

## Analisa Penjualan Rokok Dengan Metode Klasifikasi Menggunakan Metode Algoritma C4.5 Pada CV Jaya Berkah Mas

Sebastianus Aditya Narendra<sup>1</sup>, Pradita Aldi Setiawan<sup>2</sup>, Syaiful Anwar<sup>3</sup>, Mohammad Haddiel Fuad<sup>4</sup>

<sup>1,2,3,4</sup>Universitas Bina Sarana Informatika

Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10450, Indonesia

e-mail: <sup>1</sup>adityanarendra021218@gmail.com, <sup>2</sup>praditaaldisetiawan@gmail.com, <sup>3</sup>syaiful.sfa@bsi.ac.id, <sup>4</sup>mhfuad@bsi.ac.id

Artikel Info : Diterima : 10-10-2024 | Direvisi : 11-12-2024 | Disetujui : 08-01-2025

**Abstrak** - Penelitian ini bertujuan untuk memberikan wawasan yang bermanfaat bagi CV Jaya Berkah Mas dalam menentukan strategi penjualan yang optimal. Penelitian dilakukan dengan mengumpulkan data melalui observasi dan wawancara, serta melakukan pembersihan data untuk mengatasi inkonsistensi dan duplikasi. Algoritma C4.5 digunakan untuk menghasilkan pohon keputusan yang dapat memprediksi penjualan rokok terlaris. Hasil dari analisis menunjukkan bahwa metode C4.5 dapat digunakan untuk memberikan wawasan yang bermanfaat dalam menentukan strategi penjualan yang optimal. Penelitian ini memberikan kontribusi dalam bidang data mining dengan menerapkan algoritma C4.5 berupa memberikan contoh nyata bagaimana teknik data mining, khususnya algoritma C4.5, dapat diterapkan dalam industri untuk menghasilkan informasi yang berguna dan actionable. Hasil nilai akurasi tertinggi perhitungan Algoritma C4.5 dengan data sebanyak 212 entry menunjukkan tingkat Akurasi 96.73%, Presisi 98.00%, dan Recall 85.17%.

Kata Kunci : Penjualan rokok, Data Mining, Algoritma C4.5, Decision Tree, Rapidminer

**Abstracts** - This research aims to provide useful insights for CV Jaya Berkah Mas in determining the optimal sales strategy. The research was conducted by collecting data through observation and interviews, as well as cleaning the data to overcome inconsistencies and duplications. The C4.5 algorithm is used to generate a decision tree that can predict best-selling cigarette sales. The results of the analysis show that the C4.5 method can be used to provide useful insights in determining the optimal sales strategy. This research contributes to the field of data mining by applying the C4.5 algorithm in the form of providing a real example of how data mining techniques, especially the C4.5 algorithm, can be applied in the industry to produce useful and actionable information. The results of the highest accuracy value of the C4.5 Algorithm calculation with data of 212 entries showed a level of Accuracy of 96.73%, Precision of 98.00%, and Recall of 85.17%.

Keywords : Cigarette sales, Data Mining, C4.5 Algorithm,, Decision Tree, Rapidminer

## PENDAHULUAN

Industri rokok merupakan salah satu sektor ekonomi yang signifikan di Indonesia, dengan kontribusi besar terhadap pendapatan negara dan lapangan kerja. CV Jaya Berkah Mas adalah salah satu perusahaan yang bergerak dalam distribusi rokok di berbagai wilayah di Indonesia. Dalam upaya untuk meningkatkan penjualan dan memahami pola pembelian konsumen, analisis data penjualan menjadi sangat penting.

CV Jaya Berkah Mas menghadapi tantangan dalam memahami pola penjualan rokok yang bervariasi berdasarkan jenis produk, waktu, dan lokasi penjualan. Tanpa analisis yang tepat, perusahaan mungkin kehilangan peluang untuk meningkatkan penjualan dan mengoptimalkan strategi pemasaran.

Penelitian ini bertujuan untuk menganalisis data penjualan rokok di CV Jaya Berkah Mas menggunakan metode klasifikasi C4.5. Algoritma C4.5 dipilih karena kemampuannya dalam mengolah data yang kompleks dan menghasilkan pohon keputusan yang dapat digunakan untuk mendukung keputusan bisnis. (Tukino, 2019)

Penelitian ini diharapkan dapat memberikan kontribusi praktis bagi CV Jaya Berkah Mas dalam bentuk rekomendasi strategi penjualan yang lebih efektif. Selain itu, penelitian ini juga memberikan kontribusi teoretis dalam bidang data mining, khususnya dalam penerapan algoritma C4.5 untuk analisis penjualan. (Suriani, 2023)

Artikel ini disusun dengan urutan sebagai berikut: Bagian pertama adalah Pendahuluan yang menjelaskan latar belakang, masalah, tujuan, dan kontribusi penelitian. Bagian kedua adalah Metode Penelitian yang menjelaskan metodologi yang digunakan dalam analisis data. Bagian ketiga adalah Hasil dan Pembahasan yang memaparkan hasil analisis dan interpretasinya. Bagian terakhir adalah Kesimpulan yang merangkum temuan utama dan memberikan rekomendasi untuk penelitian selanjutnya.

Penelitian ini didasarkan pada beberapa penelitian terdahulu yang telah dilakukan dalam bidang yang sama, seperti:

1. Penelitian terdahulu yang dilakukan oleh (Leonardi et al., 2021): “Prediksi penjualan produk rokok menggunakan algoritma C4.5 pada PT. Indomarco Prismatama”, dengan tingkat akurasi 92,11%.
2. Penelitian terdahulu yang dilakukan oleh (Triawan et al., 2024): “Penerapan Data Mining untuk Prediksi Penjualan Spanduk Menggunakan Algoritma C4.5”, mencapai tingkat kinerja algoritma sebesar 84%.
3. Penelitian terdahulu yang dilakukan oleh (Musa et al., 2024): “Penerapan Data Mining Untuk Klasifikasi Data Penjualan Pakan Ternak Terlaris Dengan Algoritma C4.5”, menemukan bahwa kategori pakan memiliki gain tertinggi sebesar 0.306739968 dan *entropy* pakan ayam pedaging sebesar 0.99107606.
4. Penelitian terdahulu yang dilakukan oleh (Wahyudi, 2023): “Penerapan Data Mining Dengan Algoritma C4.5 Dalam Prediksi Penjualan Buku”, hasil pengukuran akurasi data menunjukkan bahwa dari 90 data pelatihan, *Confusion Matrix* menghasilkan nilai akurasi 100%, dan data pengujian juga mencapai tingkat akurasi 100%.

Pada penelitian ini menerapkan metode usulan tentang "Analisa Penjualan Rokok Dengan Metode Klasifikasi Menggunakan Formula Algoritma C4.5 Pada Cv Jaya Berkah Mas" sebagai berikut:

1. **Pengumpulan Data:** Mengumpulkan data penjualan rokok dari CV. Jaya Berkah Mas diambil dari data sekunder yang sudah pernah tercatat pada dokumen perusahaan pada awal tahun 2024. Sedangkan elemen data yang dipakai pada penelitian ini seperti tanggal penjualan, status penjualan, kode sales, nama sales, alamat pelanggan, produk rokok, total kuantitas, dan informasi lain yang relevan.
2. **Analisis Data Menggunakan Algoritma C4.5:** Meliputi persiapan data, pembuatan model, dan evaluasi model.
3. **Pengembangan Penjualan:** Mengembangkan penjualan yang berfokus pada data rokok yang telah diidentifikasi.

## METODE PENELITIAN

Pada bagian ini, penulis akan memaparkan proses penelitian langkah demi langkah, yang dimulai dengan mengidentifikasi masalah dan diakhiri dengan evaluasi:



Sumber:(Musa et al., 2024)

**Gambar 1.** Tahapan Penelitian

### 1. Pengumpulan Data

Pengumpulan data dilakukan dengan observasi langsung serta melakukan wawancara dengan Kepala Gudang CV Jaya Berkah Mas, Jl. Mustari No.17, RT.03/RW.09, Dusun 4, Penaruban, Kec. Kaligondang, Kabupaten Purbalingga, Jawa Tengah 53331.

### 2. Seleksi Data

Seleksi data untuk memilih data yang sesuai dengan kebutuhan untuk memperoleh pengetahuan. Data hasil seleksi digunakan dalam proses data mining, untuk memilih semua data yang diperlukan untuk diproses lebih lanjut. Data disimpan dalam file yang terpisah dari database produksi agar mudah digunakan nanti (Shaumi et al., 2022).

### 3. Transformasi Data

Transformasi data Disebut juga integrasi data. Fase ini mengubah data yang dipilih ke dalam format yang sesuai untuk pemrosesan semantik dengan menormalkan dan menggabungkannya (Zai, 2022).

### 4. Perhitungan Nilai *Entropy* dan *Gain*

Hitung entropi seluruh atribut/variabel, entropi persamaan (1), dan information gain persamaan (2) untuk mencari information gain tertinggi. Ini akan digunakan sebagai simpul akar ketika membangun pohon keputusan.

### 5. *Decision Tree*

Pohon keputusan merupakan hasil proses penghitungan entropi dan perolehan informasi setelah setiap atribut pohon mempunyai kelas dan proses penghitungan tersebut diulangi hingga tidak dapat dilakukan lagi.

### 6. *Role Model*

*Role Model*/Model aturan menggambarkan pohon keputusan.

### 7. Analisis Hasil Perhitungan

Analisis hasil pengujian dalam *data mining* adalah proses mengevaluasi model untuk menilai kinerja prediktif atau klasifikasinya. Pertama, bagi data menjadi set pelatihan dan set pengujian. Metode evaluasi yang umum digunakan meliputi validasi silang, matriks konfusi (untuk menghitung presisi, presisi, perolehan, dan skor F1), serta kurva ROC dan AUC untuk mengukur performa model secara keseluruhan. Masalah regresi menggunakan metrik seperti *mean absolute error* (MAE) dan *mean squared error* (MSE). Hasil evaluasi kemudian diinterpretasikan untuk menentukan seberapa baik kinerja model.

“Instrumen penelitian merupakan pedoman tertulis yang mencakup wawancara, pengamatan, dan pertanyaan yang disiapkan untuk memperoleh informasi.” (Ovan & Saputra, 2020).

Dalam penelitian ini peneliti menggunakan 3 pengumpulan data untuk mengumpulkan data-data yang penting dalam penelitian ini sebagai berikut:

#### 1. Observasi

Observasi merupakan teknik yang sangat umum digunakan dalam penelitian kualitatif.

#### 2. Wawancara

Wawancara merupakan salah satu teknik utama untuk mengumpulkan data.

#### 3. Dokumentasi

Dokumentasi merupakan catatan otentik atau dokumen asli yang dapat digunakan sebagai bukti dalam masalah hukum.

Untuk mencapai tujuan penelitian maka analisa yang digunakan adalah analisis data kuantitatif. Analisis data kuantitatif merupakan suatu analisis data yang dipergunakan apabila kesimpulan-kesimpulan yang diperoleh berupa angka-angka dan juga dalam perhitungan dipergunakan rumus yang ada hubungannya dengan analisis penulisan (Umrati & Wijaya, 2020). Dalam hal ini akan dipergunakan analisis menggunakan metode Algoritma C4.5 sebagai berikut:

#### 1. *Data Mining*

*Data Mining* proses menggali informasi dan pola yang bermanfaat dari data yang sangat besar. Proses *data mining* meliputi pengumpulan data, ekstraksi data, analisis data, dan statistik data. Proses ini juga dikenal sebagai *knowledge discovery*, *knowledge extraction*, *data/pattern analysis*, dan *information harvesting*. Keempat proses dalam data mining ini menghasilkan model atau pengetahuan yang sangat berguna (Arhami & Nasir, 2020).

#### 2. *Rapidminer*

*RapidMiner* adalah perangkat lunak sumber terbuka (open source) yang digunakan untuk analisis data mining dan prediksi. *RapidMiner* menawarkan berbagai teknik deskriptif dan prediktif untuk memberikan wawasan kepada pengguna, membantu mereka membuat keputusan yang optimal (Pratiwi et al., 2020).

### 3. Klasifikasi

Febriani dan Sulistiani mengemukakan “Klasifikasi merupakan proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan memperkirakan kelas dari suatu objek yang labelnya belum diketahui. Model tersebut dapat berupa aturan “jika maka”, pohon keputusan, atau formula matematis” (Anggi Trifani et al., 2022).

### 4. Algoritma C4.5

Algoritma C4.5 program yang memproses sekumpulan data untuk menghasilkan pohon keputusan sebagai output. Pohon keputusan yang dihasilkan kemudian divalidasi menggunakan data uji berlabel yang belum terlihat untuk menghitung generalisasinya (Sugara et al., 2019). Untuk menyelesaikan kasus pada Algoritma C4.5 supaya lebih mudah dipahami prosesnya adalah sebagai berikut:

#### a. Mencari Nilai Gain

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n - \frac{|s_i|}{s} * Entropy(S) \dots \dots (1)$$

Dimana keterangan dari rumus tersebut adalah:

- S = Himpunan Kasus
- A = atribut
- n = jumlah partisi
- pi = porsi dari si kepada s
- |si| = jumlah kasus pada nilai atribut/kriteria
- |s| = jumlah kasus atau total kasus

#### b. Mencari Nilai Entropy

$$Entropy(S) = - \sum_{i=1}^n - pi * Log_2 pi \dots \dots (2)$$

- Keterangan:
- S : Himpunan Kasus
- N: Jumlah Partisi S
- Pi: Proporsi dari Si terhadap S.

### 5. Decision Tree

*Decision Tree* (Pohon Keputusan) adalah salah satu metode klasifikasi yang paling populer karena mudah diinterpretasikan oleh manusia. *Decision Tree* merupakan model prediksi yang menggunakan struktur pohon atau hierarki. Konsep dasar dari *Decision Tree* adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan. Manfaat utama dari penggunaan *Decision Tree* adalah kemampuannya untuk memecah proses pengambilan keputusan yang kompleks menjadi lebih sederhana, sehingga mempermudah dalam menginterpretasikan solusi dari permasalahan (Fauziningrum & Suryaningsih, 2021).

## HASIL DAN PEMBAHASAN

### 1. Pengumpulan Data

Data ini diambil dari data sekunder penjualan rokok di CV Jaya Berkah Mas pada bulan April 2024. Memiliki data sebanyak 212 Entry.

| NAMA PRODUK       | Unit   | Kategori Rokok | Jenis Rokok | Aroma        | Rasa          | Ukuran Batang Rokok | Kepadatan Rokok | Hisapan Rokok | Kemasan Rokok   | Label       |
|-------------------|--------|----------------|-------------|--------------|---------------|---------------------|-----------------|---------------|-----------------|-------------|
| CENGKEH ABADI     | 6 PCS  | SKT            | Kretek      | Sangat Wangi | Blueberry     | Sedang              | Keras           | Ringan        | Plastik         | Laris       |
| CENGKEH ABADI     | 10 PCS | SKT            | Kretek      | Sangat Wangi | Blueberry     | Sedang              | Keras           | Ringan        | Plastik         | Laris       |
| HARUM CENGKEH     | 3 PCS  | SKT            | Kretek      | Sangat Wangi | Wangi Cengkeh | Sedang              | Sedang          | Berat         | Plastik         | Tidak Laris |
| DAUN DJAMBOE      | 20 PCS | SKT            | Kretek      | Sangat Wangi | Pahit         | Besar               | Sedang          | Ringan        | Plastik         | Laris       |
| REED LITE         | 12 PCS | SKM            | Filter      | Sangat Wangi | Manis         | Sedang              | Sedang          | Berat         | Kertas Grenjeng | Laris       |
| REED BOLD         | 12 PCS | SKM            | Filter      | Sangat Wangi | Manis         | Sedang              | Sedang          | Ringan        | Kertas Grenjeng | Laris       |
| CENGKEH ABADI     | 7 PCS  | SKT            | Kretek      | Sangat Wangi | Blueberry     | Sedang              | Keras           | Ringan        | Plastik         | Laris       |
| SOSROBAHU         | 5 PCS  | SKT            | Kretek      | Lembut       | Manis         | Sedang              | Sedang          | Ringan        | Plastik         | Tidak Laris |
| REED LITE         | 16 PCS | SKM            | Filter      | Sangat Wangi | Manis         | Sedang              | Sedang          | Berat         | Kertas Grenjeng | Laris       |
| SOSROBAHU         | 5 PCS  | SKT            | Kretek      | Lembut       | Manis         | Sedang              | Sedang          | Ringan        | Plastik         | Tidak Laris |
| SOSROBAHU         | 4 PCS  | SKT            | Kretek      | Lembut       | Manis         | Sedang              | Sedang          | Ringan        | Plastik         | Tidak Laris |
| DAUN DJAMBOE      | 8 PCS  | SKT            | Kretek      | Sangat Wangi | Pahit         | Besar               | Sedang          | Ringan        | Plastik         | Laris       |
| REED BOLD         | 7 PCS  | SKM            | Filter      | Sangat Wangi | Manis         | Sedang              | Sedang          | Ringan        | Kertas Grenjeng | Laris       |
| SIMUSTIKA         | 6 PCS  | SKT            | Kretek      | Lembut       | Pedes         | Sedang              | Sedang          | Ringan        | Plastik         | Laris       |
| TUJUH DAUN REFILL | 10 PCS | SKT            | Kretek      | Lembut       | Manis         | Besar               | Keras           | Ringan        | Kertas Grenjeng | Laris       |

Sumber: Penelitian Mandiri (2024)

Gambar 2. Data Penjualan Rokok

2. Tahap Pre-Processing

Dalam tahap Pre-Processing atau tahap menyiapkan data agar lebih siap untuk dilakukan lebih lanjut, peneliti melakukan seleksi data sebagai berikut:

a. Seleksi Data

Proses seleksi atribut dilakukan untuk memilih variabel yang akan digunakan dalam penelitian. Variabel yang digunakan pada penelitian yaitu Kecamatan, Kabupaten, Nama Produk, Kategori Rokok, Jenis Rokok, Aroma, Rasa, Ukuran Batang Rokok, Kepadatan Rokok, Hisapan Rokok, Kemasan Rokok, Harga Satuan, dan Total. Berikut pada gambar dibawah ini dataset yang sudah di seleksi.

| Unit   | Kategori Rokok | Jenis Rokok | Aroma        | Rasa          | Ukuran Batang Rokok | Kepadatan Rokok | Hisapan Rokok | Kemasan Rokok   | Label       |
|--------|----------------|-------------|--------------|---------------|---------------------|-----------------|---------------|-----------------|-------------|
| 6 PCS  | SKT            | Kretek      | Sangat Wangi | Blueberry     | Sedang              | Keras           | Ringan        | Plastik         | Laris       |
| 10 PCS | SKT            | Kretek      | Sangat Wangi | Blueberry     | Sedang              | Keras           | Ringan        | Plastik         | Laris       |
| 3 PCS  | SKT            | Kretek      | Sangat Wangi | Wangi Cengkeh | Sedang              | Sedang          | Berat         | Plastik         | Tidak Laris |
| 20 PCS | SKT            | Kretek      | Sangat Wangi | Pahit         | Besar               | Sedang          | Ringan        | Plastik         | Laris       |
| 12 PCS | SKM            | Filter      | Sangat Wangi | Manis         | Sedang              | Sedang          | Berat         | Kertas Grenjeng | Laris       |
| 12 PCS | SKM            | Filter      | Sangat Wangi | Manis         | Sedang              | Sedang          | Ringan        | Kertas Grenjeng | Laris       |
| 7 PCS  | SKT            | Kretek      | Sangat Wangi | Blueberry     | Sedang              | Keras           | Ringan        | Plastik         | Laris       |
| 5 PCS  | SKT            | Kretek      | Lembut       | Manis         | Sedang              | Sedang          | Ringan        | Plastik         | Tidak Laris |
| 16 PCS | SKM            | Filter      | Sangat Wangi | Manis         | Sedang              | Sedang          | Berat         | Kertas Grenjeng | Laris       |
| 5 PCS  | SKT            | Kretek      | Lembut       | Manis         | Sedang              | Sedang          | Ringan        | Plastik         | Tidak Laris |
| 4 PCS  | SKT            | Kretek      | Lembut       | Manis         | Sedang              | Sedang          | Ringan        | Plastik         | Tidak Laris |
| 8 PCS  | SKT            | Kretek      | Sangat Wangi | Pahit         | Besar               | Sedang          | Ringan        | Plastik         | Laris       |
| 7 PCS  | SKM            | Filter      | Sangat Wangi | Manis         | Sedang              | Sedang          | Ringan        | Kertas Grenjeng | Laris       |
| 6 PCS  | SKT            | Kretek      | Lembut       | Pedes         | Sedang              | Sedang          | Ringan        | Plastik         | Laris       |
| 10 PCS | SKT            | Kretek      | Lembut       | Manis         | Besar               | Keras           | Ringan        | Kertas Grenjeng | Laris       |
| 8 PCS  | SKT            | Kretek      | Sangat Wangi | Pahit         | Besar               | Sedang          | Ringan        | Plastik         | Laris       |
| 1 PCS  | SKT            | Kretek      | Sangat Wangi | Wangi Cengkeh | Sedang              | Sedang          | Berat         | Plastik         | Tidak Laris |
| 8 PCS  | SKT            | Kretek      | Sangat Wangi | Blueberry     | Sedang              | Keras           | Ringan        | Plastik         | Laris       |
| 17 PCS | SKM            | Filter      | Sangat Wangi | Manis         | Sedang              | Sedang          | Ringan        | Kertas Grenjeng | Laris       |
| 8 PCS  | SKT            | Kretek      | Lembut       | Manis         | Besar               | Keras           | Ringan        | Kertas Grenjeng | Laris       |
| 21 PCS | SKM            | Filter      | Sangat Wangi | Manis         | Sedang              | Sedang          | Berat         | Kertas Grenjeng | Laris       |

Sumber: Penelitian Mandiri (2024)

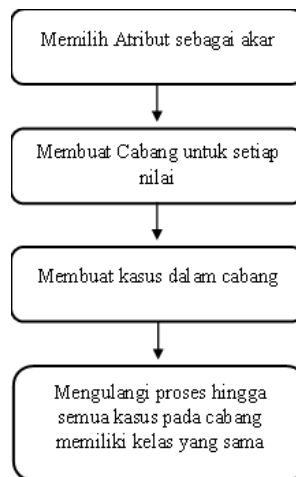
Gambar 3. Dataset yang telah diseleksi

3. Pengolahan Data Menggunakan Algoritma C4.5

Untuk mengolah data menggunakan Algoritma C4.5 terdapat dua tahap yang harus diselesaikan untuk mendapatkan hasil data yang baik sebagai berikut:

a. Melakukan Perhitungan menggunakan metode Decision Tree

Secara umum, proses membangun Decision Tree dilakukan melalui beberapa langkah sebagai berikut:



Sumber: Penelitian Mandiri (2024)

Gambar 4. Proses Decision Tree

- 1) Pilih atribut sebagai akar.
- 2) Buat cabang untuk tiap-tiap nilai.
- 3) Buat kasus dalam cabang
- 4) Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk menghitung *gain* digunakan persamaan sebagai berikut :

$$Gain(S, A) = Entropy(S) - \sum_i^n = 1 \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (1)$$

Keterangan :

- S = Himpunan kasus
- A = Jumlah Partisi A
- |S<sub>i</sub>| = Proporsi S<sub>i</sub> terhadap S
- |S| = Jumlah kasus dalam S

Sedangkan untuk menghitung nilai *entropy* digunakan persamaan sebagai berikut :

$$Entropy(S) = - \sum_i^n = 1 - p_i * \log_2 \quad (2)$$

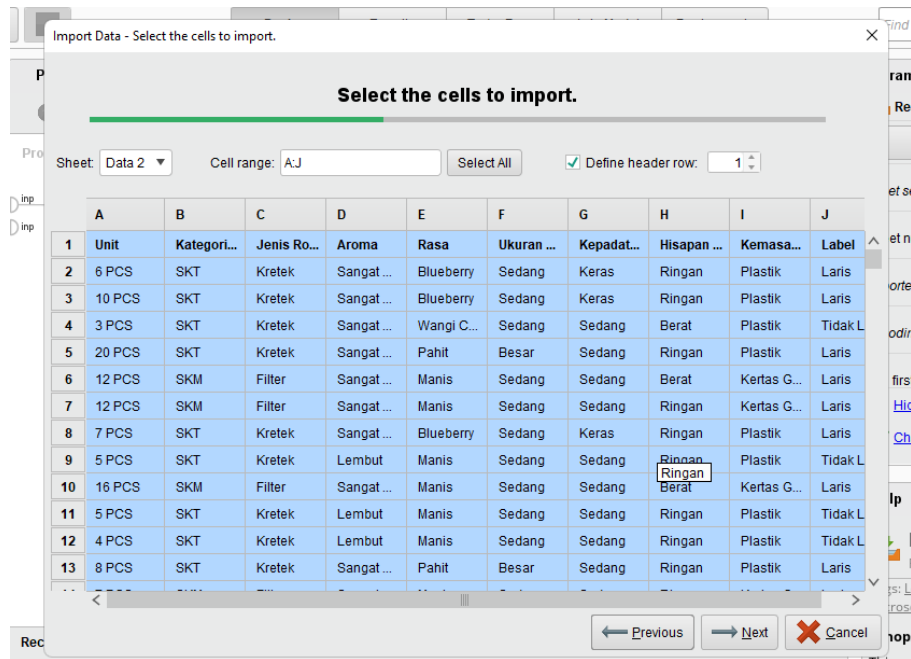
Keterangan

- S = Himpunan kasus
- n = Jumlah partisi A
- p<sub>i</sub> = proporsi S<sub>i</sub> terhadap S

b. Pengolahan Data Dengan *Decision Tree* dan Algoritma C4.5 Menggunakan Rapidminer

Berikut adalah model *Decision Tree* yang telah dibuat dengan menganalisis data menggunakan aplikasi Rapidminer:

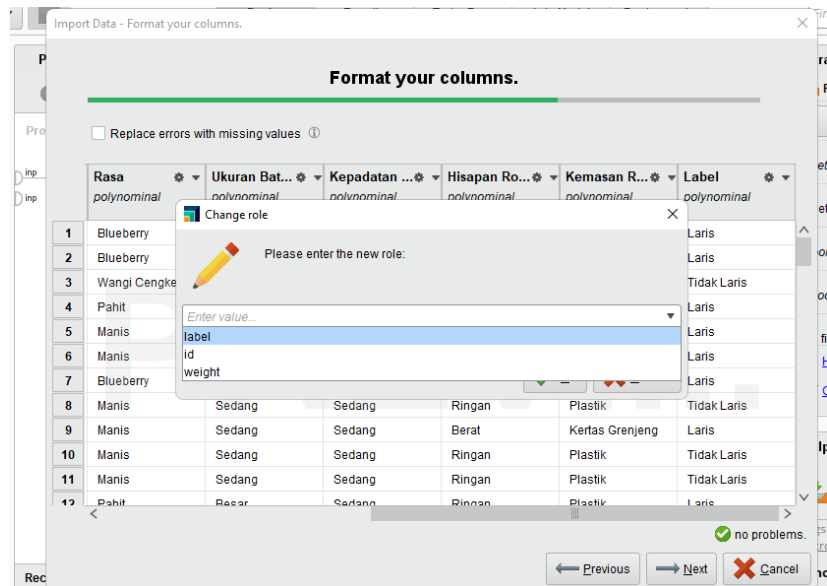
- 1) Pertama memasukan dataset untuk pengolahan Decision Tree.



Sumber: Penelitian Mandiri (2024)

Gambar 5. Dataset Penjualan Rokok

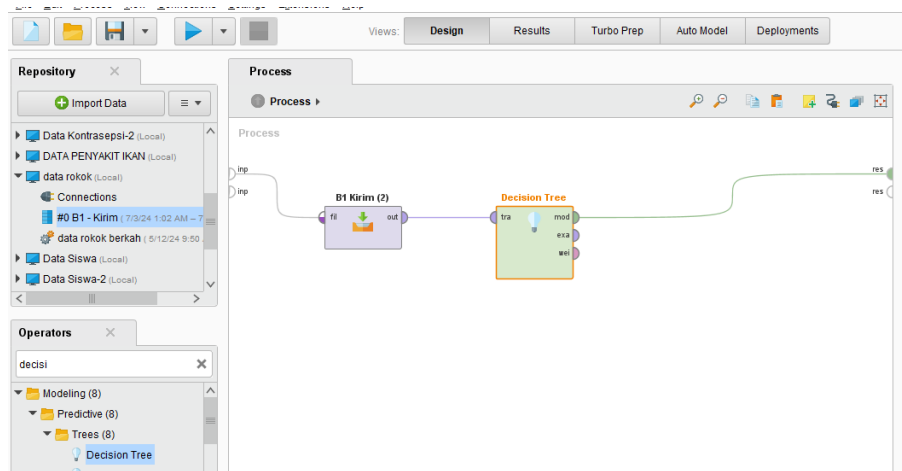
2) Kemudian menentukan *type* dan label yang digunakan pada aplikasi rapidminer.



Sumber: Penelitian Mandiri (2024)

Gambar 6. Penambahan Label Pada Atribut Label

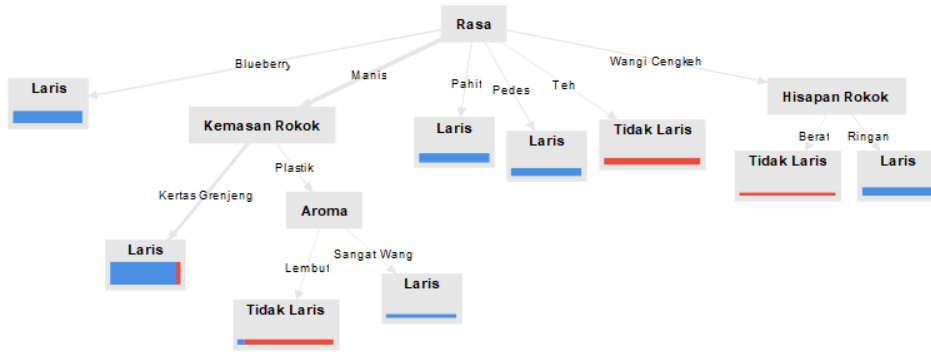
3) Selanjutnya memasukkan operator *Decision Tree* dengan mengetik di kolom *search for operator* untuk mengolah dataset menjadi model *Decision Tree*.



Sumber: Penelitian Mandiri (2024)

Gambar 7. Pengolahan *Decision Tree* Menggunakan Rapidminer

4) Berikut adalah hasil *Decision Tree* yang dibuat dengan menganalisis data menggunakan aplikasi *rapidminer*.



Sumber: Penelitian Mandiri (2024)

**Gambar 8.** Hasil *Decision Tree* Pada Rapidminer

Dari pohon keputusan pada gambar 8 didapatkan Rule untuk klasifikasi penjualan laris. Rule yang sebagai berikut :

```

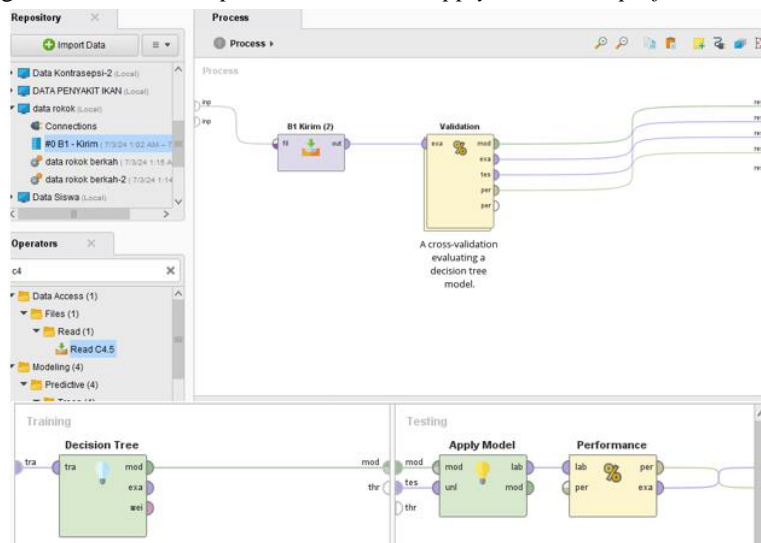
Tree

Rasa = Blueberry: Laris {Laris=34, Tidak Laris=0}
Rasa = Manis
|   Kemasan Rokok = Kertas Grenjeng: Laris {Laris=62, Tidak Laris=4}
|   Kemasan Rokok = Plastik
|   |   Aroma = Lembut: Tidak Laris {Laris=1, Tidak Laris=12}
|   |   Aroma = Sangat Wangi: Laris {Laris=8, Tidak Laris=0}
Rasa = Pahit: Laris {Laris=25, Tidak Laris=0}
Rasa = Pedes: Laris {Laris=21, Tidak Laris=0}
Rasa = Teh: Tidak Laris {Laris=0, Tidak Laris=17}
Rasa = Wangi Cengkeh
|   Hisapan Rokok = Berat: Tidak Laris {Laris=0, Tidak Laris=6}
|   Hisapan Rokok = Ringan: Laris {Laris=22, Tidak Laris=0}
    
```

Sumber: Penelitian Mandiri (2024)

**Gambar 9.** Rule hasil algoritma *Decision Tree*

5) Untuk mengetahui akurasi data dan menampilkan *Area Under Curve (AUC)* menggunakan aplikasi *rapidminer* dengan menambahkan *operator validation, apply model, dan performance*.

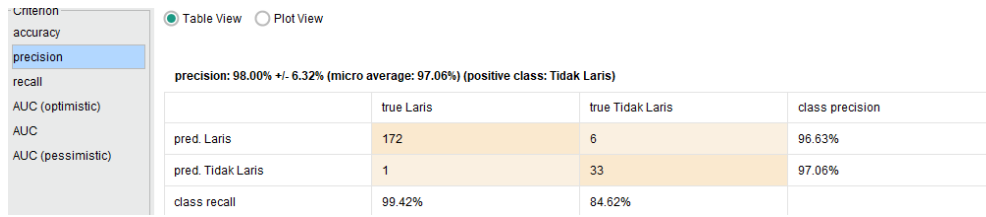


Sumber: Penelitian Mandiri (2024)

**Gambar 10.** Proses Pengolahan Akurasi Data pada Rapidminer

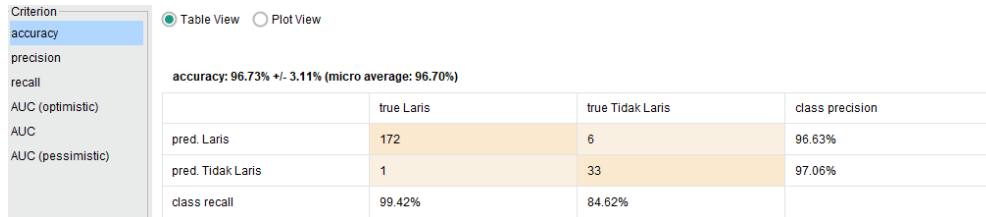


6) Berikut adalah hasil akurasi data dan *Area Under Curve* (AUC).



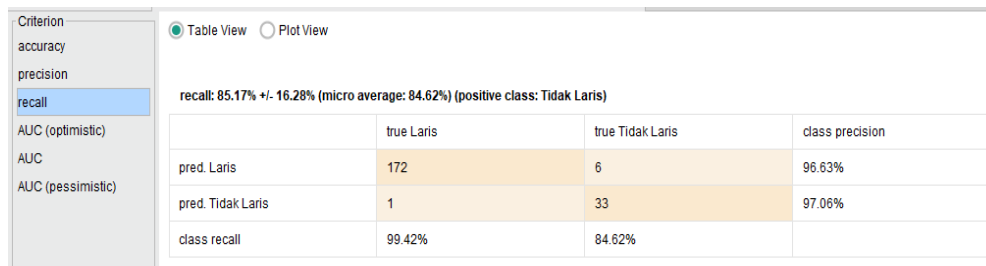
Sumber: Penelitian Mandiri (2024)

**Gambar 11.** Nilai *Accuracy* pada *Decision Tree*



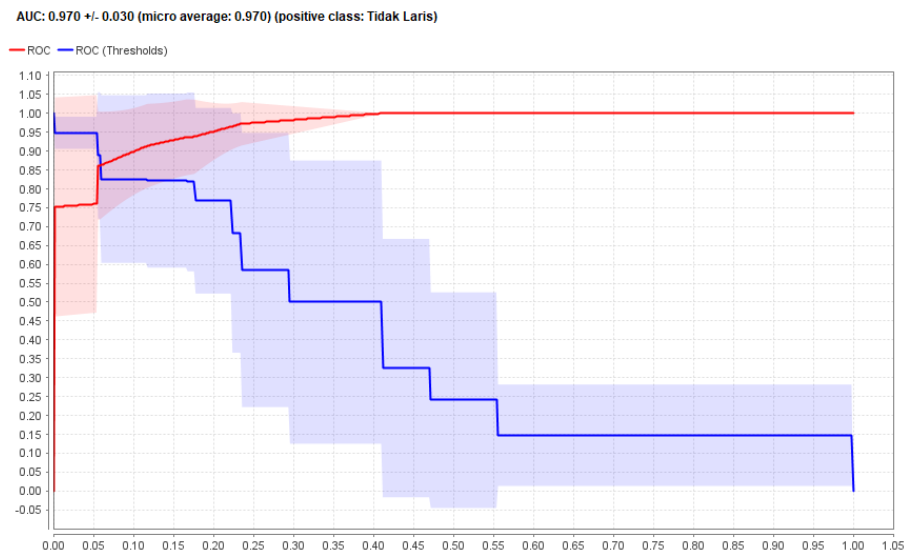
Sumber: Penelitian Mandiri (2024)

**Gambar 12.** Nilai *Precision* pada *Decision Tree*



Sumber: Penelitian Mandiri (2024)

**Gambar 13.** Nilai *Recall* pada *Decision Tree*



Sumber: Penelitian Mandiri (2024)

**Gambar 14.** Nilai *Area Under Curve* (AUC) pada *Decision Tree*

## KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma C4.5 efektif dalam menganalisis dan mengklasifikasikan data penjualan rokok, menghasilkan model yang mampu memprediksi penjualan rokok terlaris dengan akurasi yang baik. Dengan menganalisis data penjualan rokok, penelitian ini dapat membantu perusahaan CV Jaya Berkah Mas untuk memahami pola dan tren penjualan, memungkinkan mereka untuk merancang strategi pemasaran yang lebih efektif. Penelitian ini juga menunjukkan pentingnya proses pembersihan data dalam memastikan kualitas analisis data mining. Dari hasil penelitian ini juga didapatkan akurasi yang lebih tinggi dari penelitian sebelumnya, dengan menggunakan metode Algoritma C4.5 dan data sebanyak 212 *entry* menunjukkan tingkat Akurasi 96.73%, Presisi 98.00%, dan *Recall* 85.17%.

## REFERENSI

- Anggi Trifani, Agus Perdana Windarto, & Hendry Qurniawan. (2022). Penerapan Data Mining Klasifikasi C4.5 dalam Menentukan Tingkat Stres Mahasiswa Akhir. *Jurnal Riset Rumpun Ilmu Teknik*, 1(2), 92. <https://doi.org/10.55606/jurritek.v1i2.414>
- Arhami, M., & Nasir, M. (2020). *Data Mining.pdf* (I. R. Utami (ed.); 1st ed.). ANDI (IKAPI).
- Fauziningrum, E., & Suryaningsih, E. I. (2021). *Penerapan Data Mining Metode Decision Tree Untuk Mengukur Penguasaan Bahasa Inggris Maritim*. CV. PUSTAKA STIMAR AMNI SEMARANG.
- Leonardi, M., Emilda, R., Katrin, I., & Yulianto, A. (2021). Prediksi Penjualan Produk Rokok Pada PT. Indomarco Prisma Utama Menggunakan Algoritma C4.5. *Paradigma - Jurnal Komputer Dan Informatika*, 23(2), 182–190. <https://doi.org/10.31294/p.v23i2.11151>
- Musa, D. M., Sakti, D., Shantiony, K. A., Zega, S. K. P., Hamzah, S., Zega, Y. J., & Lubis, B. O. (2024). Penerapan Data Mining Untuk Klasifikasi Data Penjualan Pakan Ternak Terlaris Dengan Algoritma C4.5. *Jurnal Teknologi Informatika Dan Komputer*, 10(1), 168–182. <https://doi.org/10.37012/jtik.v10i1.1985>
- Ovan, & Saputra, A. (2020). *Aplikasi Uji Validitas dan Reliabilitas Instrumen Penelitian Berbasis Web.pdf*. Yayasan Ahmar Cendekia Indonesia.
- Pratiwi, D. A., Awangga, R. M., & Setyawan, M. Y. H. (2020). Seleksi Calon Kelulusan Tepat Waktu Mahasiswa Teknik Informatika Menggunakan Metode Naive Bayes. In R. M. Awangga (Ed.), *Kreatif* (1st ed.). Kreatif Industri Nusantara.
- Shaumi, A. R., Ali, M. F., & Simbolon, M. T. A. M. (2022). Penerapan Data Mining Menggunakan Metode Teknik Classification Untuk Melihat Penerapan Data Mining Menggunakan Metode Teknik Classification Untuk. *JUKI : Jurnal Komputer Dan Informatika*, 4(2), 174. [www.pajak.go.id](http://www.pajak.go.id)
- Sugara, B., Adidarma, D., & Budilaksono, S. (2019). Perbandingan Akurasi Algoritma C4.5 dan Naive Bayes untuk Deteksi Dini Gangguan Autisme pada Anak. *Jurnal IKRA-ITH Informatika*, 3(1), 122.
- Suriani, U. (2023). Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5. *Journalcisa*, 3(2), 56. <https://doi.org/10.51519/journalcisa.v4i2.393>
- Triawan, B., Lubis, I., & Kadim, Lina, Arliana, N. (2024). *Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper*. 3(November), 112–120.
- Tukino, T. (2019). Penerapan Algoritma C4.5 Untuk Memprediksi Keuntungan Pada PT SMOE Indonesia. *Jurnal Sistem Informasi Bisnis*, 9(1), 41. <https://doi.org/10.21456/vol9iss1pp39-46>
- Umrati, & Wijaya, H. (2020). *Analisis Data Kualitatif Teori Konsep dalam Penelitian Pendidikan.pdf*.
- Wahyudi, M. D. (2023). Penerapan Data Mining Dengan Algoritma C4. 5 Dalam Prediksi Penjualan Buku. *Jurnal Teknorama (Informatika Dan ...)*, 1(1), 1–6. <https://jurnal.stikomelrahma.ac.id/index.php/teknorama/article/view/1%0Ahttps://jurnal.stikomelrahma.ac.id/index.php/teknorama/article/download/1/1>
- Zai, C. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Jurnal Teknoinfo*, 2(1), 6. <http://portaldata.org/index.php/portaldata/article/view/107>