

Analisis Profit Penjualan Menggunakan Algoritma C4.5

A.Nurul Asyilah Rizky*, Putri Tandi Langi Paongan

Teknik Informatika, Universitas Dipa Makassar, Makassar, Indonesia

e-mail: *¹nurul.andi0117@gmail.com , ²putritandilangipaongan@gmail.com

Abstrak

Penelitian ini bertujuan untuk menganalisis pola profit penjualan pada Toko Kopi Hope menggunakan algoritma C4.5 berdasarkan data transaksi penjualan periode Mei 2025 hingga Januari 2026. Data penelitian terdiri dari 9.280 transaksi yang diolah melalui tahapan pembersihan dan transformasi data sehingga diperoleh data profit harian selama 263 hari operasional. Nilai profit dikelompokkan ke dalam tiga kategori, yaitu profit rendah, sedang, dan tinggi, menggunakan metode entropy-based discretization. Proses klasifikasi dilakukan dengan membangun model algoritma C4.5 dengan pembagian data training dan testing sebesar 70:30 menggunakan stratified sampling. Evaluasi model menggunakan confusion matrix menghasilkan tingkat akurasi sebesar 88,61%. Decision tree yang dihasilkan menunjukkan bahwa variabel Subtotal, HPP Total, dan Quantity merupakan faktor utama yang memengaruhi profit penjualan. Aturan keputusan yang dihasilkan dapat digunakan sebagai dasar pengambilan keputusan bisnis yang lebih terarah dan berbasis data.

Kata kunci—Algoritma C4.5, Klasifikasi Profit, Decision Tree, Data Mining, Entropy-Based Discretization.

Abstract

This study aims to analyze sales profit patterns at Toko Kopi Hope using the C4.5 algorithm based on transaction data from May 2025 to January 2026. The dataset consists of 9,280 transactions processed through data cleaning and transformation stages to generate daily profit data covering 263 operational days. Profit values were classified into three categories—low, medium, and high—using entropy-based discretization. Profit classification was conducted by building a C4.5 model with a 70:30 split between training and testing data using stratified sampling. Model performance evaluated using a confusion matrix achieved an accuracy of 88.61%. The resulting decision tree reveals that Subtotal, Total Cost of Goods Sold, and Quantity are the most influential factors. The generated decision rules can support data-driven business decision-making.

Keywords—C4.5 Algorithm, Profit Classification, Decision Tree, Data Mining, Entropy-Based Discretization.

1. PENDAHULUAN

Kedai kopi merupakan tempat yang menjual minuman berbahan kopi dan berbagai menu lain non-kopi, sekaligus menjadi ruang yang nyaman untuk bekerja, belajar, atau berkumpul. Jumlah kedai kopi di Indonesia terus meningkat dari tahun ke tahun. Laporan sebelumnya menunjukkan adanya kenaikan jumlah kedai kopi dari 1.083 gerai pada tahun 2016 menjadi 2.937 gerai pada tahun 2019 [1]. Peningkatan ini menggambarkan bahwa minat masyarakat, khususnya kalangan mahasiswa dan pekerja, terhadap aktivitas di kedai kopi cukup tinggi [2].

Setiap transaksi di kedai kopi menghasilkan struk penjualan yang berisi informasi seperti nama produk, jumlah pembelian atau Quantity (QTY), harga jual, metode pembayaran, Harga

Pokok Penjualan (HPP), dan profit. Jika dimanfaatkan dengan baik, data ini dapat membantu pemilik usaha memahami pola pembelian pelanggan [3]. Namun, data struk yang tersimpan biasanya belum diolah dan belum tersusun dengan baik sehingga tidak langsung memberikan informasi yang jelas untuk analisis.

Agar data struk penjualan bisa digunakan secara optimal, diperlukan proses data processing. Tahapan ini meliputi perbaikan format data, pembersihan nilai yang tidak valid, serta pembuatan variabel baru seperti Subtotal, HPP Total, dan Profit. Tanpa melalui proses processing, data tidak akan siap untuk dianalisis lebih lanjut dan pola transaksi pelanggan tidak dapat dianalisis secara akurat [4]. Data Mining dalam pemasaran memudahkan pengelola bisnis untuk mengerti tren tersembunyi di balik transaksi pembelian sehingga membantu dalam perencanaan promosi pemasaran secara lebih efisien [5].

Dalam konteks Toko Kopi Hope, data struk penjualan yang tersedia menyimpan informasi penting mengenai perilaku pembelian pelanggan dan besarnya profit pada setiap transaksi, tetapi belum pernah diolah secara sistematis untuk keperluan analisis profit penjualan. Kondisi ini menyebabkan pemilik usaha belum memperoleh gambaran yang jelas mengenai transaksi mana yang memberikan profit tinggi, sedang, atau rendah, serta variabel apa saja yang paling berpengaruh terhadap profit tersebut.

Dalam penelitian ini, algoritma C4.5 dipilih karena merupakan metode klasifikasi yang mampu membentuk model pohon keputusan yang relatif mudah dipahami. Algoritma ini menentukan atribut yang paling berpengaruh terhadap profit melalui perhitungan entropy, information gain, dan gain ratio, sehingga menghasilkan model berupa pohon keputusan yang berisi aturan-aturan (rules) untuk mengklasifikasikan transaksi ke dalam kategori profit tertentu [6]. Algoritma C4.5 juga terbukti mampu mengelola data numerik dan kategorikal sekaligus, serta menghasilkan pohon keputusan yang mudah diinterpretasikan [7].

Penelitian terdahulu oleh Falentina et al. [8] menerapkan C4.5 untuk mengklasifikasikan menu penjualan terlaris pada usaha kuliner dengan akurasi 84%. Penelitian Muktar et al. [9] menggunakan C4.5 pada data penjualan pakaian online Shopee. Sulastri [10] menerapkan C4.5 untuk identifikasi tingkat penjualan produk herbal. Hamzah et al. [11] menerapkan data mining untuk klasifikasi data penjualan pakan ternak menggunakan C4.5. Azwanti [12] menggunakan C4.5 untuk memprediksi penjualan motor. Penelitian-penelitian tersebut menunjukkan bahwa C4.5 adalah metode yang efektif dalam analisis penjualan, namun belum ada yang berfokus pada analisis profit harian kedai kopi dengan pendekatan entropy-based discretization untuk penentuan kategori.

Penelitian ini berfokus pada analisis profit penjualan melalui serangkaian proses data processing dan penerapan algoritma C4.5 untuk mengelompokkan transaksi ke dalam kategori profit tinggi, sedang, dan rendah, sehingga diharapkan dapat memberikan informasi mengenai variabel-variabel yang paling berpengaruh dan mendukung pengambilan keputusan bisnis yang lebih tepat dan berbasis data [13], [14].

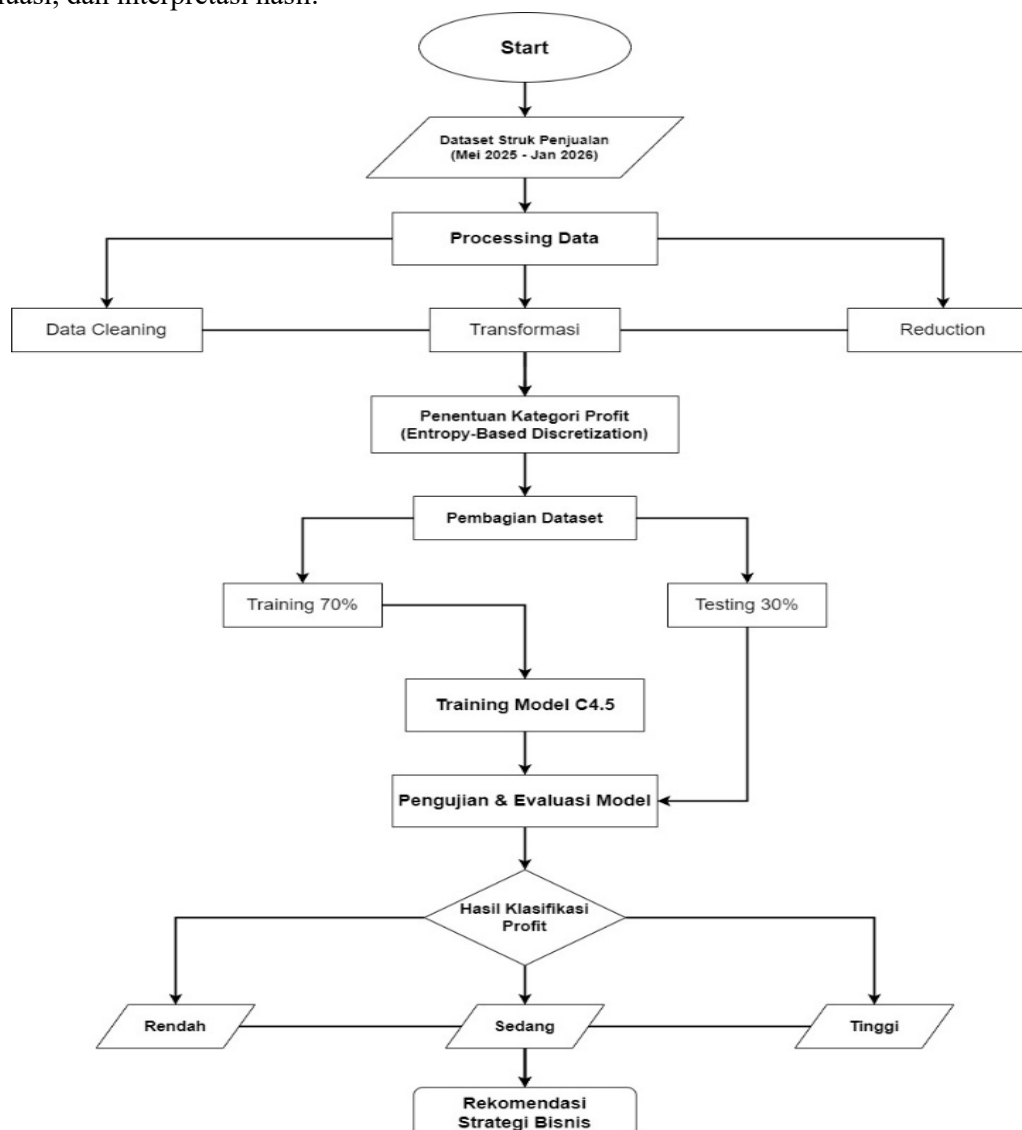
2. METODE PENELITIAN

2.1 Sumber Data dan Pengumpulan Data

Penelitian ini menggunakan data transaksi penjualan Toko Kopi Hope yang berlokasi di Gowa, Samata. Data berupa laporan penjualan bulanan dalam format Microsoft Excel yang mencakup periode Mei 2025 hingga Januari 2026. Data transaksi memuat informasi tanggal transaksi, produk dan kategori produk, Quantity (QTY), harga jual, Harga Pokok Penjualan (HPP), dan metode pembayaran. Pengumpulan data dilakukan melalui metode dokumentasi dan observasi tidak langsung terhadap catatan transaksi yang telah tersimpan dalam sistem pencatatan toko.

2.2 Prosedur Penelitian

Alur perancangan solusi pada penelitian ini ditampilkan pada Gambar 1 berikut. Prosedur penelitian meliputi data processing, penentuan kategori profit, pembangunan model C4.5, evaluasi, dan interpretasi hasil.



Gambar 1 Flowchart Perancangan Solusi

2.3 Data Peocessing

Tahapan data processing dalam penelitian ini terdiri dari tiga tahap utama, yaitu data cleaning, transformasi data, dan data reduction. Pada tahap data cleaning, dilakukan penghapusan baris kosong, penyeragaman nama kolom, konversi tipe data, pengecekan duplikasi, serta penghapusan nilai yang tidak valid. Dataset awal yang berjumlah 9.323 baris berkurang menjadi 9.280 baris setelah proses cleaning, sehingga terdapat 43 baris data yang dihapus.

Tahap kedua adalah transformasi data, yaitu pembuatan variabel turunan yang digunakan dalam analisis profit, yaitu Subtotal, HPP Total, dan Profit. Perhitungan variabel tersebut menggunakan rumus berikut:

$$\text{Profit} = \text{Subtotal} - \text{HPP Total} \quad (1)$$

$$\text{Subtotal} = \text{Harga Jual} \times \text{QTY} \quad (2)$$

$$\text{HPP Total} = \text{HPP} \times \text{QTY} \quad (3)$$

Ketiga, Data Reduction menghapus kolom yang tidak berkaitan langsung dengan analisis profit. Setelah transformasi, data diintegrasikan menjadi 263 data harian yang menjadi input utama model C4.5.

2.4 Penentuan Kategori Profit (Entropy-Based Discretization)

Karena algoritma C4.5 lebih cocok untuk data kategori, nilai profit harian yang masih numerik dikonversi menjadi tiga kelompok menggunakan metode entropy-based discretization. Metode ini menggunakan perhitungan statistik data untuk menentukan batas kategori secara objektif, bukan secara manual. Rumus entropy yang digunakan:

$$Entropy(S) = -\sum p_i \log^2(p_i) \quad (4)$$

Keterangan: S = himpunan data, p_i = proporsi data pada kelas ke-i, n = jumlah kelas.

Weighted Entropy dihitung sebelum Information Gain untuk mengetahui entropy yang tersisa setelah pemisahan:

$$WE = \left(\frac{|S^1|}{|S|}\right) \times Entropy(S^1) + \left(\frac{|S^2|}{|S|}\right) \times Entropy(S^2) \quad (5)$$

Information Gain dihitung dengan:

$$Gain(S, A) = Entropy(S) - \sum \left(\frac{|S_v|}{|S|}\right) \times Entropy(S_v) \quad (6)$$

Berdasarkan hasil perhitungan pada dataset, diperoleh dua batas kategori: Batas pertama = Rp 409.817 dan Batas kedua = Rp 425.736. Sehingga: Profit Rendah $\leq$ Rp 409.817, Profit Sedang antara Rp 409.817 – Rp 425.736, dan Profit Tinggi $>$ Rp 425.736.

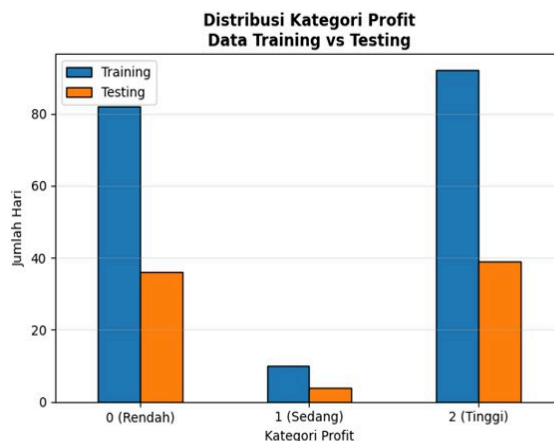
Tabel 1 Distribusi Kategori Profit Harian

Kategori Profit	Jumlah Hari	Persentase	Karakteristik
Rendah	118 hari	44.9%	Profit $\leq$ Rp 409.817
Sedang	14 hari	5.3%	Rp 409.817 – Rp 425.736
Tinggi	131 hari	49.8%	Profit $>$ Rp 425.736
TOTAL	263 hari	100.0%	-

2.5 Pembagian Data dan Algoritma C4.5

Dataset penelitian terdiri dari 263 data profit harian yang dibagi menjadi 184 data (70%) sebagai data training dan 79 data (30%) sebagai data testing menggunakan metode stratified sampling agar proporsi setiap kategori profit tetap seimbang pada kedua kelompok data. Algoritma C4.5 diimplementasikan menggunakan library scikit-learn dengan parameter `criterion='entropy'` pada Python 3.10 melalui Google Colab.

Pemilihan rasio 70:30 bertujuan menjaga keseimbangan antara proses pembelajaran dan evaluasi model. Proporsi 70% data training memberikan jumlah data yang cukup bagi model untuk mempelajari pola klasifikasi secara representatif, sedangkan 30% data testing digunakan untuk menguji kemampuan model pada data yang tidak dilibatkan dalam proses pelatihan. Pembagian ini juga umum digunakan pada penelitian klasifikasi dengan dataset berukuran kecil hingga menengah karena mampu menghasilkan evaluasi model yang lebih reliabel.



Gambar 2 Distribusi Kategori Profit Data Training vs Testing

Distribusi kategori profit pada data training (184 data) terdiri dari 82 profit rendah (44,6%), 10 profit sedang (5,4%), dan 92 profit tinggi (50,0%). Sementara pada data testing (79 data) terdapat 36 profit rendah (45,6%), 4 profit sedang (5,1%), dan 39 profit tinggi (49,4%). Proporsi yang relatif konsisten menunjukkan bahwa metode stratified sampling berhasil menjaga keseimbangan distribusi kelas sehingga model dapat dilatih tanpa bias terhadap kategori tertentu. Parameter `random_state` digunakan untuk memastikan proses pembagian data dapat direproduksi secara konsisten.

Model kemudian dilatih menggunakan data training dan diuji menggunakan data testing. Performa model dievaluasi menggunakan confusion matrix serta metrik accuracy, precision, recall, dan F1-score untuk menilai kemampuan model dalam mengklasifikasikan kategori profit.

Dalam proses pembangunan pohon keputusan menggunakan algoritma C4.5, pemilihan atribut terbaik dilakukan melalui perhitungan Entropy, Information Gain, Split Information, dan Gain Ratio. Atribut dengan nilai Gain Ratio tertinggi dipilih sebagai root node pada pohon keputusan. Rumus Split Information:

$$SplitInfo(A) = -\sum \left(\frac{|S_v|}{|S|} \right) \times \log^2 \left(\frac{|S_v|}{|S|} \right) \quad (7)$$

Rumus Gain Ratio:

$$GainRatio(A) = \frac{Gain(S, A)}{SplitInfo(A)} \quad (8)$$

Berdasarkan perhitungan, atribut Subtotal memperoleh nilai Gain Ratio tertinggi sebesar 0.8142, sehingga dipilih sebagai root node dalam pohon keputusan.

2.6 Variabel Input Model

Dataset yang digunakan untuk membangun model C4.5 terdiri dari 9 variabel input dan satu variabel target (kategori profit). Variabel input model disajikan pada Tabel 2.

Tabel 2 Variabel Input Model

No	Variabel Input	Tipe Data
1	Tanggal_Angka	Numerik (1-31)
2	Hari_Nama	Kategorikal (Senin-Minggu)
3	Bulan_Nama	Kategorikal (nama bulan)
4	Harga_Jual_QTY	Numerik (jumlah item terjual)
5	Subtotal	Numerik (total pendapatan)
6	HPP_Total	Numerik (total biaya produksi)

No	Variabel Input	Tipe Data
7	Kategori_Dominan	Kategorikal (kategori dominan)
8	Produk_Dominan	Kategorikal (produk dominan)
9	Jenis_Bayar_Dominan	Kategorikal (metode pembayaran)

2.7 Metode Evaluasi Model

Evaluasi model menggunakan confusion matrix dan empat metrik pengukuran kinerja. Rumus Accuracy:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100\% \quad (9)$$

$$Accuracy = \frac{36 + 34}{36 + 34 + 5 + 4} \times 100\% = 88,61\%$$

Rumus Precision:

$$Precision = \frac{TP}{(TP + FP)} \times 100\% \quad (10)$$

$$Precision = \frac{36}{36 + 5} \times 100\% = 87,80\%$$

Rumus Recall:

$$Recall = \frac{TP}{(TP + FN)} \times 100\% \quad (11)$$

$$Recall = \frac{36}{36 + 4} \times 100\% = 90,00\%$$

Rumus F1-Score:

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (12)$$

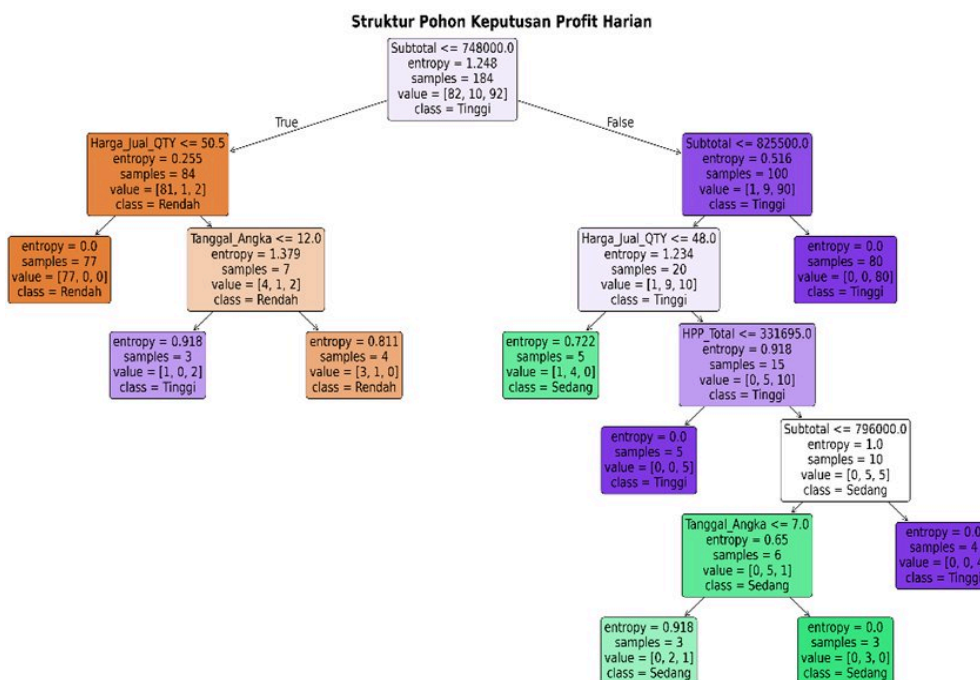
$$F1 - Score = 2 \times \frac{0,878 \times 0,90}{0,878 + 0,90} \\ = 0,889 = 88,9\%$$

Berdasarkan hasil evaluasi menggunakan confusion matrix, model klasifikasi yang dibangun dengan algoritma C4.5 menghasilkan tingkat akurasi sebesar 88,61% . Nilai presisi sebesar 87,80% menunjukkan bahwa sebagian besar data yang diprediksi sebagai kategori profit tinggi merupakan prediksi yang tepat. Sementara itu, nilai recall sebesar 90,00% menunjukkan bahwa model mampu mengidentifikasi sebagian besar data yang benar-benar termasuk dalam kategori profit tinggi. Nilai F1-Score sebesar 88,9% menunjukkan bahwa model memiliki keseimbangan yang baik antara presisi dan recall dalam proses klasifikasi.

3. HASIL DAN PEMBAHASAN

3.1 Struktur Pohon Keputusan C4.5

Hasil pelatihan model menghasilkan pohon keputusan dengan kedalaman maksimum 6 level, 17 node total, dan 9 leaf node. Atribut Subtotal terpilih sebagai root node karena memiliki nilai Gain Ratio tertinggi (0.8142). Struktur pohon keputusan ditampilkan pada Gambar 3 berikut.

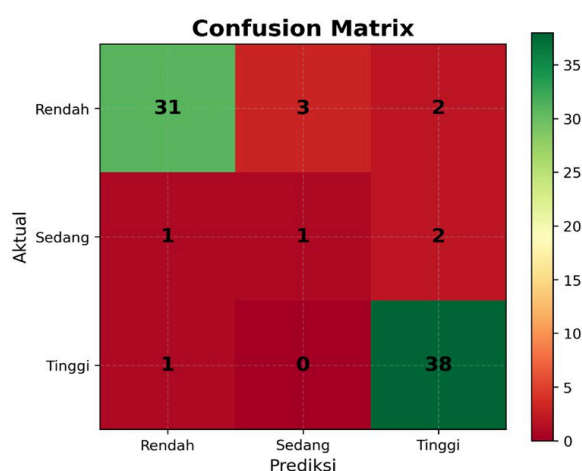


Gambar 3 Hasil Penelitian Struktur Pohon Keputusan

Berdasarkan pohon keputusan tersebut, terlihat bagaimana data profit harian dipisahkan secara bertahap berdasarkan atribut yang paling berpengaruh. Atribut Subtotal menjadi pemisah utama, kemudian dilanjutkan dengan Harga_Jual_QTY, HPP_Total, dan Tanggal_Angka sebagai atribut pendukung dalam sub-tree.

3.2 Evaluasi Model dengan Confusion Matrix

Evaluasi model dilakukan menggunakan 79 data testing. Perbandingan hasil prediksi model dengan nilai aktual setiap kategori profit ditampilkan pada Gambar 4.



Gambar 4 Confusion Matrix Hasil Evaluasi Model

Berdasarkan Berdasarkan confusion matrix, untuk kategori profit rendah, dari 36 hari data aktual, model memprediksi dengan benar 31 hari (3 hari diprediksi sebagai sedang, 2 hari sebagai tinggi). Kategori profit sedang dari 4 hari aktual, hanya 1 hari diprediksi benar. Kategori profit tinggi dari 39 hari aktual, 38 hari diprediksi benar.

Perhitungan akurasi keseluruhan berdasarkan persamaan (9):

$$Accuracy = \frac{(31 + 1 + 38)}{79} = \frac{70}{79} = 88.61\%$$

Ringkasan hasil evaluasi per kategori profit disajikan pada Tabel 3.

Tabel 3 Hasil Evaluasi Model per Kategori Profit

Kategori	Precision	Recall	F1-Score	Support
Rendah	93.94%	86.11%	89.86%	36
Sedang	25.00%	25.00%	25.00%	4
Tinggi	90.48%	97.44%	93.83%	39
Weighted Avg	89%	89%	89%	79

Nilai precision yang tinggi pada kategori profit rendah (93,94%) dan profit tinggi (90,48%) menunjukkan bahwa sebagian besar prediksi model sesuai dengan kondisi sebenarnya. Nilai recall kategori profit tinggi sebesar 97,44% menandakan hampir seluruh data berhasil dikenali dengan baik. Sebaliknya, performa pada kategori profit sedang masih rendah dengan precision dan recall sebesar 25,00%. Hal ini disebabkan oleh jumlah data kategori sedang yang relatif sedikit, yaitu hanya 14 hari (5,3%) dari total 263 hari, sehingga pada data testing hanya terdapat 4 hari. Keterbatasan data ini membuat model kurang optimal dalam mempelajari pola kategori tersebut. Selain itu, posisi kategori sedang yang berada di antara profit rendah dan tinggi menyebabkan berada pada area transisi, sehingga lebih sulit dibedakan secara konsisten.

3.3 Featur Importance dan Kontribusi Variabel

Analisis feature importance menunjukkan seberapa besar kontribusi setiap variabel dalam proses klasifikasi. Hasil analisis disajikan pada Tabel 4.

Tabel 4 Kontribusi Variabel dalam Model C4.5

Variabel	Skor Kepentingan	Persentase
Subtotal	0.8727	87.27%
Harga_Jual_QTY	0.0878	8.78%
Tanggal_Angka	0.0221	2.21%
HPP_Total	0.0174	1.74%
Variabel lainnya	0.0000	0.00%

Variabel Subtotal mendominasi proses klasifikasi dengan kontribusi 87,27%, jauh melampaui variabel lainnya. Harga_Jual_QTY berada di posisi kedua dengan 8,78%, diikuti Tanggal_Angka (2,21%) dan HPP_Total (1,74%). Variabel seperti Bulan_Nama, Hari_Nama, dan metode pembayaran tidak berkontribusi dalam model ini.

3.4 Aturan Keputusan (Decision Tree)

Dari pohon keputusan yang terbentuk, diturunkan aturan IF-THEN yang dapat digunakan sebagai panduan operasional bisnis:

Rule 1 (Profit Rendah): IF Subtotal $\leq$ Rp 748.000 AND Harga_Jual_QTY $\leq$ 50,5 THEN Profit = Rendah. Hari dengan nilai penjualan tidak lebih dari Rp 748.000 dan jumlah produk terjual sedikit cenderung menghasilkan profit rendah.

Rule 2 (Profit Tinggi-Volume): IF Subtotal $\leq$ Rp 748.000 AND Harga_Jual_QTY $>$ 50,5 AND Tanggal_Angka $\leq$ 12 THEN Profit = Tinggi. Meskipun subtotal rendah, volume tinggi pada awal bulan dapat menghasilkan profit tinggi.

Rule 3 (Profit Sedang): IF Subtotal > Rp 748.000 AND Subtotal ≤ Rp 825.500 AND Harga_Jual_QTY ≤ 48 THEN Profit = Sedang. Transaksi menengah dengan volume tidak terlalu banyak menghasilkan profit sedang.

Rule 4 (Profit Tinggi-Biaya Terkendali): IF Subtotal > Rp 748.000 AND Subtotal ≤ Rp 825.500 AND Harga_Jual_QTY > 48 AND HPP_Total ≤ Rp 331.695 THEN Profit = Tinggi. Kombinasi volume tinggi dan biaya produksi terkendali menghasilkan profit tinggi.

Rule 5 (Profit Tinggi-Penjualan Sangat Tinggi): IF Subtotal > Rp 825.500 THEN Profit = Tinggi. Ketika nilai transaksi harian sudah sangat tinggi, profit hampir selalu masuk kategori tinggi.

3.5 Rekomendasi Strategi Bisnis

Berdasarkan hasil analisis, dirumuskan tiga rekomendasi strategi utama. Pertama, peningkatan nilai transaksi (Subtotal) merupakan prioritas utama mengingat kontribusi variabel ini sebesar 87,27%. Target minimal Rp 600.000 per transaksi dapat dicapai melalui upselling produk premium, bundling kopi-snack, dan penawaran menu dengan margin tinggi.

Kedua, peningkatan volume pembelian item ($\text{Harga_Jual_QTY} \geq 30$ item per hari) dapat dicapai melalui promo Buy 2 Get 1, paket group/meeting, dan penawaran khusus pembelian dalam jumlah besar. Ketiga, efisiensi biaya produksi perlu dijaga pada batas aman $\text{HPP_Total/Subtotal} \leq 50\%$ melalui negosiasi harga bahan baku, pengurangan waste, dan evaluasi menu dengan margin profit rendah.

3.6 Pembahasan

Penelitian ini menghasilkan akurasi sebesar 88,61% menggunakan algoritma C4.5 dalam klasifikasi profit harian. Proses klasifikasi menggunakan variabel input seperti Subtotal, HPP Total, Quantity, serta atribut waktu yang telah melalui tahap *data processing*. Model dibangun melalui perhitungan *entropy*, *information gain*, dan *gain ratio* untuk menentukan atribut terbaik. Hasilnya menunjukkan bahwa Subtotal menjadi faktor utama, diikuti oleh Harga_Jual_QTY dan HPP_Total, dengan kemampuan klasifikasi yang baik pada kategori profit rendah, sedang, dan tinggi.

Dibandingkan penelitian Falentina et al. [8] yang memperoleh akurasi 84%, perbandingan ini tidak dapat dijadikan acuan langsung karena perbedaan karakteristik data. Evaluasi yang lebih tepat seharusnya menggunakan dataset yang sama dengan algoritma berbeda.

Kebaruan penelitian ini terletak pada fokus analisis profit harian pada UMKM kedai kopi, penggunaan *entropy-based discretization* untuk penentuan kategori secara objektif, serta pemanfaatan data transaksi nyata yang telah diintegrasikan menjadi data harian.

Dibandingkan penelitian Muktar et al. [9], penelitian ini lebih menekankan variabel internal transaksi seperti Subtotal, HPP Total, dan Quantity yang berhubungan langsung dengan profit, sehingga mampu menghasilkan pola yang mendukung pengambilan keputusan bisnis berbasis data.

4. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan sebagai berikut:

- Proses pemrosesan data struk penjualan berhasil dilakukan melalui tahapan data cleaning, transformasi, dan agregasi transaksi menjadi data harian. Proses ini menghasilkan 263 data harian dari 9.280 transaksi, dengan pola transaksi pelanggan dan distribusi kategori profit yang dapat dianalisis secara terstruktur.
- Penerapan algoritma C4.5 berhasil membangun model klasifikasi profit dengan akurasi 88,61%. Model menunjukkan bahwa variabel Subtotal merupakan faktor paling berpengaruh (87,27%), diikuti Harga_Jual_QTY (8,78%) dan HPP_Total (1,74%). Aturan keputusan IF-

THEN yang dihasilkan dapat digunakan sebagai dasar pengambilan keputusan bisnis yang terarah dan berbasis data.

5. SARAN

Beberapa saran untuk penelitian lebih lanjut:

- a. Proses pengolahan data penjualan sebaiknya dilakukan secara berkala menggunakan skrip Python yang telah dikembangkan agar pembuatan laporan profit harian dapat dilakukan lebih cepat dan efisien.
- b. Penelitian selanjutnya dapat menggunakan metode perbandingan algoritma klasifikasi lain seperti Random Forest atau Naive Bayes untuk mengevaluasi performa yang lebih komprehensif.
- c. Penambahan data dari periode yang lebih panjang dapat meningkatkan representativitas model, khususnya untuk kategori profit sedang yang masih memiliki jumlah data terbatas.
- d. Model dapat dikembangkan menjadi sistem pendukung keputusan berbasis web agar pemilik usaha dapat menggunakan hasil analisis secara real-time.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Dipa Makassar atas dukungan akademik selama penelitian, kepada pemilik Toko Kopi Hope yang telah menyediakan data transaksi penjualan, serta kepada seluruh pihak yang telah membantu kelancaran penelitian ini.

DAFTAR PUSTAKA

- [1] W. Aulia and Z. Zainarti, "Pengaruh Aktualisasi Diri Dan Gaya Hidup Hangout Terhadap Keputusan Pembelian Di Kalamera Coffee Space Pada Generasi Milenial Kota Medan," *Jurnal Pendidikan, Sains Sosial, Dan Agama*, vol. 8, no. 2, pp. 719–734, 2023.
- [2] N. Nursiti and S. Syafarudin, "Harga Dan Promosi Berdampak Terhadap Keputusan Pembelian Kopi Kenangan," *Jurnal Industri Kreatif Dan Kewirausahaan*, vol. 6, no. 1, pp. 75–84, 2023.
- [3] I. Supriatna and F. R. Ridha, "Penjualan Dan Kasir Digital Berbasis Android Loyverse POS Dalam Penyusunan Laporan Penjualan," *Indonesian Accounting Literacy Journal*, vol. 3, no. 3, pp. 309–316, 2023.
- [4] F. B. Nauw, "Pemanfaatan Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori," *Jurnal Sistem Cerdas Dan Rekayasa (JSCR)*, vol. 5, no. 1, 2023.
- [5] I. G. I. Sudipa, I. G. M. Darmawiguna, I. M. Dendi, et al., *Buku Ajar Data Mining*. PT, April 2024.
- [6] Tukino, S. Ani Arnomo, and A. Rahman Hakim, "Implementasi Algoritma C4.5 Dalam Penentuan Pola," *CBIS Journal*, vol. 10, pp. 13–21, 2022.
- [7] Y. Asri, D. Kuswardani, W. N. Suliyanti, and C. M. Tambunan, *Algoritma C4.5 : Klasifikasi Titik Dan Jenis Gangguan Pada Jaringan Distribusi Penyulang*. Uwais Inspirasi Indonesia, 2023.

- [8] F. G. Falentina, G. Yohan, Y. Wabdaron, D. Andiyani, and M. Sendoni, "Analysis Of Data Mining Implementation For Classifying Best-Selling Food Sale Using The Decision Tree (C4.5) Algorithm," vol. 4, no. 2, pp. 382–395, 2026.
 - [9] M. R. Muktar, M. Rifqi, F. Syam, A. Kunda, and M. S. Natsir, "Analisis Penerapan Data Mining Dalam Klasifikasi Penjualan Pakaian Pada Toko Online Shopee Menggunakan Algoritma C4.5," vol. 1, 2023.
 - [10] R. Sulastri, "Identifikasi Tingkat Penjualan Produk Herbal HWI Menggunakan Algoritma C4.5," *Jurnal Informatika Ekonomi Bisnis*, vol. 4, no. 3, pp. 84–89, 2022.
 - [11] S. Hamzah, Y. J. Zega, B. O. Lubis, et al., "Penerapan Data Mining Untuk Klasifikasi Data Penjualan Pakan Ternak Terlaris dengan Algoritma C4.5," vol. 10, no. 1, pp. 168–182, 2024.
 - [12] N. Azwanti, "Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada PT. Capella Dinamik Nusantara Cabang Muka Kuning," vol. 13, no. 1, 2018.
 - [13] G. Maulani, C. S. H. Rito, and D. K. Tania, "Penerapan Data Mining Di Berbagai Bidang," 2023.
 - [14] S. Yugi Wargiyo, Susliansyah, H. Sumarno, H. Priyono, and L. Maulida, "Rancangan Aplikasi Algoritma C4.5 Pada Stunting Balita," *REMIK: Riset Dan E-Jurnal Manajemen Informatika Komputer*, vol. 9, no. 1, pp. 254–265, 2025.
-