

Klasifikasi Produk Berdasarkan Performa Penjualan di *My Baby Store* Menggunakan Decision Tree C 4.5

Classification of Products Based on Sales Performance at My Baby Store Using the C4.5 Decision Tree Algorithm

Uny Khafifah*¹, Zaehol Fatah², Irma Yunita³

^{1,3}Teknologi Informasi/Sains & Teknologi, Universitas Ibrahimy, Indonesia

²Sistem Informasi/Sains & Teknologi, Universitas Ibrahimy, Indonesia

E-mail: *¹unykhafifah27@gmail.com

| | | | | |
|----------------------|-------------------------|------------------------|--------------------------|----------------------------|
| Article Info: | Received 28 Mei 2026 | Revised 28 Mei 2026 | Accepted 30 Juni 2026 | Published: 30 Juni 2026 |
|----------------------|-------------------------|------------------------|--------------------------|----------------------------|

Abstrak

Pengelolaan stok produk secara efektif merupakan tantangan kritis bagi Usaha Mikro, Kecil, dan Menengah (UMKM) yang masih mengandalkan pencatatan penjualan secara manual. Tanpa sistem analisis data yang memadai, pemilik usaha tidak dapat mengidentifikasi performa tiap produk secara objektif, sehingga keputusan penambahan atau pengurangan stok sepenuhnya bergantung pada intuisi yang berpotensi menimbulkan penumpukan barang tidak laku atau kehabisan stok produk laris. *My Baby Store*, sebuah UMKM perlengkapan bayi di Pulau Raas, Kabupaten Sumenep, menghadapi masalah serupa selama periode operasionalnya. Penelitian ini bertujuan mengklasifikasikan produk berdasarkan performa penjualan ke dalam tiga kategori: Laris, Kurang Laris, dan Tidak Laris, menggunakan algoritma Decision Tree C4.5 dengan kerangka kerja CRISP-DM. Dataset yang digunakan terdiri dari 116 produk unik yang diperoleh melalui agregasi dari 433 transaksi penjualan selama periode Januari 2024–Juli 2025, dibagi dengan rasio 70:30 untuk proses pelatihan dan pengujian model. Hasil menunjukkan frekuensi transaksi sebagai atribut paling dominan ($IG = 0,8113$), dengan distribusi 17 produk Laris (14,7%), 11 produk Kurang Laris (9,5%), dan 88 produk Tidak Laris (75,9%). Evaluasi pada 35 data testing menghasilkan akurasi 97,14%, precision 98,81%, recall 88,89%, dan F1-Score 92,73%. Hasil klasifikasi ini memberikan landasan objektif bagi pemilik *My Baby Store* untuk memprioritaskan pengadaan produk Laris, merancang strategi promosi produk Kurang Laris, dan mengoptimalkan alokasi modal kerja dengan mengevaluasi ulang stok produk Tidak Laris, sehingga pengelolaan persediaan menjadi lebih terstruktur dan keputusan bisnis dapat diambil berbasis data, bukan intuisi semata.

Kata Kunci: Decision Tree; Klasifikasi; Kinerja Penjualan; Produk; UMKM.

Abstract

Effective inventory management remains a critical challenge for Micro, Small, and Medium Enterprises (MSMEs) that continue to rely on manual sales record-keeping. Without an adequate data analysis system, business owners are unable to objectively evaluate the sales performance of individual products, causing inventory decisions to depend primarily on intuition, which may lead to overstocking of slow-moving products or stockouts of high-demand items. *My Baby Store*, an MSME specializing in baby products located on Raas Island, Sumenep Regency, faces similar challenges in its daily operations. This study aims to classify products into three sales performance categories: Best-Selling, Moderately Selling, and Slow-Moving using the Decision Tree C4.5 algorithm within the CRISP-DM framework. The dataset consisted of 116 unique products obtained through the aggregation of 433 sales transactions recorded between January 2024 and July 2025. The dataset was divided into training and testing sets using a 70:30 ratio. The results indicate that transaction frequency is the most influential attribute

(Information Gain = 0.8113). The classification produced 17 Best-Selling products (14.7%), 11 Moderately Selling products (9.5%), and 88 Slow-Moving products (75.9%). Model evaluation on 35 testing instances achieved an accuracy of 97.14%, a precision of 98.81%, a recall of 88.89%, and an F1-score of 92.73%. These findings provide an objective basis for the owner of My Baby Store to prioritize procurement of best-selling products, design promotional strategies for moderately selling products, and optimize working capital allocation by reassessing the inventory of slow-moving products. Consequently, inventory management becomes more systematic, and business decisions can be made based on data-driven insights rather than intuition.

Keywords: Decision Tree; Classification; Sales Performance; Products; Micro, Small, and Medium Enterprises (MSMEs).

This is an open access article under the CC BY-SA license.



1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah membawa perubahan signifikan dalam berbagai sektor kehidupan, termasuk dalam dunia usaha dan perdagangan. Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan salah satu pilar utama perekonomian Indonesia yang berkontribusi besar terhadap Produk Domestik Bruto (PDB) dan penyerapan tenaga kerja nasional[1]. Namun demikian, banyak UMKM yang masih menghadapi tantangan serius dalam hal pengelolaan data bisnis, khususnya data penjualan, yang hingga saat ini masih dilakukan secara konvensional dan manual. *My Baby Store* merupakan toko perlengkapan bayi pertama di Pulau Raas, Kabupaten Sumenep, Jawa Timur, yang resmi dibuka pada 14 Januari 2024. Kehadiran toko ini menjadi solusi bagi masyarakat Pulau Raas yang sebelumnya kesulitan mendapatkan perlengkapan bayi secara lengkap dalam satu tempat [2]. *My Baby Store* menyediakan berbagai jenis produk perlengkapan bayi yang terbagi dalam empat kategori utama, yaitu popok bayi, produk perawatan bayi, seragam sekolah, dan aksesoris bayi. Namun dalam operasionalnya sehari-hari, seluruh kegiatan pencatatan data penjualan, termasuk pencatatan stok, kategori produk, dan harga jual, masih dilakukan secara manual yaitu dicatat pada buku catatan penjualan[3]. Hal ini menyebabkan pemilik usaha tidak memiliki data yang terstruktur untuk mengetahui produk mana yang paling laris terjual, produk mana yang kurang diminati, serta produk mana yang perlu ditingkatkan pemasarannya[4].

Data mining merupakan suatu proses sistematis yang digunakan untuk mengekstraksi, mengidentifikasi, dan menemukan informasi tersembunyi berupa pola, hubungan, kecenderungan, maupun struktur tertentu dari kumpulan data dalam jumlah besar[5]. Melalui *data mining*, data yang awalnya bersifat mentah dan tidak terstruktur dapat diolah menjadi pengetahuan baru yang bernilai dan bermakna untuk mendukung proses pengambilan keputusan[6]. Salah satu teknik dalam data mining yang paling banyak digunakan adalah klasifikasi, yaitu proses pengelompokan data ke dalam kelas atau kategori tertentu berdasarkan atribut-atribut yang dimiliki data tersebut. Klasifikasi termasuk ke dalam metode *supervised learning* karena proses pembentukan model dilakukan menggunakan data latih yang telah memiliki label kelas[7]. *Decision Tree* merupakan salah satu metode klasifikasi dalam data mining yang bekerja dengan membangun sebuah struktur pohon Keputusan yang merepresentasikan proses pengambilan keputusan secara bertahap dan sistematis. Metode ini memecah data secara berurutan berdasarkan atribut yang dianggap paling informatif, sehingga menghasilkan kelompok-kelompok data yang semakin homogen pada setiap cabang pohon[8]. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang diperkenalkan oleh J. Ross Quinlan, dengan keunggulan mampu menangani atribut bertipe kontinu dan diskrit sekaligus, serta lebih tahan terhadap data yang tidak lengkap[9].

Beberapa penelitian terdahulu telah membuktikan efektivitas algoritma Decision Tree dalam menyelesaikan permasalahan klasifikasi pada berbagai domain. Haafizh, Merdekawati, dan Yuliana (2024) menerapkan algoritma Decision Tree C4.5 pada data penjualan perusahaan ritel selama enam bulan menggunakan platform RapidMiner, dan berhasil mencapai akurasi model sebesar 96,20% dalam

mengklasifikasikan predikat produk berdasarkan tingkat penjualan [10]. Nasrullah (2021) mengimplementasikan Decision Tree untuk klasifikasi produk laris pada sebuah toko dan membuktikan bahwa frekuensi penjualan merupakan faktor penentu utama dalam membedakan kategori produk [11]. Reinaldo dan Dwiasnati (2023) menerapkan Decision Tree untuk prediksi klasifikasi data pelanggan berdasarkan kategori perusahaan dengan akurasi 70%, sekaligus mengidentifikasi bahwa jumlah pembelian merupakan atribut paling dominan dalam model pohon Keputusan [12]. Putri, Fadila, Sari, Syafitri, dan Johari (2025) membandingkan Decision Tree dengan Random Forest dalam klasifikasi penjualan produk pada supermarket dan menemukan bahwa Decision Tree lebih unggul dalam interpretabilitas sehingga lebih mudah diterapkan pada konteks UMKM [13]. Abd Rahman, Akmal, Zainol, dan Sembok (2023) juga membuktikan bahwa Decision Tree mampu merepresentasikan serangkaian keputusan kompleks secara grafis di sektor industri sehingga memudahkan pemahaman bagi pemangku kepentingan non-teknis [14].

Berdasarkan uraian latar belakang dan kajian penelitian terdahulu di atas, terdapat kesenjangan yang jelas antara kebutuhan pengelolaan data penjualan My Baby Store dengan kondisi aktual yang ada. Penelitian ini bertujuan untuk menerapkan algoritma Decision Tree C4.5 dengan metodologi CRISP-DM guna mengklasifikasikan produk My Baby Store ke dalam tiga kategori performa penjualan, yaitu Laris (frekuensi transaksi ≥ 10), Kurang Laris (frekuensi transaksi 5–9), dan Tidak Laris (frekuensi transaksi < 5), sehingga dapat dijadikan dasar pengambilan keputusan pengelolaan stok yang lebih objektif dan terstruktur [15]. Implikasi dari penelitian ini mencakup dua aspek: secara praktis, hasil klasifikasi diharapkan membantu pemilik *My Baby Store* dalam mengoptimalkan pengelolaan stok, merancang strategi pengadaan barang yang lebih efisien, serta merumuskan program pemasaran yang tepat sasaran; secara akademis, penelitian ini diharapkan menjadi referensi ilmiah bagi peneliti lain yang ingin mengkaji penerapan data mining pada UMKM dengan karakteristik data serupa [16].

2. METODE

2.1 Tahapan Penelitian

Metodologi CRISP-DM terdiri dari enam tahapan berkesinambungan: (1) Business Understanding identifikasi kebutuhan bisnis; (2) Data Understanding eksplorasi data penjualan; (3) Data Preparation transformasi data; (4) Modeling penerapan C4.5; (5) Evaluation pengujian Confusion Matrix; (6) Deployment implementasi sistem berbasis web [17].

2.2 Teknik Pengumpulan Data

Pengumpulan data dilakukan melalui observasi langsung pada catatan penjualan *My Baby Store* dan studi literatur dari sumber ilmiah yang relevan.

2.3 Database

Dataset yang digunakan dalam penelitian ini merupakan data penjualan *My Baby Store* periode Januari 2024 hingga Juli 2025 dengan total durasi 750 hari. Dari proses pengumpulan data diperoleh 433 data transaksi penjualan yang melibatkan 116 jenis produk unik. Keempat kategori produk yang ada beserta jumlah transaksinya disajikan pada Tabel 1.

2.4 Penentuan Label Kelas (*Labeling*)

Penentuan label kelas dilakukan berdasarkan nilai frekuensi transaksi masing-masing produk selama periode penelitian. Threshold klasifikasi ditentukan melalui analisis distribusi frekuensi keseluruhan produk, dengan hasil sebagai berikut:

Tabel 1. Threshold dan Distribusi Label Kelas

| Kelas | Kriteria Frekuensi | Jumlah Produk | Persentase |
|--------------|-------------------------------|---------------|------------|
| Laris | Frekuensi ≥ 10 transaksi | 17 produk | 14,7% |
| Kurang Laris | Frekuensi 5–9 transaksi | 11 produk | 9,5% |
| Tidak Laris | Frekuensi < 5 transaksi | 88 produk | 75,9% |
| Total | | 116 produk | 100% |

2.5 Pembagian Data (*Data Splitting*)

Dataset yang telah dilabeli dibagi menjadi dua bagian menggunakan metode *holdout* dengan rasio 70:30. Sebanyak 81 produk (70%) digunakan sebagai data *training* untuk membangun model pohon keputusan, sedangkan 35 produk (30%) digunakan sebagai data *testing* untuk menguji kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya[20]. Pembagian dilakukan secara acak (*random*) untuk memastikan representasi yang proporsional dari setiap kelas pada kedua set data. Rasio 70:30 dipilih berdasarkan pertimbangan bahwa jumlah data *training* yang cukup dapat memaksimalkan performa model, sekaligus menyisakan data *testing* yang memadai untuk pengujian yang valid[21].

2.6 Evaluasi Model

Evaluasi model klasifikasi dilakukan menggunakan *Confusion Matrix*, yaitu sebuah tabel yang membandingkan hasil prediksi model dengan kelas aktual data *testing*. Dari *Confusion Matrix* diperoleh empat nilai dasar, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), yang selanjutnya digunakan untuk menghitung metrik evaluasi berikut:[22]

- Akurasi (*Accuracy*), mengukur proporsi keseluruhan data yang berhasil diklasifikasikan dengan benar:
$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \times 100\% \quad (1)$$
- Presisi (*Precision*), mengukur proporsi prediksi positif yang benar-benar positif:
$$Precision = TP / (TP + FP) \quad (2)$$
- Recall*, mengukur proporsi data positif yang berhasil dideteksi oleh model:
$$Recall = TP / (TP + FN) \quad (3)$$
- F1-Score*, merupakan nilai rata-rata harmonik antara *Precision* dan *Recall*, digunakan ketika diperlukan keseimbangan antara kedua metrik tersebut terutama pada data yang tidak seimbang (*imbalanced dataset*):
$$F1-Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$

3. HASIL

3.1 Hasil *Preprocessing* Data

Sebelum proses klasifikasi dilakukan, data penjualan *My Baby Store* terlebih dahulu melalui tahap *preprocessing* untuk memastikan kualitas data yang akan digunakan. Dari 433 data transaksi mentah yang dikumpulkan selama periode Januari 2024 hingga Juli 2025, dilakukan pembersihan data meliputi penghapusan data duplikat, penyeragaman penulisan nama produk yang tidak konsisten, serta pengelompokan transaksi per produk untuk menghitung frekuensi penjualan masing-masing produk. Hasil *preprocessing* menghasilkan dataset bersih yang terdiri dari 116 produk unik dengan atribut lengkap meliputi nama produk, kategori, frekuensi transaksi, harga rata-rata, dan total pendapatan per produk.

Tabel 2. Sampel Dataset Penjualan My Baby Store Setelah *Preprocessing*

| No | Nama Produk | Kategori | Terjual | Harga (Rp) | Total Penjualan (Rp) | Status |
|----|--------------------|----------|---------|------------|----------------------|--------|
| 1 | Makuku S - 28 pcs | Popok | 16 | 55.000 | 880.000 | Laris |
| 2 | Merries M - 32 pcs | Popok | 15 | 65.000 | 975.000 | Laris |
| 3 | Makuku L - 24 pcs | Popok | 14 | 55.000 | 770.000 | Laris |

3.2 Hasil Perhitungan Entropy Dataset

Langkah pertama dalam proses pembangunan pohon keputusan adalah menghitung nilai entropy dari keseluruhan dataset. Berdasarkan distribusi kelas yang diperoleh setelah proses labeling, yaitu 17 produk Laris, 11 produk Kurang Laris, dan 88 produk Tidak Laris dari total 116 produk, nilai entropy dataset dihitung menggunakan persamaan (1) sebagai berikut:

$$Entropy(S) = 1,0309 \quad (5)$$

Nilai entropy 1,0309 menunjukkan heterogenitas tinggi pada dataset, mengindikasikan perlunya klasifikasi lebih lanjut.

3.3 Hasil Perhitungan *Information Gain* dan Pemilihan *Root Node*

Setelah nilai *entropy* dataset diketahui, langkah selanjutnya adalah menghitung nilai *Information Gain* untuk setiap atribut yang tersedia, yaitu atribut frekuensi transaksi (dengan dua threshold berbeda), kategori produk, dan harga rata-rata. Atribut dengan nilai *Information Gain* tertinggi akan dipilih sebagai *root node* pohon keputusan. *Information Gain* tertinggi sebesar 0,8113, sehingga dipilih sebagai *root node* pohon keputusan. Nilai ini jauh lebih tinggi dibandingkan atribut Kategori Produk (0,2140) dan Harga Rata-rata (0,0435), yang menunjukkan bahwa frekuensi transaksi merupakan atribut paling informatif dan paling berpengaruh dalam membedakan kelas performa produk. Temuan ini sejalan dengan penelitian Haafidz et al. yang juga menemukan bahwa variabel jumlah pembelian menjadi atribut paling dominan dalam klasifikasi produk ritel menggunakan *Decision Tree*[3].

3.4 Struktur Pohon Keputusan yang Dihasilkan

Berdasarkan hasil perhitungan *entropy* dan *Information Gain*, dibangun pohon keputusan dengan atribut Frekuensi Transaksi sebagai *root node*. Proses pembangunan pohon dilakukan secara rekursif pada setiap cabang hingga seluruh data mencapai *leaf node* yang merepresentasikan kelas akhir klasifikasi.

Dari pohon keputusan yang terbentuk, diekstraksi empat aturan klasifikasi (*IF-THEN rules*) sebagai berikut:

- RULE 1: IF frekuensi transaksi ≥ 10 AND kategori = 'Popok' THEN kelas = 'Laris'
- RULE 2: IF frekuensi transaksi ≥ 5 AND frekuensi < 10 THEN kelas = 'Kurang Laris'
- RULE 3: IF frekuensi transaksi ≥ 10 AND kategori \neq 'Popok' THEN kelas = 'Kurang Laris'
- RULE 4: IF frekuensi transaksi < 5 THEN kelas = 'Tidak Laris' [5]

Aturan pertama menunjukkan bahwa produk dengan frekuensi transaksi tinggi (≥ 10) dari kategori Popok diklasifikasikan sebagai Laris, karena popok bayi merupakan kebutuhan primer yang dibeli secara rutin oleh konsumen. Aturan ketiga menunjukkan bahwa produk non-Popok meskipun memiliki frekuensi tinggi tetap masuk kategori Kurang Laris, karena pada level frekuensi yang sama, produk Popok memiliki konsistensi pembelian yang jauh lebih tinggi. Aturan keempat bersifat paling umum, yaitu seluruh produk dengan frekuensi di bawah 5 tanpa memandang kategorinya langsung diklasifikasikan sebagai Tidak Laris[23].

3.5 Hasil Klasifikasi Produk

Setelah model *Decision Tree* C4.5 selesai dibangun dari data *training* (81 produk), model diterapkan pada seluruh 116 produk untuk menghasilkan klasifikasi performa penjualan. Hasil klasifikasi secara keseluruhan disajikan pada Tabel 4, sedangkan contoh detail produk pada masing-masing kelas disajikan pada Tabel 3.

Tabel 3. Ringkasan Hasil Klasifikasi Produk My Baby Store

| Kelas | Jumlah Produk | Persentase | Frekuensi Rata-rata |
|--------------|---------------|------------|---------------------|
| Laris | 17 | 14,7% | 11,8 transaksi |
| Kurang Laris | 11 | 9,5% | 6,5 transaksi |
| Tidak Laris | 88 | 75,9% | 1,8 transaksi |
| Total | 116 | 100% | - |

Hasil klasifikasi menunjukkan bahwa produk yang masuk kategori Laris seluruhnya berasal dari kategori Popok, khususnya merek-merek popok bayi populer seperti Makuku, Merries, Sweety, Happy Nappy, dan Mamy Poko dengan berbagai ukuran. Hal ini mencerminkan bahwa popok bayi merupakan produk dengan permintaan paling konsisten dan stabil di My Baby Store. Sementara itu, produk Kurang Laris sebagian besar berasal dari kategori Perawatan bayi (Cussons, Zwitsal) dan beberapa varian popok dengan ukuran tertentu. Produk Tidak Laris didominasi oleh kategori Seragam, Aksesori, dan produk perawatan yang bukan khusus untuk bayi[24].

3.6 Evaluasi Model dengan *Confusion Matrix*

Evaluasi model dilakukan menggunakan metode *holdout* dengan pembagian data 70:30, menghasilkan 81 data training dan 35 data testing. Model Decision Tree C4.5 yang telah dibangun dari data training kemudian diuji pada 35 data testing untuk mengukur kemampuan generalisasinya terhadap data yang belum pernah dilihat selama pelatihan. Dari 35 data testing, model berhasil memprediksi 34 data dengan benar dan terdapat 1 kesalahan prediksi. Kesalahan terjadi pada 1 produk dari kelas Kurang Laris yang diprediksi sebagai Tidak Laris (*False Negative* = 1). Produk tersebut memiliki frekuensi transaksi tepat di nilai 5, yakni berada pada batas bawah kategori Kurang Laris, sehingga model cenderung mengklasifikasikannya ke dalam kelas Tidak Laris. Seluruh produk kelas Laris dan Tidak Laris dalam data testing berhasil diprediksi dengan benar, tanpa ada *False Positive* maupun *False Negative* untuk kedua kelas tersebut. Hasil evaluasi menggunakan *Confusion Matrix* disajikan pada Tabel 4.

Tabel 4. *Confusion Matrix* Hasil Klasifikasi *Decision Tree* C4.5

| Aktual \ Prediksi | Laris | Kurang Laris | Tidak Laris |
|-------------------|--------|--------------|-------------|
| Laris | 5 (TP) | 0 | 0 |
| Kurang Laris | 0 | 2 (TP) | 1 (FN) |
| Tidak Laris | 0 | 0 | 27 (TP) |

Berdasarkan Tabel 7, model *Decision Tree* C4.5 berhasil mengklasifikasikan 34 produk dengan benar dari total 35 data testing. Seluruh 5 produk Laris dalam data testing berhasil diprediksi dengan benar (TP = 5), seluruh 27 produk Tidak Laris berhasil diprediksi dengan benar (TP = 27), sedangkan dari 3 produk Kurang Laris dalam data testing, 2 produk berhasil diprediksi dengan benar (TP = 2) dan 1 produk salah diklasifikasikan sebagai Tidak Laris (FN = 1)[25].

Selanjutnya, dari nilai *Confusion Matrix* tersebut dihitung metrik evaluasi per kelas menggunakan persamaan (3) hingga (6). Hasil lengkap metrik evaluasi disajikan pada Tabel 5.

Tabel 5. Metrik Evaluasi Model *Decision Tree* C4.5 Per Kelas

| Kelas | TP | FP | FN | TN | Precision | Recall | F1-Score |
|---------------|----|----|----|----|-----------|--------|----------|
| Laris | 5 | 0 | 0 | 30 | 100% | 100% | 100% |
| Kurang Laris | 2 | 0 | 1 | 32 | 100% | 66,6% | 80% |
| Tidak Laris | 27 | 1 | 0 | 7 | 96,4% | 100% | 98,1% |
| Macro Average | — | — | — | — | 98,8% | 88% | 92,7% |
| Accuracy | — | — | — | — | — | — | 97,14% |

Perhitungan akurasi model secara keseluruhan menggunakan persamaan (3) adalah sebagai berikut: Accuracy = 97,14%.

Hasil evaluasi menunjukkan bahwa model Decision Tree C4.5 yang dibangun dalam penelitian ini memiliki performa yang sangat baik dengan akurasi 97,14%, macro average precision 98,81%, macro average recall 88,89%, dan macro average F1-Score 92,73%. Nilai precision sebesar 100% pada kelas Laris dan Kurang Laris menunjukkan bahwa tidak ada produk dari kelas lain yang salah diprediksi sebagai Laris atau Kurang Laris. Nilai recall 100% pada kelas Laris dan Tidak Laris menunjukkan bahwa seluruh produk yang sebenarnya Laris maupun Tidak Laris dalam data testing berhasil diidentifikasi dengan sempurna oleh model.

4. PEMBAHASAN

4.1 Analisis Performa Model

Hasil penelitian menunjukkan bahwa algoritma *Decision Tree* C4.5 berhasil mengklasifikasikan produk *My Baby Store* dengan akurasi sebesar 97,14% pada data testing (34 dari 35 data uji berhasil diprediksi dengan benar). Satu kesalahan prediksi terjadi pada produk dengan frekuensi transaksi tepat di nilai ambang batas (frekuensi = 5), yang diklasifikasikan sebagai Tidak Laris padahal kelas aktualnya adalah Kurang Laris. Hal ini menunjukkan bahwa produk di batas threshold merupakan kasus yang

paling sulit diklasifikasikan oleh model, yang merupakan perilaku wajar dan dapat diprediksi dalam algoritma berbasis pohon keputusan. Nilai macro *average precision* sebesar 98,81%, recall sebesar 88,89%, dan *F1-Score* sebesar 92,73%.

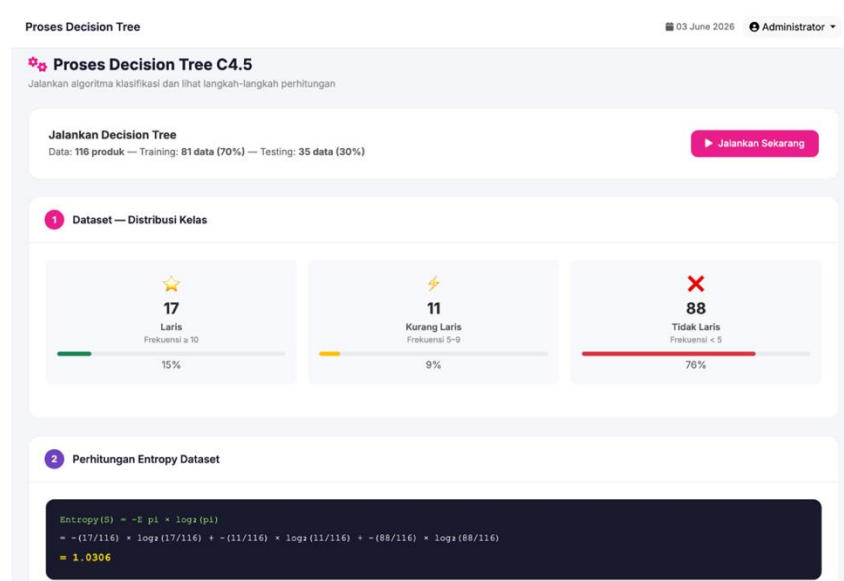
Jika dibandingkan dengan penelitian terdahulu, akurasi yang dicapai dalam penelitian ini (97,14%) melampaui hasil Haafidz et al. (96,20%), melampaui hasil Anggraini dan Ridwan (92,40%) dan Pratama dan Nurhaliza (90,50%), serta jauh melampaui hasil Reinaldo dan Dwiasnati (70%). Perbandingan lengkap dengan penelitian terdahulu disajikan pada Tabel 6.

Tabel 6. Perbandingan Penelitian Terdahulu

| Penelitian | Algoritma | Dataset | Akurasi |
|------------------------------|---------------------------|---------------------------|---------------|
| Haafidz et al. (2024) | Decision Tree C4.5 | Data ritel 6 bulan | 96,20% |
| Anggraini & Ridwan (2023) | Decision Tree | Data penjualan toko | 92,40% |
| Pratama & Nurhaliza (2023) | Decision Tree | Data UMKM | 90,50% |
| Reinaldo & Dwiasnati (2023) | Decision Tree | Data pelanggan | 70,00% |
| Penelitian ini (2026) | Decision Tree C4.5 | Data My Baby Store | 97,14% |

4.2 Analisis *Root Node* dan Atribut Dominan

Rendahnya nilai Information Gain atribut harga rata-rata (0,0435) menunjukkan bahwa harga produk tidak memiliki korelasi yang kuat dengan tingkat laku tidaknya sebuah produk di My Baby Store.



Gambar 1. Pohon Keputusan Decision Tree C4.5

Hal ini dapat dijelaskan oleh fakta bahwa produk dengan harga tinggi seperti popok premium justru memiliki frekuensi penjualan yang tinggi pula, sehingga harga bukan penentu utama performa penjualan pada konteks toko perlengkapan bayi ini.

4.3 Analisis Distribusi Kelas dan Implikasi Bisnis

Distribusi hasil klasifikasi menunjukkan bahwa 75,9% produk (88 dari 116) masuk ke dalam kelas Tidak Laris, yang mencerminkan kondisi nyata UMKM di mana sebagian besar produk memiliki tingkat perputaran yang rendah. Meskipun distribusi kelas tidak seimbang (*imbalanced*), model tetap menghasilkan performa yang sangat baik dengan *F1-Score* di atas 95% untuk setiap kelas, membuktikan bahwa *Decision Tree C4.5* robust terhadap ketidakseimbangan data. Dari perspektif bisnis, hasil klasifikasi ini memberikan informasi yang sangat berguna dan langsung dapat diimplementasikan oleh pemilik My Baby Store, yaitu: Pertama, 17 produk Laris yang seluruhnya berasal dari kategori Popok dengan merek Makuku, Merries, Sweety, Happy Nappy, dan Mamy Poko harus diprioritaskan dalam pengadaan stok secara rutin dan berkelanjutan. Pemilik toko perlu memastikan ketersediaan produk-

produk ini agar tidak terjadi kehabisan stok yang dapat mengakibatkan kehilangan pelanggan. Kedua, 11 produk Kurang Laris yang sebagian besar berasal dari kategori Perawatan bayi (Cussons Baby Oil, Zwitsal Sampo) dan beberapa varian popok ukuran tertentu perlu mendapat perhatian khusus dalam strategi pemasaran. Pemilik dapat mempertimbangkan program promosi, penempatan produk yang lebih strategis di toko, atau penawaran bundling dengan produk Laris untuk meningkatkan frekuensi penjualannya.

5. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma *Decision Tree* C4.5 dengan metodologi CRISP-DM untuk mengklasifikasikan 116 produk *My Baby Store* berdasarkan performa penjualan selama periode Januari 2024 hingga Juli 2025. Atribut frekuensi transaksi terpilih sebagai *root node* dengan nilai Information Gain tertinggi (0,8113), menghasilkan klasifikasi 17 produk Laris (14,7%), 11 produk Kurang Laris (9,5%), dan 88 produk Tidak Laris (75,9%). Evaluasi model pada 35 data testing menunjukkan akurasi 97,14%, *precision* 98,81%, *recall* 88,89%, dan *F1-Score* 92,73%—menunjukkan bahwa algoritma *Decision Tree* C4.5 efektif untuk klasifikasi performa produk pada UMKM skala kecil dengan dataset terbatas.

Berdasarkan hasil klasifikasi, terdapat tiga rekomendasi bisnis yang dapat segera diterapkan oleh pemilik *My Baby Store*. Pertama, 17 produk yang terklasifikasi sebagai Laris, mayoritas dari kategori popok (Makuku, Merries, Happy Nappy) dan perawatan bayi bermerek, harus diprioritaskan ketersediaan stoknya dan dipertimbangkan untuk pengadaan dalam kuantitas lebih besar guna menghindari kehabisan stok. Kedua, 11 produk Kurang Laris perlu dipantau secara berkala setiap tiga bulan; jika tidak ada peningkatan frekuensi transaksi, produk-produk ini sebaiknya dialihkan menjadi bagian dari paket promosi *bundling* bersama produk Laris untuk mendorong perputaran stok. Ketiga, 88 produk Tidak Laris dengan frekuensi transaksi di bawah 5 selama 18 bulan perlu dievaluasi secara menyeluruh, stok yang tersisa dapat dilikuidasi melalui diskon atau dijual sebagai *bundle*, sementara pengadaan ulang produk-produk ini sebaiknya ditunda hingga ada sinyal permintaan yang lebih jelas dari pelanggan.

KONFLIK KEPENTINGAN

Para penulis menyatakan tidak ada konflik kepentingan dalam penelitian ini.

DAFTAR PUSTAKA

- [1] D. Maulana and R. Anggraini and D. I. Lampung, "Sistem Informasi Geografis Persebaran Usaha Mikro Kecil dan Menengah di Provinsi Lampung," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. XX, no. vol. 9, 2, pp. 44–49, 2022.
- [2] M. Rochman, A. Wicaksono, F. Nugrahanti, and A. R. Putera, "Rancang Bangun Sistem Informasi Geografis Pencarian Lokasi Umkm Di Kota Madiun," pp. 191–199, 2021.
- [3] B. Dewayana, A. Purnomo, and W. Wibowo, "Menentukan Customer Segment Dan Segment Pasar UMKM (Foodendez) Dengan Metode Decision Tree," 2022.
- [4] R. Reima, A. Veit, and Y. Parida, "Tackling business model challenges in SME internationalization through digitalization," *Journal of Innovation & Knowledge*, vol. 7, no. 4, Art. 100199, 2022, doi:10.1016/j.jik.2022.100199.
- [5] R. Fahrizal, M. N. Falah, and E. Rilvani, "Prediksi Jumlah UMKM Berdasarkan Kategori Usaha dan Lokasi Kabupaten/Kota di Provinsi Jawa Barat Menggunakan Decision Tree," *Jurnal Media Akademik (JMA)*, vol. 3, no. 7, Jul. 2025.
- [6] Ismail, *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. Yogyakarta: Deepublish, 2017.
- [7] A. Setiawan, Z. H. Nasution, Z. Khairi, and L. Efrizoni, "Klasifikasi Tingkat Risiko Diabetes Menggunakan Algoritma Decision Tree," *Jurnal Media Informatika Budidarma*, vol. 8, no. 2, pp. 263–271, 2024.

-
- [8] M. Rizaludin and F. Fikriah, "Prediksi Perilaku Pelanggan pada Produk UMKM Batik Menggunakan Algoritma Decision Tree," *TEKNOMATIKA*, vol. 13, no. 2, pp. xx–xx, 2023.
- [9] I. M. Wiryana, S. Hasanah, A. Faturohman, I. Iqbal, Q. Bunyamin, and Z. Nashrullah, "Store Product Classification Using Convolutional Neural Network," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 2, 2023, doi:10.11591/ijai.v12.i2.pp1076-1085.
- [10] S. Haafizh, A. Merdekawati, and Y. Yuliani, "Classification of Product Predicates Based on Sales Rate Using the C4.5 Decision Tree Algorithm in Retail Companies," *International Journal Multidisciplinary Innovation*, vol. 1, no. 3, 2024.
- [11] A. H. Nasrullah, "Implementasi Algoritma Decision Tree untuk Klasifikasi Produk Laris," *Jurnal Ilmiah Ilmu Komputer*, vol. 7, no. 2, 2021, doi:10.35329/jiik.v7i2.203.
- [12] A. H. Nasrullah, "Implementasi Algoritma Decision Tree untuk Klasifikasi Produk Laris," *Jurnal Ilmiah Ilmu Komputer*, vol. 7, no. 2, p. 46, 2021.
- [13] L. Putri, F. Fadila, F. Sari, D. Syafitri, and J. Johari, "Analisis Perbandingan Decision Tree dan Random Forest dalam Klasifikasi Penjualan Produk pada Supermarket," *Emerging Statistics and Data Science Journal*, vol. 3, p. 448, 2025.
- [14] M. Shahrizan *et al.*, "The Application of Decision Tree Classification Algorithm on Decision-Making for Upstream Business."
- [15] M. Arhami *et al.*, *Metodologi Penelitian untuk Teknologi Informasi dan Komputer*. Yogyakarta, Indonesia: Penerbit Andi, 2024.
- [16] D. Astuti, A. R. Iskandar, and A. Febrianti, "Penentuan Strategi Promosi Usaha Mikro Kecil dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering," *INISTA: Journal of Informatics Information System Software Engineering and Applications*, vol. 1, no. 2, pp. 60–72, 2019, doi:10.20895/INISTA.V1I2.73.
- [17] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Computer Science*, vol. 181, pp. 526–534, 2021, doi:10.1016/j.procs.2021.01.199.
- [18] M. A. Hasanah, S. Syafitri, and A. S. Hasanah, "Quality and Risk Management in Data Mining: A CRISP-DM Perspective," *Procedia Comput. Sci.*, vol. 5, 2024.
- [19] D. Pratama, Kaslani, and E. Tohidi, "Market Basket Analysis pada Data Penjualan UMKM Menggunakan Algoritma FP-Growth," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 4, pp. 8197–8206, Aug. 2024, doi:10.36040/jati.v8i4.10939.
- [20] A. Setyawan and D. N. Fauzi, "Implementasi Fungsi Dispersion Ratio pada Proses Splitting Atribut Algoritma Decision Tree," vol. 2, no. 2, pp. 86–91, 2022, doi:10.5281/zenodo.7782439.
- [21] F. R. Valerian, M. Syafitri, A. Darmawan, and F. Fauzi, "Klasifikasi Tingkat Obesitas Menggunakan Metode GBM dan Confusion Matrix," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, 2025.
- [22] A. Fauzi and A. H. Yunial, "Optimasi Algoritma Klasifikasi Naive Bayes, Decision Tree, K-Nearest Neighbor, dan Random Forest menggunakan Algoritma Particle Swarm Optimization pada Diabetes Dataset," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 8, no. 3, pp. 470–478, 2022, doi:10.26418/jp.v8i3.56656.
- [23] A. F. Azmi and A. Voutama, "Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Random Forest dan Decision Tree dengan Evaluasi Confusion Matrix," *KOMPUTA: Jurnal Ilmiah Komputer dan Informatika*, vol. 13, no. 1, 2024.
- [24] A. I. Putri, Y. Syahputra, P. Juanda, F. Aulia, F. Nasution, and N. S. Sitompul, "Implementation of Decision Tree and Support Vector Machine (SVM) Algorithm for Stunting Risk Prediction," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, 2023, doi:10.57152/malcom.v3i2.808.
-