selama periode Januari hingga Desember 2024. Setelah proses pelabelan dilakukan, atribut Jumlah Terjual tidak digunakan sebagai atribut prediktor karena telah digunakan sebagai dasar pembentukan label kelas. Penggunaan atribut tersebut secara bersamaan sebagai input klasifikasi berpotensi menyebabkan data leakage dan menghasilkan pohon keputusan yang bias, di mana algoritma C4.5 cenderung memilih atribut tersebut sebagai simpul akar (root node). Kondisi ini menyebabkan model hanya berfokus pada jumlah penjualan dan mengurangi kontribusi atribut lain dalam proses klasifikasi. Oleh sebab itu, atribut Jumlah Terjual dieliminasi dari proses pembentukan model guna menghasilkan struktur pohon keputusan yang lebih representatif dan interpretatif.
3. Rekayasa Fitur (*Feature Engineering*): Tahap *feature engineering* dilakukan dengan menambahkan tiga atribut kunci yang relevan dengan karakteristik segmen B2S: Periode Akademik (Ganjil Awal, Ganjil Akhir, Genap Awal, Genap Akhir), Jenis Pengadaan (Rutin/Persiapan/Berkala). Rutin merupakan pembelian yang dilakukan secara terus menerus atau terencana. Persiapan merupakan pembelian volume besar yang biasanya dilakukan sekolah menjelang awal semester baru. Berkala merupakan pemebelian yang dilakukan berdasarkan jadwal tetap dalam jangka waktu yang lebih lama. Jenis Produk (terbagi dalam 7 kategori utama). Penambahan ini esensial agar algoritma C4.5 dapat mengidentifikasi variasi pola permintaan institusional sekolah pada setiap periode transaksi. Penambahan atribut ini didapat dari metode wawancara kepada pemiliki perusahaan.

d. Data Mining

Pada tahap ini dilakukan proses ekstraksi pola menggunakan algoritma C4.5. Tahap data mining dilakukan dengan menerapkan dua skenario pemilihan atribut (*splitting criteria*). Skenario pertama menggunakan *Information Gain* sebagai metode standar, sedangkan skenario kedua menggunakan *Gain Ratio* sebagai metode pembanding. Pemilihan *Gain Ratio* bertujuan untuk mengatasi kelemahan *Information Gain* yang cenderung bias terhadap atribut dengan banyak nilai unik, seperti variasi nama produk atau rentang harga ATK, sehingga diharapkan menghasilkan pohon keputusan yang lebih stabil dan akurat**.**[14].

e. Evalusi Model Evaluasi

Evaluasi model dilakukan menggunakan *confusion matrix* dengan metrik *accuracy, precision*, dan *recall* karena metode ini mampu mengukur performa klasifikasi secara komprehensif pada model decision tree[15][16]. Pengujian dilakukan dengan teknik **10-Fold Cross Validation** untuk memastikan validitas hasil[17]

f. Interpretasi Hasil

Tahap terakhir adalah menafsirkan hasil dari model yang telah dievaluasi. Hasil dijelaskan secara sistematis agar dapat mengidentifikasi fitur signifikan dalam penentuan kelas, pohon keputusan yang dibuat diperiksa[13]

### 2.2 Data Mining

Data mining adalah teknik untuk mengekstrak pola dan informasi penting dari kumpulan data yang sangat besar sehingga untuk membantu pengambilan keputusan[18]. Dengan algoritma C4.5, data mining dapat menyederhanakan masalah analisis karena kecepatan proses dan struktur aturan yang mudah dipahami[19] Dalam penelitian ini, data mining untuk menganalisis data penjualan ATK, sehingga pola produk terlaris dapat ditentukan dengan lebih akurat daripada hanya mengandalkan perkiraan subjektif.

### 2.3 Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang diperkenalkan oleh Ross Quinlan untuk membangun model klasifikasi berbasis pohon keputusan (*decision tree*)[20][21]. Keunggulan utama algoritma ini terletak pada kemampuannya menangani data numerik (kontinu), data diskrit, serta mengatasi nilai yang hilang (*missing values*) dalam dataset[9]. Proses pembentukan pohon dilakukan dengan memilih atribut yang memiliki nilai *Gain* atau *Gain Ratio* tertinggi sebagai *Root Node*[22]. Pemilihan atribut ini dilakukan secara rekursif pada setiap cabang hingga seluruh data terklasifikasi ke dalam kelas yang sama atau tidak ada lagi atribut yang dapat dibagi[22]. Hasil akhir dari model ini adalah struktur pohon yang dapat diterjemahkan ke dalam aturan keputusan berbentuk *If-Then*, sehingga memberikan tingkat interpretabilitas yang tinggi bagi manajemen dalam pengambilan keputusan strategis[22]. Proses pembentukan pohon keputusan dilakukan melalui tahapan berikut:

a. Perhitungan Nilai Entropy

Perhitungan entropy digunakan untuk mengetahui tingkat keragaman data dalam setiap atribut[23]. Adapun rumus entropy sebagai berikut (1):

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

Keterangan Rumus:

| | | |
|---|---|---|
| S | : | Keseluruhan data(dataset utama) |
| Pi | : | Proposi atau peluang data pada kelas ke-i |
| N | : | Total kelas kategori/ label dalam dataset (misalnya: laris dan tidak laris, berarti n=2) |

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9702
Hal 909-918
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

| | | |
|---|---|---|
| $Log_2(pi)$ | : | Logaritma basis 2 dari peluang kelas tersebut |
| $\sum$ | : | Penjumlahan untuk semua kelas |

b. Perhitungan Information Gain
Information gain digunakan untuk menilai sejauh mana suatu atribut efektif dalam mengelompokkan sampel data data[24]. Berikut ini Rumus untuk mencari Gain (2):

$$Gain(S, A) = Entropy(S) - \sum_{v=1}^{n} \frac{|S_v|}{|S|} \times Entropy(S_v) \quad (2)$$

Keterangan Rumus:

c. Perhitungan Split Info

$$Split\ Info(A) = -\sum_{v=1}^{n} \frac{|S_v|}{|S|} \mathrm{X} \log_2 \frac{|S_v|}{|S|} \quad (3)$$

d. Perhitungan Gain Ratio

$$Gain\ Ratio\ (A) = \frac{\text{Information Gain(S,A)}}{Split\ Info(A)} \quad (4)$$

### 2.4 Evaluasi Model

Evaluasi model dilakukan untuk mengukur efektivitas klasifikasi algoritma C4.5 dalam mendukung keputusan bisnis[25]. Kinerja model diukur menggunakan metode Confusion Matrix yang membandingkan hasil prediksi dengan data aktual melalui empat indikator utama:

1. *True Positive* (TP): Data positif yang diprediksi benar sebagai positif.
2. *True Negative* (TN): Data negatif yang diprediksi benar sebagai negatif.
3. *False Positive* (FP): Data negatif yang diprediksi salah sebagai positif.
4. *False Negative* (FN): Data positif yang diprediksi salah sebagai negatif

Berdasrkan komponen tersebut, kinerja model diukur menggunakan 3 metrik utama yaitu *metrik accuracy, precision*, dan *recall* untuk mengukur performa klasifikasi model[26]. Dengan rumus sebagai berikut rumus sebagai berikut:

a. *Accuracy* untuk mengukur tingkat ketepatan model dalam mengklasifikasikan seluruh data (baik laris maupun tidak laris) secara keseluruhan. Dengan rumus perhitungan sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} x\ 100\% \quad (5)$$

b. *Precision* untuk Mengukur tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem (fokus pada prediksi kelas "Laris"). Dengan rumus perhitungan sebagai berikut:

$$Precision = \frac{TP}{TP+FP} x\ 100\% \quad (6)$$

c. *Recall* untuk Mengukur keberhasilan model dalam menemukan kembali informasi yang benar (fokus pada data yang benar-benar "Laris"). Dengan rumus perhitungan sebagai berikut:

$$Recall = \frac{TP}{TP+FN} x\ 100\% \quad (7)$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil Pengolahan Data

Setelah melalui tahapan *preprocessing* dan *transformasi* data, diperoleh dataset final yang siap digunakan untuk proses *data mining*. Dataset ini mencakup 647 rekaman transaksi penjualan ATK periode Januari hingga Desember 2024. Data tersebut telah melalui proses pembersihan, konversi format tanggal menjadi bulan, serta rekayasa fitur untuk menambahkan atribut yang relevan dengan karakteristik segmen *Business-to-School* (B2S). Berikut adalah Tabel 1 yang menyajikan sampel data hasil transformasi yang akan digunakan dalam pemodelan algoritma C4.5:

**Tabel 1.** Dataset Final

| Bulan | Periode Akademik | Nama Barang | Jenis Produk | Jenis Pengadaan | Kategori |
|---|---|---|---|---|---|
| Januari | Genap Awal | Baterai Alkaline AA | Peralatan lainnya | Persiapan | Laris |
| Januari | Genap Awal | Amplop Coklat/Amplop Soal | Penyimpanan File | Persiapan | Laris |
| Januari | Genap Awal | Amplop Panjang Putih | Penyimpanan File | Persiapan | Laris |
| ... | ... | ... | ... | ... | ... |
| Desember | Ganjil Akhir | Tinta Spidol Snowman Merah | Tinta & Stempel | Berkala | Tidak Laris |
| Desember | Ganjil Akhir | Tinta Stempel Otomatis | Tinta & Stempel | Berkala | Tidak Laris |

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9702
Hal 909-918
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Tabel 1 di atas merepresentasikan struktur data yang akan diolah untuk memprediksi atribut target, yaitu Kategori, yang telah dilabeli menjadi "Laris" dan "Tidak Laris" Atribut-atribut seperti Bulan, Periode Akademik, Jenis Produk, dan Jenis Pengadaan digunakan sebagai prediktor karena kemampuannya dalam menangkap pola musiman permintaan institusional sekolah. Data ini selanjutnya akan menjadi basis perhitungan Algoritma C4.5.

### 3.2 Perhitungan Algoritma C4.5

Tahap awal pemodelan pohon keputusan dimulai dengan menentukan *root node* melalui perhitungan nilai *entropy* total. Diketahui dataset sebanyak 647 rekaman transaksi (267 kategori Laris dan 380 Tidak Laris),

$$Entropy(S) = -\sum_{i=0}^{n} pi \log_2(pi)$$

Penyelesaian:

Entropy (total) = ((-267/647)*$\log_2$ (267/647) + (-380/647)*$\log_2$ (380/647))
= ((-267/647)* (-1.2769) + (-380/647)* (-0.7677))
= ((-0.412673)* (-1.2769) + (-0.587326)* (-0.7677))
= 0.526942 + 0.450890
= 0.9779

Langkah selanjutnya adalah menentukan simpul akar (*root node*) dengan menghitung nilai *Entropy*, *Information Gain* dan *Gain Ratio* untuk setiap atribut prediktorhasil perhitungan kriteria pemilihan atribut disajikan pada Tabel 2 berikut ini

**Tabel 2.** Hasil Perhitungan Algoritma C4.5

| Atribut | Value | Jlh Kasus | Laris | Tidak Laris | Entropy | Information Gain | Split Info | Gain Ratio |
|---|---|---|---|---|---|---|---|---|
| Total | | 647 | 267 | 380 | 0,9779 | | | |
| Bulan | | | | | | 0,292 | | |
| | Januari | 58 | 37 | 21 | 0,9444 | | | |
| | Februari | 46 | 9 | 37 | 0,7131 | | | |
| | Maret | 28 | 6 | 22 | 0,7496 | | | |
| | April | 40 | 14 | 26 | 0,9341 | | | |
| | Mei | 38 | 8 | 30 | 0,7425 | | | |
| | Juni | 50 | 7 | 43 | 0,5842 | | 3,5054 | 0,0833 |
| | Juli | 100 | 95 | 5 | 0,2864 | | | |
| | Agustus | 80 | 35 | 45 | 0,9887 | | | |
| | September | 50 | 34 | 16 | 0,9044 | | | |
| | Oktober | 45 | 2 | 43 | 0,2623 | | | |
| | November | 51 | 10 | 41 | 0,714 | | | |
| | Desember | 61 | 10 | 51 | 0,6436 | | | |
| Jenis Produk | | | | | | 0,0122 | | |
| | Alat Pemotong & Pengukur | 43 | 19 | 24 | 0,9902 | | | |
| | Alat Tulis & Gambar | 115 | 53 | 62 | 0,9956 | | | |
| | Kertas & Buku | 158 | 61 | 97 | 0,9622 | | | |
| | Penyimpanan File | 82 | 43 | 39 | 0,9983 | | 2,6724 | 0,0046 |
| | Peralatan Lainnya | 40 | 19 | 21 | 0,9982 | | | |
| | Perekat & Pengikat | 116 | 42 | 74 | 0,9444 | | | |
| | Tinta & Stempel | 93 | 30 | 63 | 0,9072 | | | |
| Peroiode Akademik | | | | | | 0,1786 | | |
| | Ganjil Akhir | 157 | 22 | 135 | 0,5846 | | | |
| | Ganjil Awal | 230 | 164 | 66 | 0,8648 | | 1,9565 | 0,0913 |
| | Genap Akhir | 128 | 29 | 99 | 0,772 | | | |
| | Genap Awal | 132 | 52 | 80 | 0,9673 | | | |
| Jenis Pengadaan | | | | | | 0,2105 | | |
| | Persiapan | 158 | 132 | 26 | 0,6451 | | 1,5314 | 0,1374 |
| | Berkala | 191 | 79 | 112 | 0,9784 | | | |
| | Rutin | 298 | 56 | 242 | 0,6971 | | | |

*Sumber: Hasil Pengolahan Data Menggunakan Microsoft Excel (2016)*

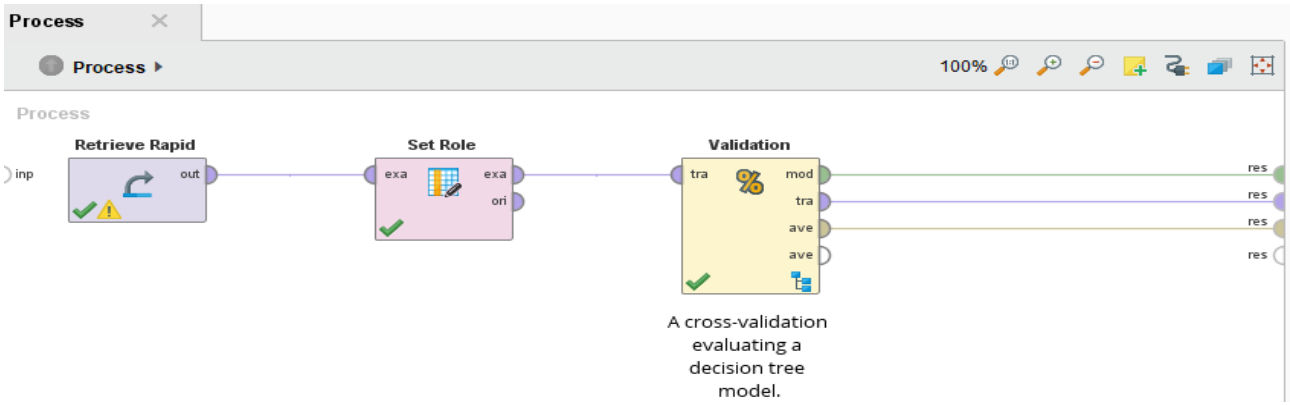
Berdasarkan hasil perhitungan pada Tabel 2, terlihat adanya perbedaan prioritas atribut antara kedua kriteria pemilihan. Atribut Bulan memiliki nilai Information Gain tertinggi (0,292), namun cenderung memiliki nilai *Split Info*

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9702
Hal 909-918
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

yang besar (3,5054) karena memiliki 12 kategori nilai unik. Sebaliknya, atribut Jenis Pengadaan menghasilkan nilai Gain Ratio tertinggi (0,1374), yang menunjukkan bahwa metode ini berhasil melakukan normalisasi untuk mengurangi bias terhadap atribut dengan banyak variasi nilai. Hasil ini menjadi dasar bagi pembentukan dua skenario pohon keputusan yang akan dievaluasi kinerjanya.

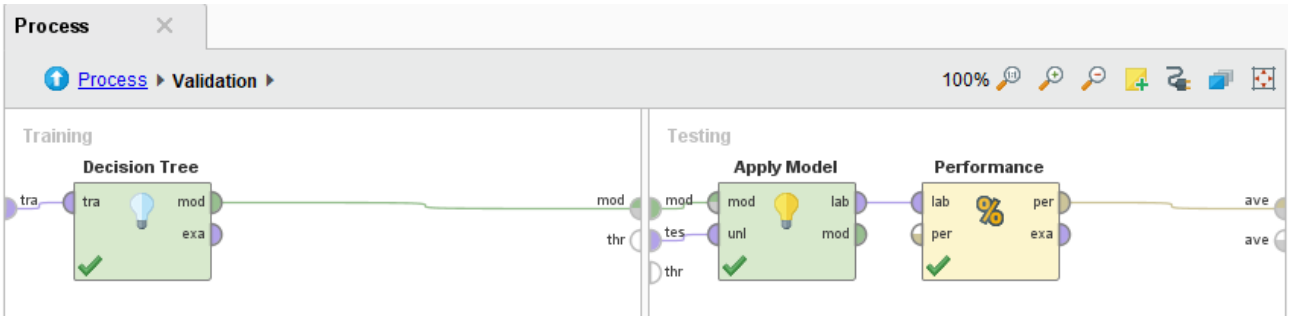
### 3.3 Implementasi Model Menggunakan RapidMiner

Implementasi algoritma C4.5 dalam penelitian ini dilakukan menggunakan perangkat lunak RapidMiner dengan menerapkan prosedur ***10-Fold Cross Validation*** sesuai dengan rancangan pada metodologi. Secara teknis, alur proses pembentukan model klasifikasi, mulai dari pemanggilan dataset hingga proses evaluasi, ditunjukkan pada Gambar 2 berikut:




**Gambar 2.** Tahapan Proses Pembentukan Model Pada RapidMiner

Gambar 2 merepresentasikan rangkaian operator utama dalam lingkungan RapidMiner. Proses diawali dengan operator Retrieve untuk memuat dataset transaksi yang telah diproses sebelumnya. Selanjutnya, operator *Set Role* diterapkan untuk menetapkan atribut Kategori sebagai variabel target (*label*) yang akan diprediksi. Inti dari arsitektur ini terletak pada penggunaan operator *Cross Validation*, yang menerapkan teknik ***10-fold cross validation*** untuk memastikan model memiliki kemampuan generalisasi yang baik serta meminimalkan risiko *overfitting* dalam proses evaluasi. Untuk memberikan gambaran yang lebih mendalam mengenai mekanisme pengujian model, Gambar 3 berikut menyajikan detail sub-proses di dalam operator *Cross Validation* yang membagi dataset menjadi tahap pelatihan (*training*) dan pengujian (*testing*):



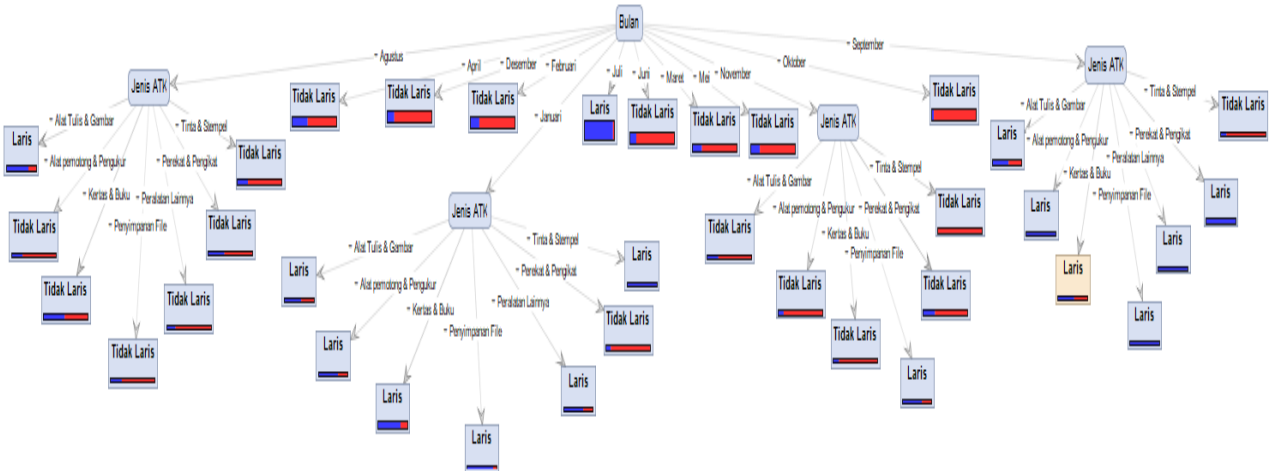

**Gambar 3**. Proses Training dan Testing Model

Berdasarkan visualisasi pada Gambar 3, tahap training menggunakan operator *Decision Tree* untuk mengekstraksi pola dari data latih menggunakan algoritma C4.5. Hasil dari tahap ini berupa model pohon keputusan yang kemudian divalidasi pada tahap testing menggunakan operator *Apply Model* terhadap sisa data uji. Terakhir, operator *Performance* digunakan untuk menghasilkan metrik evaluasi berupa nilai *akurasi, presisi*, dan *recall* yang menjadi indikator efektivitas klasifikasi dalam mendukung keputusan bisnis

### 3.4 Hasil Klasifikasi Skenario 1: Information Gain

Hasil pemodelan algoritma C4.5 pada skenario pertama divisualisasikan dalam bentuk pohon keputusan untuk memberikan gambaran logis mengenai hierarki atribut dalam menentukan klasifikasi produk ATK, sebagaimana ditunjukkan pada Gambar 4 berikut:

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9702
Hal 909-918
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

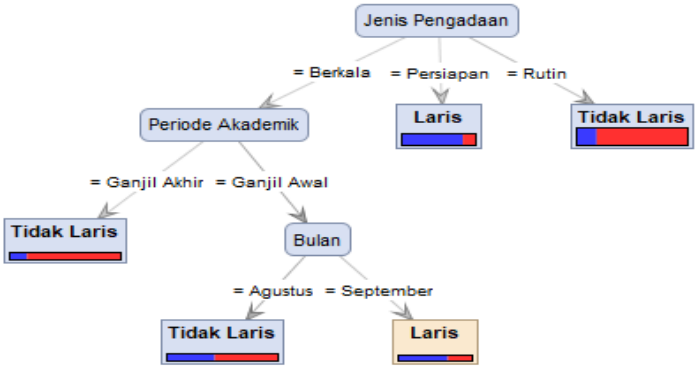



**Gambar 4.** Decision Tree Menggunakan Information Gain

Berdasarkan struktur pohon pada Gambar 4, terlihat bahwa pemilihan Bulan sebagai simpul akar menyebabkan model memiliki percabangan yang luas dan kompleks guna mengakomodasi 12 variasi nilai kategori bulan. Meskipun struktur yang dihasilkan cukup rumit, model ini secara efektif berhasil menangkap pola permintaan musiman yang kuat, seperti pada bulan Juli, serta memberikan performa klasifikasi yang tinggi dengan tingkat akurasi mencapai 78,98% berdasarkan hasil evaluasi ***10-fold cross validation*.** Hasil penelitian ini sejalan dengan penelitian oleh Aditya Narendra Sebastianus et al. yang menunjukkan bahwa algoritma C4.5 mampu menghasilkan performa klasifikasi yang tinggi pada analisis penjualan rokok dengan tingkat akurasi yang baik dalam mendukung pengambilan keputusan bisnis retail[6]. Temuan tersebut memperkuat bahwa Information Gain mampu menangkap pola data penjualan secara detail, terutama pada dataset dengan karakteristik musiman.

### 3.4 Hasil Klasifikasi Skenario 2: Gain Ratio

Implementasi skenario kedua dilakukan dengan menerapkan kriteria *Gain Ratio* guna mengatasi kelemahan kriteria sebelumnya yang cenderung bias terhadap atribut dengan banyak nilai unik. Melalui proses normalisasi menggunakan *Split Information*, diperoleh atribut Jenis Pengadaan sebagai simpul akar (*root node*). Struktur pohon keputusan yang dihasilkan dari skenario ini divisualisasikan pada Gambar 5 berikut ini:




**Gambar 5.** Decision Tree Menggunakan Gain Ratio

Berdasarkan visualisasi pada Gambar 5, terlihat bahwa penggunaan *Gain Ratio* menghasilkan struktur pohon yang jauh lebih ringkas, sederhana, dan stabil dibandingkan skenario pertama. Model ini menghasilkan aturan keputusan (*rules*) yang sangat aplikatif bagi manajemen; sebagai contoh, jika Jenis Pengadaan bersifat Persiapan, maka produk diklasifikasikan sebagai Laris dengan dukungan 132 data, sedangkan pengadaan bersifat Rutin diklasifikasikan sebagai Tidak Laris dengan 242 data. Meskipun skenario ini menghasilkan tingkat akurasi sebesar 73,26% yang secara teknis sedikit lebih rendah dari *Information Gain* model ini dinilai lebih optimal untuk pengambilan keputusan bisnis strategis karena memiliki tingkat interpretabilitas yang lebih tinggi dan lebih mudah dipahami oleh pengguna awam dalam operasional distributor ATK. Temuan ini sejalan dengan penelitian Widowati yang menjelaskan bahwa pendekatan pemilihan atribut berbasis normalisasi pada decision tree mampu mengurangi risiko *overfitting* serta menghasilkan model yang lebih stabil dan mudah diinterpretasikan[14]. Hal ini memperlihatkan bahwa *Gain Ratio* lebih efektif digunakan pada dataset dengan atribut yang memiliki banyak variasi nilai.

### 3.5 Evaluasi dan Perbandingan Performa Model

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9702
Hal 909-918
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Evaluasi performa kedua skenario dilakukan menggunakan *Confusion Matrix* melalui teknik ***10-Fold Cross Validation*** untuk menjamin validitas model. Ringkasan perbandingan metrik performa disajikan pada Tabel 3:

**Tabel 3.** Perbandingan Peforma Model

| Skenario | Accuracy | Precision | Recall |
|---|---|---|---|
| Information Gain | 78,98 % | 78,86 % | 87,90 % |
| Gain Ratio | 73,26 % | 75,16 % | 83,17 % |

*Sumber: Hasil Pengolahan Data Menggunakan RapidMiner (2016)*

Hasil pengujian menunjukkan skenario *Information Gain* unggul secara teknis dengan akurasi 78,98%, sementara *Gain Ratio* memperoleh 73,26%. Fokus utama pada kelas Tidak Laris memperlihatkan nilai *Recall* yang tinggi pada kedua skenario (87,89% dan 83,16%), yang mengindikasikan kemampuan model dalam memprediksi produk berisiko rendah terjual guna menekan kerugian akibat *overstock*. Meskipun skenario pertama unggul secara numerik, skenario kedua tetap menjadi pertimbangan penting karena menghasilkan struktur pohon yang lebih sederhana dan aplikatif bagi manajemen. Hasil evaluasi performa pada penelitian ini memperlihatkan bahwa kedua skenario memiliki kemampuan klasifikasi yang cukup baik berdasarkan nilai *accuracy*, *precision*, dan *recall*. Temuan ini sejalan dengan penelitian oleh Zelvi Gustiana yang menyatakan bahwa evaluasi model menggunakan *confusion matrix* mampu memberikan gambaran komprehensif terhadap performa algoritma C4.5 dalam proses klasifikasi data[26]. Selain itu, penelitian Musa Dandi Muhamad et al. juga menunjukkan bahwa algoritma C4.5 efektif digunakan untuk klasifikasi data penjualan karena mampu menghasilkan performa klasifikasi yang stabil berdasarkan atribut penjualan yang relevan[4]

### 3.6 Pembahasan dan Interpretasi Hasil

Tahap interpretasi hasil merupakan bagian akhir dari siklus *Knowledge Discovery in Database* (KDD) yang bertujuan untuk mentransformasi pola-pola yang ditemukan menjadi pengetahuan yang aplikatif bagi manajemen. Berdasarkan hasil penelitian menggunakan algoritma C4.5 pada dataset produk ATK segmen B2S, berikut adalah analisis mendalam terhadap temuan tersebut:

a. Efektivitas Kriteria Pemilihan Atribut (Information Gain vs. Gain Ratio)
   Analisis komparatif antara kedua skenario menunjukkan perbedaan mendasar pada pemilihan simpul akar (*root node*) yang berdampak langsung pada kualitas model. Skenario *Information Gain* menetapkan atribut Bulan sebagai akar karena memiliki nilai *gain* tertinggi sebesar 0,292. Namun, pemilihan ini memicu bias karena atribut tersebut memiliki 12 kategori unik, sehingga model secara agresif memetakan pola transaksi spesifik tahun 2024. Dampaknya, meskipun *Information Gain* mencapai akurasi teknis lebih tinggi yaitu 78,98%, model menghasilkan struktur pohon yang sangat kompleks dan rimbun yang mengindikasikan gejala *overfitting*. Sebaliknya, kriteria *Gain Ratio* berhasil mengoreksi bias tersebut melalui mekanisme normalisasi menggunakan *Split Information*. Pendekatan ini menggeser prioritas akar ke atribut Jenis Pengadaan yang memiliki nilai *Gain Ratio* tertinggi sebesar 0,1374. Hasilnya adalah struktur pohon yang jauh lebih ringkas, sederhana, dan stabil dengan tingkat akurasi 73,26%. Walaupun secara numerik akurasinya sedikit di bawah *Information Gain*, model *Gain Ratio* terbukti lebih handal untuk pengambilan keputusan bisnis karena hanya menggunakan atribut dengan daya pembeda kuat yang relevan dengan logika operasional distributor. Temuan ini mengonfirmasi bahwa pada dataset dengan atribut bervariasi banyak, *Gain Ratio* lebih unggul dalam menghasilkan model yang stabil, mudah diinterpretasikan, dan terhindar dari *overfitting.*
b. Pola Permintaan Musiman dan Faktor Dominan
   Penelitian ini berhasil mengidentifikasi bahwa Jenis Pengadaan adalah faktor paling dominan dalam menentukan kelarisan produk ATK pada segmen B2S. Berdasarkan pola yang diekstraksi dari skenario *Gain Ratio*, ditemukan aturan keputusan yang sangat signifikan:
   1. Jika Jenis Pengadaan bersifat "Persiapan", maka produk diprediksi "Laris" (didukung oleh 132 kasus Laris dan hanya 26 Tidak Laris)
   2. Jika Jenis Pengadaan bersifat "Rutin", maka produk secara konsisten diprediksi "Tidak Laris" (56 Laris vs 242 Tidak Laris)

   Pola ini mencerminkan karakteristik musiman segmen B2S, di mana lonjakan permintaan terjadi pada periode persiapan awal semester. Atribut Periode Akademik juga terbukti memperkuat prediksi pada jenis pengadaan tertentu, memperlihatkan keterkaitan erat antara kalender pendidikan dengan volume transaksi distributor.
c. Rekomendasi Model untuk Pengambilan Keputusan Bisnis
   Meskipun skenario *Information Gain* secara teknis memiliki akurasi sedikit lebih tinggi (78,98%) dibandingkan *Gain Ratio* (73,26%), penelitian ini merekomendasikan model Gain Ratio sebagai metode yang lebih optimal untuk implementasi bisnis strategis. Hal ini didasarkan pada aspek interpretabilitas model; aturan *If-Then* yang dihasilkan oleh *Gain Ratio* jauh lebih ringkas dan mudah dipahami oleh manajemen perusahaan. Dalam konteks distribusi ATK, kemampuan manajer untuk memahami logika model jauh lebih berharga daripada selisih akurasi yang kecil, karena model tersebut dapat langsung digunakan untuk mengatur stok secara akurat guna menghindari risiko *overstock* atau *stock-out*. Keseimbangan antara performa teknis dan nilai praktis ini menjadi kontribusi utama penelitian dalam membantu distributor mengelola inventaris secara lebih saintifik.

# 4. KESIMPULAN

Penelitian ini berhasil menyimpulkan bahwa algoritma C4.5 efektif dalam mengklasifikasikan produk ATK terlaris pada segmen *Business-to-School* (B2S), di mana hasil analisis komparatif menunjukkan kriteria *Information Gain* unggul secara teknis dengan akurasi sebesar 78,98%, sementara kriteria *Gain Ratio* menghasilkan akurasi 73,26%. Meskipun memiliki akurasi yang sedikit lebih rendah, *Gain Ratio* ditetapkan sebagai metode yang paling optimal untuk pengambilan keputusan bisnis strategis karena melalui normalisasi *Split Information*, kriteria ini menghasilkan struktur pohon keputusan yang jauh lebih sederhana, stabil, dan mudah diinterpretasikan oleh pihak manajemen dibandingkan *Information Gain* yang cenderung kompleks dan bias terhadap atribut bervariasi banyak. Atribut Jenis Pengadaan teridentifikasi sebagai faktor paling dominan dalam menentukan tingkat kelarisan produk, dengan pola pengadaan "Persiapan" yang secara signifikan berkorelasi dengan kategori "Laris". Hasil ini memberikan kontribusi praktis bagi distributor dalam meminimalisir risiko *overstock* dan *stock-out* secara saintifik. Namun, penelitian ini memiliki keterbatasan pada penggunaan dataset yang hanya mencakup satu periode tahunan (Januari–Desember 2024), sehingga disarankan bagi penelitian selanjutnya untuk memperluas cakupan data lintas tahun guna memperkuat validitas pola musiman jangka panjang. Selain itu, pengembangan kedepannya dapat dilakukan melalui eksperimen komparatif dengan algoritma klasifikasi lain seperti *Naïve Bayes*, *Random Forest*, atau *Support Vector Machine* (SVM) untuk mengeksplorasi potensi peningkatan akurasi tanpa mengorbankan aspek kemudahan interpretasi model bagi pengguna awam

# UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada dosen pembimbing atas arahan dan bimbingan pada selama proses penelitian. Penulis menyampaikan terima kasih pada perusahaan yang telah bersedia memberikan data penjualan untuk keperluan penelitian ini. Selain itu, penulis mengapresiasi pihak Program Studi atas dukungan, arahan, dan masukan terkait penyusunan penelitian. Ucapan terima kasih disampaikan pada semua pihak yang telah membantu dalam pengumpulan data, pengolahan data, hingga penyusunan artikel ini sehingga penelitian bisa diselesaikan dengan baik.

# REFERENCES

[1] M. Y. Zidane, B. N. Sari, I. Maulana, A. Primaya, and Garno, "Penerapan Data Mining Dalam Klasifikasi Data Transaksi Produk Koperasi Di SMK PGRI 2 Karawang," JATI (Jurnal Mahasiswa Teknik Informatika), Feb. 2025.

[2] Mulyanda Sandy, Defit Sarjon, and Sumijan, "Analisis Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Harga Pasar Mobil Bekas," *Jurnal KomtekInfo*, no. 3, pp. 116–121, Sep. 2023, doi: 10.35134/komtekinfo.v10i3.427.

[3] Aditya Narendra Sebastianus, Aldi Setiawan Pradita, Anwar Syaiful, and Haddiel Fuad Mohammad, "Analisa Penjualan Rokok Dengan Metode Klasifikasi Menggunakan Metode Algoritma C4.5 Pada CV Jaya Berkah Mas," *Jurnal INSAN (Journal of Information Systems Management Innovation*, vol. 4, no. 2, pp. 91–100, Dec. 2024, [Online]. Available: http://jurnal.bsi.ac.id/index.php/jinsan

[4] Musa Dandi Muhamad *et al.*, "Penerapan Data Mining Untuk Klasifikasi Data Penjualan Pakan Ternak Terlaris Dengan Algoritma C4.5," *Jurnal Teknologi Informatika dan Komputer MH. Thamrin*, vol. 10, no. 1, pp. 168–1862, Mar. 2024, doi: 10.37012/jtik.v10i1.1985.

[5] N. S. Sitorus and D. Leman, "Implementasi Data Mining Dalam Klasifikasi Produk Laris Dan Tidak Laris Dengan Algoritma C4.5 Implementation Of Data Mining In Classification Of Best-Selling And Non-Selling Products With The C4.5 Algorithm," *Tanjung Mulia, Kec.Medan Deli*, vol. 3, no. 2, pp. 423–436, 2025, doi: 10.22303/upu.1.1.2021.01-10.

[6] Aditya Narendra Sebastianus, Aldi Setiawan Pradita, Anwar Syaiful, and Haddiel Fuad Mohammad, "Analisa Penjualan Rokok Dengan Metode Klasifikasi Menggunakan Metode Algoritma C4.5 Pada CV Jaya Berkah Mas," *Jurnal INSAN (Journal of Information Systems Management Innovation)*, vol. 4, no. 2, pp. 91–100, Dec. 2024, [Online]. Available: http://jurnal.bsi.ac.id/index.php/jinsan

[7] Triawan Bagus, Lubis Imran, and Kadim Lina Arlina Nur, "Penerapan Data Mining Untuk Prediksi Penjulan Spanduk Menggunakan Algoritma C4.5," *Journal of Mathematics and Technology (MATECH)*, vol. 3, no. 2, pp. 149–157, Nov. 2024.

[8] A. J. Kahfi, F. A. Djendra, Y. P. Ananda, Z. Wijayanti, N. Khoerotunnisa, and B. O. Lubis, "ANALISIS DATA PENJUALAN PRODUK HERBAL MENGGUNAKAN ALGORITMA C4.5 PADA E-COMMERCE AZKA JAISY STORE," Apr. 2025.

[9] Nasrullah Asmaul Husnah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris," *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas AL Asyariah Mandar*, vol. 7, no. 2, pp. 45–51, Sep. 2021, [Online]. Available: http://ejournal.fikom-unasman.ac.id

[10] A. B. Almagribi and S. Redjeki, "Clustering and Classification of Retail Sales Data: A Big Data and Data Mining Analysis," *Journal Innovations Computer Science*, vol. 4, no. 2, pp. 242–253, Nov. 2025, doi: 10.56347/jics.v4i2.303.

[11] D. L. S. Purnama and U. Apsiswanto, "Analysis of C4.5 Algorithm Performance for Predicting Student Achievement Based on Socio-Economic Status, Motivation, Discipline, and Past Achievement," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 7, no. 1, pp. 190–199, Jan. 2025, doi: 10.47709/cnahpc.v7i1.5143.

[12] C. C. Aggarwal, *Data Mining: The Textbook*. Springer International Publishing, 2015. doi: 10.1007/978-3-319-14142-8.

[13] F. G. Falentina, G. Y. Y. Wabdaron, D. Andiyani, M. S. Wondiwoi, and H. Sutejo, "Analisis Penerapan Data Mining Untuk Klasifikasi Penjualan Makanan Terlaris Menggunakan Algoritma Decision Tree (C4.5)," *Jurnal Ilmiah Sistem Informasi*, vol. 4, no. 2, pp. 382–395, Jan. 2026, doi: 10.51903/f2jegk76.

[14] F. U. Widowati, "Application of C4.5 algorithm with PSO Feature Selection and Bagging Technique on Breast Cancer Classification," *International Journal of Management Science and Information Technology*, vol. 4, no. 2, pp. 312–320, Aug. 2024, doi: 10.35870/ijmsit.v4i2.3061.

[15] Z. Gustiana, "Performance Evaluation Algoritma C4.5 Pada Klasifikasi Data," *Jurnal Teknologi Informasi*, vol. 5, no. 2, 2024, doi: 10.46576/djtechno.
[16] A. Kristina and S. Rukiastiandari, "PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI KELAYAKAN PENERIMA PROGRAM INDONESIA PINTAR (PIP) DI SD NEGERI 13 JONGKONG," *JURNAL TEKNOLOGI INFORMASI*, no. 2, pp. 156–169, Dec. 2025, doi: 10.52972/hoaq.vol16no2.
[17] K. A. Putri, D. Febriawan, and F. N. Hasan, "Implementation of Data Mining to Predict Student Study Period with Decision Tree Algorithm (C4.5)," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 13, no. 1, pp. 31–39, Feb. 2024, doi: 10.32736/sisfokom.v13i1.1943.
[18] I. Gede Iwan Sudipa *et al.*, *DATA MINING*. 2023. [Online]. Available: www.globaleksekutifteknologi.co.id
[19] L. Winda Sari Siburian, D. Saripurna, and S. Kusnasari, "Analisis Tingkat Kepuasan Masyarakat Terhadap Pelayanan Kantor Desa Dengan Menggunakan Algoritma C4.5," *Jurnal Sistem Informasi TGD*, vol. 3, no. 2, pp. 263–273, Mar. 2024, [Online]. Available: https://ojs.trigunadharma.ac.id/index.php/jsi
[20] F. Pirmansyah and T. Wahyudi, "Implementasi Data Mining Menggunakan Algoritma C4.5 untuk Prediksi Evaluasi Anggota Satuan Pengamanan : Studi Kasus PT. YIMM Pulogadung," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 3, no. 8, pp. 343–356, Sep. 2023, doi: 10.52436/1.jpti.294.
[21] Arupandani Widya Windaru, Taufik Faisal, and Mahyuni Rina, "Implementasi Data Mining Menentukan Penerimaan Bantuan Sosial Pangan (BSP) Menggunakan Algoritma C4.5," *Jurnal Sistem informasi TGD*, vol. 2, no. 5, pp. 705–715, Sep. 2023, [Online]. Available: https://ojs.trigunadharma.ac.id/index.php/jsi
[22] E. Adinda Yestina *et al.*, "Implementasi Algoritma C4.5 pada Analisis Faktor Risiko Penyakit Jantung Koroner," *(Jurnal Riset Komputer)*, vol. 12, no. 5, pp. 2407–389, 2025, doi: 10.30865/jurikom.v12i5.8852.
[23] T. Y. Pratama and A. Armansyah, "Decision Tree C4.5 dengan Teknik Information Gain Untuk Klasifikasi Pemilihan Program Studi Tingkat Lanjut," *Journal of Information System Research (JOSH)*, vol. 5, no. 4, pp. 1042–1052, Jul. 2024, doi: 10.47065/josh.v5i4.5643.
[24] E. Adhi Guna, M. Davin Diza Ghifary, E. Fransiska Sihombing, and A. Pius Datubara, "Implementasi Algoritma Decision Tree untuk Klasifikasi Data Evaluation Car Menggunakan Python," *Jurnal Sistem Informasi dan Ilmu Komputer*, vol. 1, no. 4, pp. 167–177, 2023, doi: 10.59581/jusiik-widyakarya.v1i4.
[25] Putri Kirana Alyssa, Febriawan Dimas, and Hasan Firman Noor, "Implementation of Data Mining to Predict Student Study Period with Decision Tree Algorithm (C4.5)," *Jurnal SISFOKOM (Sistem Informasi dan Komputer)*, vol. 13, no. 1, pp. 39–47, Feb. 2024, doi: 10.32736/sisfokom.v13i1.1943.
[26] Gustiana Zelvi, "PERFORMANCE EVALUATION ALGORITMA C 4.5 PADA KLASIFIKASI DATA," *Jurnal Teknologi Informasi*, vol. 5, no. 2, pp. 289–296, Aug. 2024, doi: 10.46576/djtechno.