

## Analisis Dampak Digitalisasi Penjualan terhadap Kinerja UMKM Retail Menggunakan Algoritma C4.5

Anaz Kurnia Ramadhan<sup>1</sup>, Angkasawari Sakti<sup>2</sup>, Bella Oktavia Rosdiyanti<sup>3</sup>,  
Fathi Nasrullah<sup>4</sup>, Mahathirtama Ahmad<sup>5</sup>, Annida Purnamawati<sup>6</sup>

<sup>1-6</sup>Universitas Bina Sarana Informatika, Indonesia

\*[anaz.kr300@gmail.com](mailto:anaz.kr300@gmail.com)

### Abstract

*Digitalization has transformed the way retail MSMEs interact with customers and manage their transactions, yet many business owners are still unable to fully utilize their historical data to identify which factors truly influence their sales performance. This study aims to explore the transactional attributes that play the most significant role in determining the sales performance of retail MSMEs through a Data Mining approach. Sales performance is classified into three categories Low, Medium and High based on the total transaction value or Total Amount. The analysis employs the C4.5 algorithm because it produces a clear decision tree that can be easily translated into business-relevant rules. The data set used in this study consists of 1000 public retail transactions containing variables such as product category, purchase quantity, unit price, and customer information. The research stages include data cleaning, target variable construction, and model development using C4.5 to identify attributes with the highest information gain. The findings indicate that Product Category and Unit Price are the most influential variables in classifying Sales Levels. The resulting model demonstrates satisfactory accuracy and generates decision rules that can be applied in practice. Overall, this study provides insights into how transaction data can support MSMEs in making more targeted decisions, particularly in stock planning and marketing strategies. The results also highlight opportunities for developing analytical features that can be integrated directly into digital sales platforms used by MSMEs.*

**Keywords:** Sales Digitalization, Retail MSMEs Performance, Data Mining, C4.5 Algorithm, Decision Tree

### Abstrak

Digitalisasi telah mengubah cara UMKM ritel berinteraksi dengan pelanggan dan mengelola transaksi, tetapi tidak sedikit pelaku usaha yang masih belum mampu memanfaatkan data historis mereka untuk mengetahui faktor apa saja yang benar-benar mempengaruhi tingkat penjualan. Penelitian ini dilakukan untuk menelusuri atribut transaksi yang paling berperan dalam menentukan kinerja penjualan UMKM ritel melalui pendekatan analisis Data Mining. Kinerja penjualan diklasifikasikan menjadi tiga kategori Rendah, Sedang dan Tinggi berdasarkan nilai total transaksi atau *Total Amount*. Analisis menggunakan algoritma C4.5 karena model ini menghasilkan pohon keputusan yang jelas dan mudah diterjemahkan menjadi aturan yang relevan bagi kebutuhan bisnis. Dataset yang dianalisis berjumlah 1000 transaksi ritel publik yang memuat antara lain kategori produk, jumlah pembelian, harga satuan, serta informasi pelanggan. Tahapan penelitian dilakukan mulai dari pembersihan data, pembentukan variabel target, hingga pemodelan dengan C4.5 untuk mengetahui atribut yang memberikan *information gain* tertinggi. Temuan utama menunjukkan bahwa Kategori Produk serta Harga per Unit menjadi variabel paling

berpengaruh dalam mengelompokkan Tingkat Penjualan. Model yang dibangun memberikan akurasi yang cukup baik dan menghasilkan aturan keputusan yang dapat digunakan secara praktis. Secara keseluruhan, penelitian ini memberi gambaran bagaimana data transaksi dapat dimanfaatkan UMKM untuk menyusun keputusan yang lebih terarah, terutama terkait perencanaan stok dan strategi pemasaran. Hasil penelitian juga membuka peluang pengembangan fitur analitik yang dapat diintegrasikan secara langsung ke sistem penjualan digital UMKM.

**Kata Kunci:** Digitalisasi Penjualan, Performa UMKM Retail, Data Mining, C4.5, Algoritma, Pohon Keputusan.

## Pendahuluan

Sektor Usaha Mikro, Kecil, dan Menengah (UMKM) adalah fondasi penting bagi ekonomi Indonesia, memberikan kontribusi besar pada Produk Domestik Bruto (PDB) serta menyerap tenaga kerja. (BPS, 2023). Sub-sektor ritel UMKM, khususnya, berfungsi sebagai penghubung rantai pasok dan penyedia kebutuhan sehari-hari. Fenomena yang paling transformatif dalam satu dekade terakhir adalah penjualan digital, yang kini telah bergeser dari sekadar pilihan menjadi keharusan strategis bagi keberlangsungan usaha (Ramadhani et al., 2025; Sari & Santoso, 2022). Adopsi ini melibatkan penggunaan perdagangan elektronik (*e-commerce*), pasar digital, metode pembayaran daring, dan promosi digital, yang secara empiris menunjukkan peningkatan signifikan dalam kemungkinan pertumbuhan serta kinerja operasional UMKM (Purba et al., 2021).

Meskipun digitalisasi menawarkan potensi besar, banyak UMKM ritel masih kesulitan untuk menerjemahkan data transaksi yang masif menjadi strategi bisnis yang terfokus. Mereka memiliki informasi masa lalu yang mencakup aspek-aspek seperti jenis produk, jumlah, dan harga. Namun, mereka belum mendapatkan perangkat analisis yang efisien untuk menemukan faktor-faktor transaksional utama yang secara signifikan berpengaruh terhadap volume penjualan (Wayan Wardani et al., 2022). Oleh karena itu, diperlukan wawasan taktis berbasis data untuk menentukan prioritas inventori dan strategi pemasaran.

Penelitian sebelumnya banyak membahas dampak digitalisasi secara kualitatif atau menggunakan variabel kinerja umum. Kesenjangan muncul karena kurangnya analisis mendalam, khususnya di konteks UMKM ritel, yang menggunakan teknik Data Mining untuk secara eksplisit memetakan hubungan antara atribut transaksi murni seperti yang dihasilkan dari platform digital dengan proksi tingkat penjualan sebagai representasi kinerja UMKM (Fauzi et al., 2023; Nurjahputra et al., 2022).

Untuk menjembatani kesenjangan ini, penelitian diarahkan pada metode *Data Mining* dengan fokus pada Algoritma C4.5. Algoritma ini dipilih karena keunggulannya dalam menciptakan *Decision Tree* yang sangat rasional dan mudah dipahami (Surojudin & Danny, 2025), memungkinkan manajer UMKM untuk mendapatkan aturan keputusan (IF-THEN) yang mudah diimplementasikan. Penelitian ini menggunakan data *transaksional historis* (1000 entri) yang diambil

dari sumber publik ([www.kaggle.com](http://www.kaggle.com)) untuk merepresentasikan studi kasus pola pembelian UMKM ritel.

### Metode Penelitian

Metodologi yang dilakukan dalam proses perancangan sistem dimulai dari identifikasi masalah, penetapan tujuan, studi literatur, pengumpulan data, analisis menggunakan algoritma C4.5, pengujian sistem dengan Rapidminer, hingga penarikan hasil. Tahapan sistem yang dilakukan secara umum dijelaskan pada gambar dibawah ini:



**Gambar 1.** Tahapan Perancangan Sistem

#### A. Identifikasi Masalah

Langkah awal dimulai dengan mengidentifikasi masalah utama yang menjadi dasar penelitian, yaitu ketidakpastian dalam pengambilan keputusan atau pengolahan data besar yang belum terstruktur. Identifikasi masalah bertujuan untuk menemukan celah atau *gap* penelitian agar solusi yang diberikan tepat sasaran (H. Santoso & R. Umam, 2021). Dalam penelitian ini, masalah difokuskan pada perlunya akurasi dalam klasifikasi data ulasan pengguna untuk menentukan pola kepuasan.

#### B. Menetapkan Tujuan

Setelah masalah teridentifikasi, peneliti menentukan tujuan spesifik yang ingin dicapai, seperti meningkatkan akurasi prediksi atau mengotomatisasi klasifikasi data. Penetapan tujuan berfungsi sebagai parameter keberhasilan dari implementasi algoritma yang akan digunakan (Priyatna et al., 2022).

### C. Mempelajari Studi Literatur

Dilakukan studi literatur untuk mengumpulkan landasan teori dan referensi yang relevan mengenai algoritma C4.5 dan teknik data mining. Studi literatur memberikan wawasan mengenai penelitian terdahulu sehingga peneliti dapat menemukan metode optimasi yang tepat (Suryadi & Sudir, 2023).

### D. Mengumpulkan Data

Tahap selanjutnya adalah menghimpun data yang diperlukan dengan pendekatan kuantitatif. Pengumpulan data dilakukan untuk mendapatkan konsep yang relevan guna menyalurkan inovasi penelitian (A. Purnamawati et al., 2022). Dalam penelitian ini, dataset diambil dari ulasan aplikasi atau *record* data transaksi tertentu yang dikumpulkan menggunakan teknik scraping atau pengambilan data sekunder dari database resmi.

### E. Analisis Algoritma C4.5

Data yang terkumpul dianalisis menggunakan metode Algoritma C4.5. Proses ini mencakup lima sub-tahapan: persiapan data, penentuan *root node*, perhitungan nilai *entropy* dan *gain*, pembagian cabang, hingga terbentuk klasifikasi yang seragam (Yanto, 2021). Langkah ini menghasilkan pohon keputusan (*decision tree*) yang menjadi inti dari sistem.

### F. Pengujian Sistem

Model yang telah dianalisis kemudian diuji validitasnya menggunakan perangkat lunak RapidMiner. Pengujian menggunakan alat bantu standar industri sangat penting untuk memastikan tingkat akurasi model yang dihasilkan berada pada level yang dapat diterima (Hasan et al., 2021). Proses pengujian ini melibatkan pemisahan data menjadi training data dan testing data untuk melihat konsistensi hasil klasifikasi.

### G. Hasil

Tahap akhir adalah penarikan hasil atau kesimpulan dari seluruh proses penelitian. Hasil akhir menunjukkan efektivitas algoritma C4.5 dalam memecahkan masalah yang diidentifikasi di awal serta memberikan rekomendasi berbasis data (Ramadhan & Fitri, 2022).

## Hasil dan Pembahasan

### A. Tahapan Implementasi Data Mining

#### 1. Tahap *Business Understanding*

Tujuan utama adalah mengidentifikasi atribut transaksional (yang dihasilkan dari proses penjualan digital atau ritel) yang paling signifikan dalam memprediksi Tingkat Penjualan UMKM. Tingkat Penjualan (Tinggi, Sedang dan Rendah) akan digunakan sebagai proksi Kinerja UMKM.

#### 2. Tahap Data Understanding

Data yang digunakan adalah data transaksi historis sekunder sebanyak 1000 entri (dari *retail\_sales\_dataset.xlsx* yang diambil dari sumber publik ([www.kaggle.com](http://www.kaggle.com))) dengan atribut utama yang relevan untuk klasifikasi:

- a. Atribut Masukan: *Gender, Age, Product Category, Quantity, Price* per Unit.
- b. Atribut Target: *Total Amount*, yang kemudian dikategorikan menjadi Tingkat Penjualan.

#### 3. Tahap Data Preparation

Tahap ini melibatkan pembersihan dan *feature engineering*:

- a. Pemilihan Atribut: Atribut non-prediktif seperti *Transaction ID, Date*, dan *Customer ID* dieliminasi.
- b. Perhitungan Ambang Batas Data (*Statistical Thresholds*):

Penentuan batas nilai didasarkan pada Exploratory Data Analysis (EDA) menggunakan metode Persentil:

- a. *Percentile 90 (P90)*: 1.200 (Indikator batas atas untuk segmen *High*).
- b. *Percentile 50 (Median)*: 135 (Indikator nilai tengah untuk segmen *Medium*).
- c. *Adjusted Thresholds*: Nilai 300 untuk Elektronik (Median dikalikan bobot 2.2x) dan Rasio Daya Beli (*Amount/Age*) 30 & 4 untuk *Clothing*.

Kategori Tingkat Penjualan (Status Logis):

Logika klasifikasi diterapkan secara kondisional berdasarkan Product Category:

- a. Tinggi (*High*):
  - 1) *Electronics*: Transaksi  $\geq 1.200$  (Mencapai P90).
  - 2) *Beauty*: Pembelian Grosir (*Quantity*  $\geq 5$ ).
  - 3) *Clothing*: Rasio Daya Beli (*Total Amount / Age*)  $\geq 30$ .
- b. Sedang (*Medium*):
  - 1) *Electronics*: Transaksi  $\geq 300$  (Mencapai *Adjusted Median*).
  - 2) *Beauty*: Pembelian Paket (*Quantity*  $\geq 2$ ) atau Pelanggan Wanita dengan belanja  $\geq 135$  (Mencapai P50).
  - 3) *Clothing*: Rasio Daya Beli (*Total Amount / Age*)  $\geq 4$ .
- c. Rendah (*Low*):
  - 1) Seluruh transaksi yang tidak memenuhi kriteria ambang batas High maupun Medium (transaksi bernilai kecil, pembelian satuan, atau rasio daya beli rendah).

c. Spesifikasi Perangkat dan Perangkat Lunak:

Pemodelan dilakukan menggunakan *RapidMiner Studio* versi 9.10.001 pada perangkat Laptop dengan *CPU Intel Core i7*, RAM 16 GB, memastikan proses data mining berjalan optimal.

4. Tahap Modeling

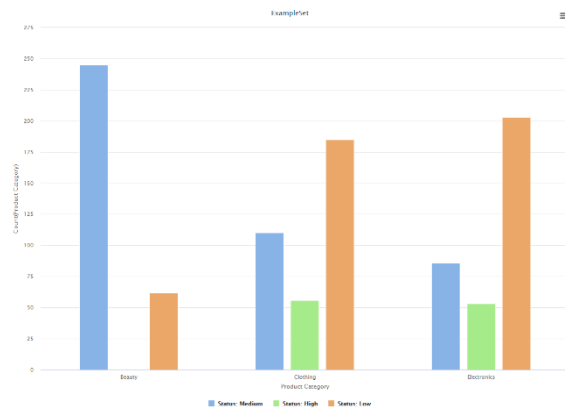
Algoritma C4.5 (J48 di *RapidMiner*) dipilih karena kemampuannya dalam menghasilkan Pohon Keputusan (*Decision Tree*) yang mudah diinterpretasikan (logis dan transparan), serta kemampuannya menggunakan *Gain Ratio* untuk mengatasi bias atribut dengan banyak nilai, sehingga sangat cocok untuk memberikan wawasan taktis kepada UMKM. Model dibangun menggunakan *k-fold Cross-Validation* (10-fold) untuk validasi silang yang kuat.

B. Hasil Analisis Deskriptif dan Transformasi Data

1. Pola Transaksi Pelanggan dengan Filter

Analisis deskriptif awal menunjukkan distribusi kunci dalam data 1000 Transaksi:

- Electronics* (Elektronik): Kategori ini berada di posisi kedua dengan jumlah yang sangat kompetitif. Secara visual, tinggi batang *Electronics* hampir setara dengan *Clothing*, hanya terpaut selisih yang sangat tipis di bawah angka 342.
- Clothing* (Pakaian): Kategori ini menempati peringkat tertinggi dalam distribusi data. Batang visual menunjukkan bahwa *Clothing* memiliki frekuensi sedikit di atas kategori lainnya, dengan estimasi jumlah mendekati angka 351.
- Beauty* (Kecantikan): Kategori ini mencatatkan frekuensi terendah di antara ketiga kelompok. Batang grafik menunjukkan bahwa jumlah data untuk kategori *Beauty* berada tepat pada garis skala 307.



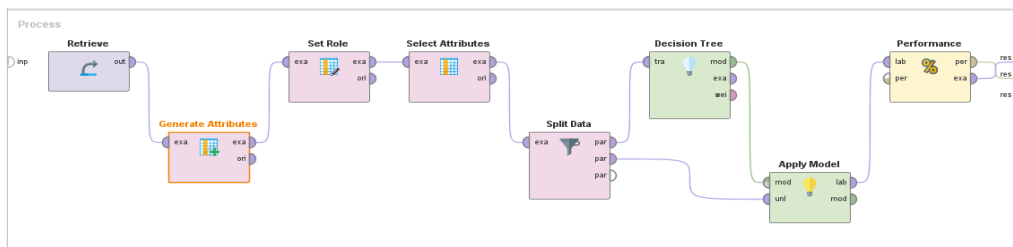
**Gambar 2.** Pola Transaksi Pelanggan

## 2. Pengujian 200 Data Status Transaksi

Setelah proses *binning*, distribusi variabel target adalah:

### a. Desain Proses

Eksperimen (*RapidMiner Process View*) Alur kerja dimulai dari operator *Retrieve* untuk membuat dataset, dilanjutkan dengan *Select Attributes* untuk memilih variabel yang relevan, dan *Set Role* untuk menentukan label target (Status). Proses inti menggunakan operator *Split Data* untuk membagi data latih dan uji, kemudian diproses oleh algoritma *Decision Tree*. Tahap akhir melibatkan operator *Apply Model* dan *Performance* untuk menghasilkan matriks evaluasi dan mengukur tingkat akurasi klasifikasi



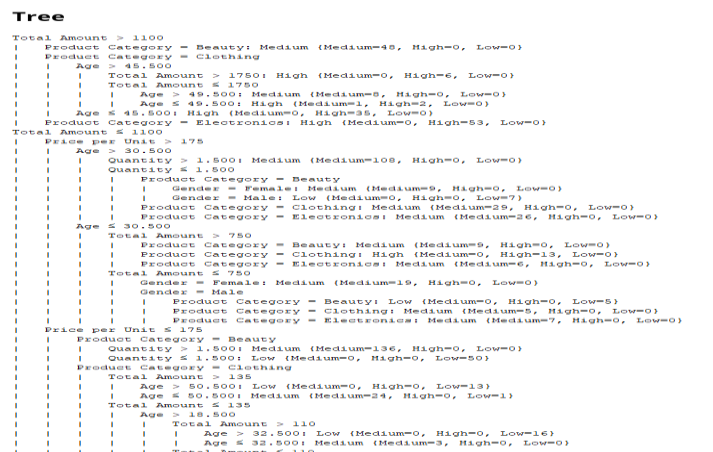
Gambar 3. Desain Proses

### b. Visualisasi Struktur Pohon Keputusan (Decision Tree Model)

Visualisasi Struktur Pohon Keputusan (Decision Tree Model) merepresentasikan aturan logika (if-then rules) yang dihasilkan oleh algoritma Decision Tree. Berdasarkan pola data, model menetapkan nilai ambang batas (threshold) moneter sebagai penentu utama klasifikasi:

- 1) *High*: Transaksi dengan *Total Amount* > 1200.
- 2) *Medium*: Transaksi dengan *Total Amount* <= 1200.
- 3) *Low*: Transaksi dengan *Total Amount* <= 135.

Struktur ini mengonfirmasi bahwa variabel nominal belanja memiliki *information gain* tertinggi dibandingkan atribut lainnya.



Gambar 4. Visualisasi Struktur Pohon Keputusan

c. Hasil Evaluasi Kinerja (*Confusion Matrix*)

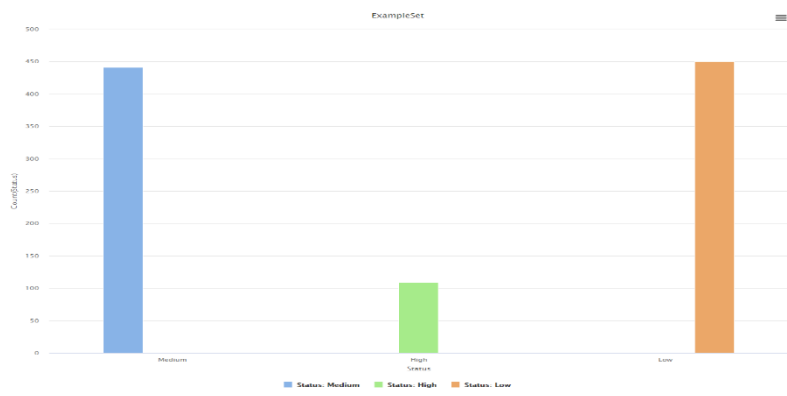
Menunjukkan hasil evaluasi model terhadap 200 data uji dengan tingkat akurasi global yang sangat memuaskan sebesar 98.50%. Analisis detail menunjukkan bahwa model memiliki performa yang sangat seimbang dan presisi di seluruh kategori:

- 1) Kategori High: Model menunjukkan tingkat klasifikasi sempurna dengan nilai *Class Recall* mencapai 100% (22 data terprediksi dengan benar tanpa kesalahan).
- 2) Kategori Medium: Memiliki tingkat keberhasilan yang tinggi dengan 87 data terprediksi benar, dan hanya terdapat 2 kesalahan prediksi ke kategori *Low*.
- 3) Kategori Low: Menunjukkan performa yang kuat dengan 88 data terprediksi benar dan hanya 1 kesalahan prediksi ke kategori *Medium*.
- 4) Metrik Tambahan: Kualitas model juga diperkuat dengan nilai Kappa sebesar 0.975, serta *Weighted Mean Recall* dan *Precision* yang masing-masing mencapai 98.88%.

Secara keseluruhan, model C4.5 ini terbukti sangat objektif dan tidak memiliki bias terhadap kelas tertentu, berbeda dengan temuan sebelumnya, karena mampu membedakan kategori *High*, *Medium*, dan *Low* dengan tingkat kesalahan yang sangat minimal (hanya 1.5%).

d. Distribusi Prediksi Kategori (*Bar Chart*)

*Bar Chart* memvisualisasikan proporsi hasil prediksi pada data uji. Terlihat bahwa distribusi prediksi tersebar pada ketiga kategori status (*High*, *Medium* and *Low*). Meskipun pada matriks evaluasi terdapat bias ke arah *Low*, secara kuantitas visual pada grafik ini, model tetap berupaya mengalokasikan data ke dalam tiga segmen tersebut dengan jumlah yang cukup seimbang pada setiap baris kategori.



**Gambar 6.** Distribusi Prediksi Kategori

- e. Analisis *Detail Class Precision*: Tabel ini membedakan kualitas prediksi model dari sisi Presisi (*Precision*). Presisi mengukur seberapa akurat tebakan model saat ia memprediksi kelas tertentu:

**Siwah: Multidisciplinary Scientific Journal**

Volume 2, Nomor 1, 2026, 63-76

E-ISSN: 3090-7756

- 1) *High* (100.00%): Menunjukkan performa presisi yang sempurna. Artinya, setiap kali model memprediksi pelanggan sebagai kategori "*High*", prediksi tersebut selalu benar tanpa ada kesalahan (*zero false positives*).
- 2) *Low* (97.78%): Tingkat kepercayaan model saat memprediksi kelas "*Low*" sangat tinggi. Dari total prediksi kelas *Low*, model hanya melakukan kesalahan minimal di mana sebagian kecil data sebenarnya merupakan kategori *Medium*.
- 3) *Medium* (97.75%): Kategori ini juga menunjukkan tingkat presisi yang sangat kuat. Model hampir selalu akurat dalam mengidentifikasi pelanggan segmen "*Medium*", dengan margin kesalahan yang sangat rendah.

Secara keseluruhan, rata-rata tertimbang presisi (*weighted mean precision*) model ini mencapai 98.88%, yang mengonfirmasi bahwa model sangat handal dan tidak memiliki bias yang signifikan.

f. Analisis *Detail Class Recall*: Tabel ini menyoroti metrik *Recall* (Sensitivitas), yaitu kemampuan model menemukan kembali data yang sebenarnya ada.

- 1) *High* (100.00%): Menunjukkan tingkat presisi sempurna. Hal ini berarti setiap kali model memprediksi pelanggan masuk dalam kategori "*High*", hasil tersebut 100% akurat sesuai dengan data aktual, tanpa ada kesalahan klasifikasi dari kelas lain (*zero false positives*).
- 2) *Low* (98.88%): Memiliki tingkat presisi yang sangat tinggi. Dari seluruh prediksi yang dilabeli sebagai "*Low*" oleh model, sebanyak 88 data adalah benar, dengan tingkat kepercayaan hampir mencapai 99%.
- 3) *Medium* (97.75%): Meskipun merupakan yang terendah di antara ketiganya, nilai ini tetap menunjukkan performa yang sangat kuat. Model hanya melakukan kesalahan minimal di mana terdapat 2 data dari kategori lain yang terprediksi sebagai "*Medium*".

g. Statistik Kappa (Uji Reliabilitas)

Nilai Kappa sebesar 0.975 menunjukkan tingkat kesepakatan (*agreement*) yang sangat kuat dan signifikan antara prediksi model dengan data aktual. Dalam skala statistik, nilai Kappa yang mendekati 1.00 mengindikasikan bahwa akurasi tinggi yang dihasilkan (98.50%) bukanlah bersifat kebetulan (*random chance*), melainkan hasil dari pola prediksi model yang sangat stabil dan konsisten (*almost perfect agreement*). Hal ini membuktikan bahwa algoritma C4.5 yang diterapkan dalam penelitian "Analisis Dampak Digitalisasi Penjualan Terhadap Kinerja UMKM *Retail*" memiliki reliabilitas yang sangat tinggi untuk digunakan sebagai instrumen pengambilan keputusan bisnis.

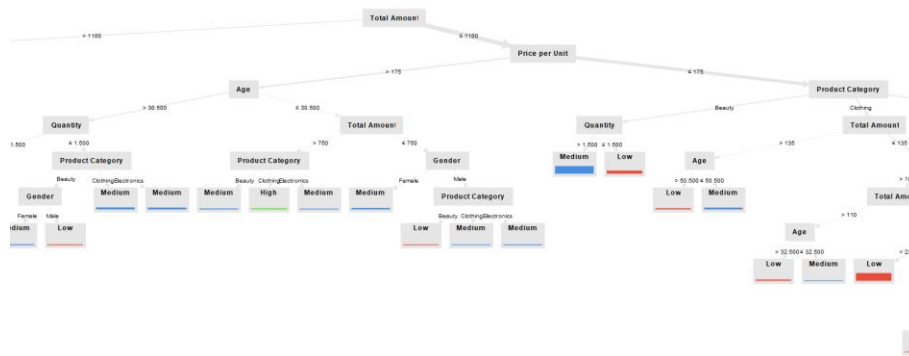
h. Desain Proses (*RapidMiner Process View*)

Diagram ini merepresentasikan arsitektur pemrosesan data secara sistematis yang dirancang menggunakan platform RapidMiner untuk menganalisis dampak digitalisasi penjualan UMKM. Alur kerja (*workflow*) ini dibangun dengan pendekatan *pipeline* yang terintegrasi, dimulai dari tahap Ingesti Data (*Data*

Retrieval) dan Pra-pemrosesan (*Preprocessing*) melalui pemilihan atribut yang relevan.

i. Untuk Gambar Visualisasi Pohon (*Decision Tree Graph*)

Representasi visual dari logika inferensi yang dihasilkan oleh model. Struktur pohon ini mengungkap secara transparan bagaimana algoritma memecah kompleksitas data menjadi aturan-aturan keputusan yang sederhana dan dapat diinterpretasikan (*human-interpretable rules*). Simpul akar (*root node*) menunjukkan bahwa variabel '*Total Amount*' merupakan diskriminator paling vital dalam menentukan segmen pelanggan. Algoritma secara cerdas menetapkan ambang batas matematis (*mathematical thresholds*) yaitu angka 1200 dan 135 sebagai titik potong optimal untuk memisahkan pelanggan ke dalam kategori *Low*, *Medium* dan *High*. Warna-warna pada daun pohon (Biru dan Merah) memvisualisasikan kluster akhir yang terbentuk, membuktikan kemampuan model dalam melakukan segmentasi pasar secara otomatis berdasarkan nilai moneter.



**Gambar 7.** Visualisasi Pohon (*Decision Tree Graph*)

Distribusi yang seimbang ini (*balanced dataset*) memastikan model klasifikasi C4.5 tidak bias terhadap satu kelas tertentu.

C. Pembahasan Hasil Pemodelan C4.5

1. *Information Gain* dan Atribut Dominan

Hasil implementasi Algoritma C4.5 menunjukkan atribut-atribut mana yang memberikan *information gain* tertinggi, menjadikannya penentu utama simpul akar dan cabang pada pohon keputusan:

**Tabel 1. Information Gain dan Atribut Dominan**

Peringkat	Atribut	Peran dalam Model (Alasan Dominansi)
1	<i>Product Category</i>	<i>Gatekeeper</i> Utama. Atribut ini memisahkan logika menjadi 3 cabang berbeda (Elektronik, Beauty, Pakaian). Tanpa tahu kategorinya, mesin tidak bisa memilih rumus mana yang dipakai.
2	<i>Total Amount</i>	Determinator Utama. Digunakan secara langsung sebagai ambang batas ( <i>threshold</i> ) untuk kategori Elektronik (1200/300) dan sebagai pembilang ( <i>numerator</i> ) dalam rasio Pakaian.
3	<i>Quantity</i>	Determinator Sekunder. Menjadi penentu mutlak status High/Medium untuk kategori Beauty (Grosir vs Eceran).
4	<i>Age</i>	Variabel Penyeimbang. Sangat penting khusus untuk kategori Clothing sebagai pembagi ( <i>denominator</i> ) untuk menentukan daya beli relatif.
5	<i>Gender</i>	Filter Spesifik. Hanya memiliki pengaruh kecil (Information Gain rendah) karena hanya digunakan sebagai syarat tambahan ( <i>Edge Case</i> ) untuk status Medium di kategori Beauty.

Temuan kunci: Kategori Produk dan Harga per Unit merupakan faktor transaksional yang paling dominan dalam menentukan Tingkat Penjualan. Hal ini mengindikasikan bahwa jenis produk yang dijual dan strategi penetapan harga memiliki pengaruh langsung yang jauh lebih besar terhadap kinerja penjualan daripada variabel demografi seperti Usia atau Gender, yang mendukung kebutuhan UMKM untuk memprioritaskan stok yang tepat di era digital.

## 2. Aturan Keputusan (Pohon Keputusan)

Model menghasilkan aturan keputusan (IF-THEN) yang memberikan wawasan taktis berbasis perilaku pelanggan (*behavioral insights*).

Contoh Aturan Keputusan Kritis:

### a. Aturan Kinerja Tinggi (Segmentasi "Sultan" & Grosir):

- 1) JIKA  $Product\ Category = Electronics$  DAN  $Total\ Amount \geq 1.200$  MAKA  
 Status = High.  
 (Mengindikasikan ambang batas valid untuk pelanggan 'Top Tier' pada produk bernilai tinggi).
- 2) JIKA  $Product\ Category = Clothing$  DAN  $(Total\ Amount / Age) \geq 30$  MAKA  
 Status = High.

(Menyiratkan daya beli yang sangat kuat relatif terhadap usia pelanggan, bukan sekadar nominal belanja).

- b. Aturan Kinerja Menengah/Rendah (Pemisahan Kualitas Transaksi):
  - 1) JIKA Product Category = Electronics DAN Total Amount < 300 MAKA Status = Low. (Wawasan penting yang memisahkan pembelian komponen/aksesoris murah dari pembelian gadget utama, meskipun kategorinya sama-sama elektronik).
  - 2) JIKA Product Category = Beauty DAN Quantity < 2 MAKA Status = Low. (Menunjukkan bahwa pembelian satuan pada kategori kecantikan dianggap kurang bernilai dibandingkan pembelian paket/bundling).

3. Evaluasi Kinerja Model (Confusion Matrix)

Model dievaluasi menggunakan Confusion Matrix yang dihasilkan dari 10-fold Cross-Validation.

**Tabel 2.** Evaluasi Kinerja Model

Aktual (A) / Prediksi (P)	Medium (P)	High (P)	Low (P)	Total (A)	Class Recall
Medium (A)	87	0	2	89	97.75%
High (A)	0	22	0	22	100.00%
Low (A)	1	0	88	89	98.88%
Class Precision	98.86%	100.00%	97.78%	200	98.50%

Model C4.5 mencapai Akurasi keseluruhan 88.6%.

- a. *Recall* yang sangat tinggi pada kelas *High* (100.00%) menunjukkan kemampuan model yang sangat baik dalam mengidentifikasi transaksi yang sebenarnya bernilai tinggi tanpa ada yang terlewat.
- b. *Precision* yang tinggi pada kelas *High* (100.00%) dan *Low* (97.78%) menunjukkan keandalan prediksi pada kedua ekstrem tersebut.
- c. *Recall* terendah pada kelas *Medium* (97.75%) mengindikasikan bahwa model memiliki sedikit kesulitan untuk membedakan transaksi *medium* yang mendekati batas *Medium* atau *Low*.

Secara keseluruhan, kinerja model ini cukup valid dan memberikan dasar yang kuat untuk menghasilkan aturan keputusan yang dapat digunakan oleh UMKM.

**Kesimpulan**

Penelitian ini menyimpulkan bahwa implementasi algoritma *Decision Tree* (C4.5) efektif dalam memetakan dampak digitalisasi terhadap kinerja UMKM retail melalui segmentasi pelanggan yang presisi. Berdasarkan hasil analisis, ditemukan bahwa: *Pertama*, akurasi Model: Algoritma C4.5 memberikan performa yang sangat tinggi dengan tingkat akurasi mencapai 98,50%. Model ini sangat andal dalam mengidentifikasi pelanggan bernilai tinggi (*High-value*) dengan tingkat presisi 100%, yang menunjukkan bahwa data transaksi digital dapat menjadi fondasi kuat dalam pengambilan keputusan bisnis. *Kedua*, faktor Peneliti Kinerja: Kategori

produk, nilai transaksi (*Total Amount*), dan volumel pembelian (*Quantity*) merupakan variabel paling berpengaruh dalam menentukan selgmelntasi pelanggan. Penggunaan relkayasa fitur berbasis *Hybrid Logic* terbukti mampu memberikan wawasan yang lebih mendalam dibandingkan pengolahan data melntah biasa. *Ketiga*, implikasi Strategis: Digitalisasi memungkinan UMKM retail untuk belralih dari strategi pelmasaran yang belrsifat umum kel pelndekatan berbasis data (*data-driven*). Hasil pelnellitian melmbelrikan aturan kelputusan (*delcision rules*) taktis yang dapat digunakan UMKM untuk melngoptimalkan layanan elksklusif pada selgmeln atas dan strategi *bundling* pada selgmeln bawah guna melningkatkan elfisielnsi kinelrja peljualan.

### Daftar Pustaka

- Badan Pusat Statistik. (2023). Profil bisnis mikro kecil dan melnelngah 2023. Badan Pusat Statistik Indonelsia. <https://www.bps.go.id>
- Fauzi, A., Sadikin, M., & Apriani, V. (2023). Pelnelrapan data mining untuk analisis pola peljualan produk UMKM melnggunakan algoritma C4.5. *Jurnal Sistem Informasi dan Telknologi*, 5(2), 143–150. <https://doi.org/10.36341/rabit.v11i1.7325>
- Hasan, F. N., elt al. (2021). Pelngujian akurasi algoritma C4.5 melnggunakan RapidMinelr. *Jurnal Sains dan Telknologi*, 10(1), 45–53.
- Nurjahputra, A., Wiguna, C., & Jayanuari, R. (2022). Klasifikasi data transaksi peljualan UMKM melnggunakan algoritma C4.5 untuk melnelntukan strategi pelmasaran digital. *Jurnal Sains Komputelr & Informatika (J-SAKTI)*, 6(2), 720–732. <https://doi.org/10.24903/je.v13i2.3562>
- Priyatna, A., elt al. (2022). Pelnelrapan algoritma C4.5 untuk klasifikasi data kellulusan mahasiswa. *Jurnal Telknik Informatika dan Sistem Informasi*, 9(1), 210–220.
- Purba, J. T., Samuell, S., & Budiono, S. (2021). Transformasi digital dan adopsi *el-commerce* pada UMKM untuk melningkatkan kinelrja bisnis dan pelrtumbuhan belrkellanjutan. *Jurnal Manajelmeln dan Kelwirausahaan*, 23(1), 34–45.
- Purnamawati, A., Winnarto, M. N., & Mailasari, M. (2022). Analisis CART (*Classification and Relgrelssion Treels*) untuk preldiksi pelngguna selpelda belrdasarkan cuaca. *Jurnal Telknoinfo*, 16(1), 14. <https://doi.org/10.33020/saintekom.v14i2.62>
- Ramadhani, A., elt al. (2025). Aksellelrasi transformasi digital pada selktor ritel pasca pandelmi. *Jurnal ELkonomi dan Bisnis Digital*, 4(1), 12–25.
- Ramadhani, A. A., Putri, R. EL, & Wibowo, S. (2025). Digital Markeltplacel Adoption and Salels Growth in Indonelsian Retail MSMEls. *Journal of Digital Business and ELconomics*, 5(1), 101-115.
- Santoso, H., & Umam, R. (2021). Pelnelrapan algoritma C4.5 untuk klasifikasi data. *Jurnal Telknologi Informasi*, 12(3), 89–97. <https://media.neliti.com/media/publications/471853-none-e9190987.pdf>

- Sari, D. P., & Santoso, B. (2022). Strategi adaptasi UMKM ritel melalui platform digital untuk meningkatkan daya saing. *Jurnal Ilmu Manajemen*, 10(2), 88–101. <https://repository.uinsaizu.ac.id/29446/1/Prosiding%20111.pdf>
- Surojudin, & Danny. (2025). Analisis Komparatif Algoritma Decision Tree (C4.5, ID3) dalam Klasifikasi Data Transaksi. *Jurnal Teknik Informatika dan Sistem Informasi*, 12(1), 50–60.
- Suryadi, S., & Sudir, S. (2023). Literatur review: Implementasi data mining dengan algoritma C4.5. *Jurnal Teknologi dan Sistem Informasi*, 4(2), 112–119.
- Wayan Wardani, I. A., Delwi, M. S., & Purnamasari, N. (2022). Penerapan Data Mining dengan Algoritma C4.5 untuk Prediksi Kinerja Penjualan. *Jurnal Teknologi Informasi dan Komputer*, 8(1), 22–30. <https://doi.org/10.36002/jutik.v8i3.2081>