

## PENERAPAN DATA MINING UNTUK KLASIFIKASI DATA PENJUALAN SEMBAKO TERLARIS DENGAN ALGORITMA C4.5

Mikaela Alta Melfia<sup>1</sup>, Keisha Athaya Ramadhini<sup>2</sup>, Muhammad Ajmal Maulana<sup>3</sup>,  
Arya Laura Dasiva<sup>4</sup>, Andika Aria Briantoro<sup>5</sup>, Baginda Oloan Lubis<sup>6</sup>

<sup>1,2,3,4,5</sup> Sistem Informasi, Bina Sarana Informatika

<sup>6</sup> Rekayasa Perangkat Lunak, Bina Sarana Informatika

Jl. Kramat Raya No. 98 Jakarta, Indonesia

19221202@bsi.ac.id

### ABSTRAK

Permasalahan dalam penjualan sembako yang paling laris meliputi ketidakstabilan permintaan akibat faktor musim, tren pasar, dan kondisi ekonomi. Pengelolaan stok menjadi tantangan karena permintaan tinggi dapat menyebabkan stok habis, sementara kelebihan stok meningkatkan risiko kerugian. Penentuan harga juga krusial dalam menjaga daya saing dan daya beli konsumen. Selain itu, distribusi sering menghadapi hambatan logistik, serta perubahan preferensi konsumen terhadap produk tertentu turut memengaruhi pola penjualan. Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menerapkan dan mengevaluasi metode data mining menggunakan algoritma C4.5 dalam mengklasifikasikan produk sembako terlaris. Peningkatan volume data penjualan menuntut analisis yang lebih efektif dalam menentukan produk paling diminati konsumen. Algoritma C4.5 dipilih karena kemampuannya menangani atribut kontinu, mengatasi missing values, dan menghasilkan pohon keputusan yang dapat diinterpretasikan dengan baik. Penelitian ini mencakup pengumpulan data dari beberapa toko selama satu tahun, praproses data, penerapan algoritma C4.5, dan evaluasi model menggunakan k-fold cross-validation. Data yang digunakan mencakup atribut jenis produk, harga, jumlah penjualan, dan waktu penjualan. Hasil penelitian menunjukkan bahwa algoritma C4.5 berhasil mengidentifikasi pola penjualan dengan akurasi klasifikasi mencapai 85%. Pohon keputusan yang dihasilkan memberikan wawasan bagi manajemen dalam optimalisasi stok dan strategi pemasaran. Temuan ini diharapkan dapat membantu pelaku usaha meningkatkan efisiensi operasional dan memenuhi kebutuhan konsumen dengan lebih tepat.

**Kata kunci :** Penjualan Sembako, Klasifikasi, Algoritma C4.5, Data Mining

### 1. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong peningkatan penggunaan data dalam pengambilan keputusan bisnis, termasuk dalam sektor perdagangan sembako. Dengan semakin besarnya volume data penjualan yang dihasilkan setiap hari, diperlukan metode analisis yang efektif untuk mengidentifikasi produk yang paling diminati konsumen. Informasi ini sangat penting bagi pelaku usaha dalam mengoptimalkan strategi pemasaran, pengelolaan stok, serta penentuan harga agar tetap kompetitif di pasar.

Namun, salah satu tantangan utama dalam manajemen penjualan sembako adalah ketidakstabilan permintaan yang dipengaruhi oleh faktor musim, tren pasar, harga bahan baku, dan perilaku konsumen. Kesalahan dalam peramalan produk terlaris dapat menyebabkan kehabisan stok atau kelebihan persediaan, yang berakibat pada kerugian bisnis. Oleh karena itu, diperlukan pendekatan berbasis data untuk mengatasi permasalahan tersebut secara lebih akurat dan sistematis.

Metode data mining menjadi salah satu solusi yang efektif dalam menganalisis data penjualan untuk menemukan pola tersembunyi dan membuat prediksi berbasis historis. Algoritma C4.5 dipilih dalam penelitian ini karena kemampuannya dalam membangun model klasifikasi yang dapat menangani atribut kontinu, menangani missing values, serta

menghasilkan pohon keputusan yang mudah dipahami.[1]

Dengan menerapkan algoritma ini, diharapkan dapat dihasilkan model klasifikasi yang mampu mengidentifikasi produk sembako terlaris berdasarkan pola penjualan sebelumnya. Pengertian penjualan secara umum adalah kegiatan jual beli dijalankan oleh dua belah pihak atau lebih dengan alat pembayaran yang sah. Tujuan yang utama tentunya mendatangkan keuntungan dari produk atau barang yang dijual.[2]

Perkembangan persaingan di bidang bisnis disertai kemajuan di bidang teknologi informasi mengantarkan perusahaan-perusahaan ritel saat ini, termasuk di Indonesia, pada persaingan yang ketat dan lebih terbuka. Setiap perusahaan dituntut untuk mengambil keputusan secara tepat dalam strategi pemasaran dengan melihat kondisi pasar.[3]

Dalam konteks penjualan sembako, klasifikasi produk terlaris menjadi kebutuhan yang mendesak untuk membantu bisnis memahami preferensi pelanggan. Dengan mengetahui kategori produk yang paling laris, perusahaan dapat mengelola stok, merancang strategi promosi, dan meningkatkan efisiensi operasional. Namun, banyak perusahaan masih menghadapi kesulitan dalam mengolah data penjualan secara manual karena kerumitan dan volume datanya.

Penelitian ini menggunakan algoritma C4.5, yang merupakan salah satu metode data mining berbasis

pohon keputusan. Algoritma ini mampu mengklasifikasikan data penjualan sembako menjadi kategori produk yang laris atau tidak laris berdasarkan atribut seperti harga, kuantitas, dan kategori produk.[4]

Proses ini melibatkan penghitungan nilai entropy dan gain untuk menentukan atribut yang paling signifikan dalam klasifikasi. [5]

Tujuan utama penelitian ini adalah untuk menghasilkan pohon keputusan yang dapat digunakan untuk mengidentifikasi produk sembako terlaris dengan tingkat akurasi yang tinggi. [6]

Hasil dari penelitian ini diharapkan tidak hanya membantu dalam memahami pola penjualan, tetapi juga memberikan dasar yang kuat untuk pengambilan keputusan strategis dalam pengelolaan bisnis sembako.

## 2. TINJAUAN PUSTAKA

Penelitian ini membahas tentang penerapan data mining untuk mengklasifikasikan data penjualan sembako terlaris menggunakan algoritma C4.5. Penelitian ini tidak terlepas dari hasil penelitian-penelitian terdahulu, yang pernah dilakukan sebagai bahan perbandingan dan kajian lebih lanjut. Adapun penelitian terdahulu yang diambil oleh peneliti.

Berdasarkan hasil dari penelitian yang dilakukan oleh [7] yang di mana melakukan penelitian mengenai Penerapan Data Mining Untuk Klasifikasi Data Penjualan Pakan Ternak Terlaris dengan menggunakan Algoritma C4.5 pada PT. Fresh X. Penelitian ini berfokus pada penerapan algoritma C4.5 untuk mengklasifikasikan data penjualan pakan ternak di PT Fresh X, yang bertujuan mengidentifikasi produk yang paling laris. Proses klasifikasi dilakukan untuk menghasilkan pola yang dapat mendukung pengambilan keputusan strategis dalam bisnis. Dalam penelitian tersebut, analisis data penjualan telah digunakan untuk mengidentifikasi produk yang paling diminati oleh konsumen berdasarkan atribut-atribut seperti kategori, harga, dan kuantitas. Teknik klasifikasi berbasis pohon keputusan telah terbukti membantu dalam merancang strategi pemasaran dan pengelolaan stok. Penelitian ini menerapkan algoritma C4.5, yang menghasilkan pohon keputusan berdasarkan nilai entropy dan gain dari setiap atribut. Penemuan gain tertinggi pada atribut tertentu memberikan wawasan tentang faktor utama yang memengaruhi penjualan, sebagaimana hasil penelitian yang menunjukkan bahwa pakan ayam pedaging merupakan produk paling laris. Melalui tahapan seperti pengolahan data, transformasi, penerapan algoritma, dan visualisasi dalam bentuk pohon keputusan, penelitian ini memberikan hasil yang relevan untuk mendukung pengelolaan bisnis. Hasil klasifikasi ini bermanfaat dalam memahami karakteristik produk yang laris dan memberikan dasar yang kuat untuk strategi bisnis yang lebih efektif.[8]

### 2.1. Klasifikasi

Klasifikasi adalah proses identifikasi dan pengelompokan suatu objek ke dalam kelompok atau kategori yang sama. Klasifikasi dapat digunakan untuk mengelompokkan suatu dataset yang berukuran besar, beberapa contoh metode klasifikasi yang umum digunakan adalah CART (Classification And Regression Tree) dan Naïve Bayes.[9]

### 2.2. Algoritma C4.5 dalam Data Mining

Data Mining adalah proses penggalian pola atau informasi berharga dari kumpulan data besar. [7] Dalam kasus ini, algoritma C4.5 diterapkan untuk menganalisis potensi pengaruh buruk gelombang elektromagnetik yang dihasilkan oleh telepon genggam terhadap otak. Fokusnya adalah mengidentifikasi faktor-faktor yang paling berpengaruh serta membuat model prediksi. [10]

### 2.3. Klasifikasi dengan Algoritma C4.5

Klasifikasi merupakan salah satu fungsi utama dalam data mining yang bertujuan untuk mengelompokkan data ke dalam kelas-kelas tertentu.[11] Algoritma C4.5 beroperasi dengan membagi data berdasarkan information gain tertinggi, yang memudahkan dalam memilih atribut terbaik untuk setiap pembagian dalam pohon keputusan. [12] Metode ini telah banyak diterapkan pada berbagai jenis data untuk meningkatkan akurasi klasifikasi dan mendukung pengambilan keputusan yang berbasis data. [13]

### 2.4. Pemanfaatan Entropy dan Information Gain

Entropy dan Information Gain adalah konsep kunci dalam algoritma C4.5 untuk menentukan atribut terbaik dalam membagi dataset.[5]

Entropy mengukur tingkat ketidakpastian dalam dataset, sementara Information Gain berfungsi untuk mengurangi ketidakpastian tersebut dengan memilih atribut yang memberikan pemisahan terbaik. [14]

Gabungan kedua konsep ini memastikan bahwa pohon keputusan yang dihasilkan menjadi lebih efisien dan akurat dalam proses klasifikasi.[15]

### 2.5. Menghitung Entropy untuk Atribut Utama

Entropy adalah ukuran ketidakpastian atau kekacauan dalam data. Dalam konteks pengambilan keputusan, terutama dalam pembelajaran mesin, seperti pohon keputusan, entropy digunakan untuk mengevaluasi tingkat impurity (ketidakmurnian) suatu dataset berdasarkan distribusi label target.

Secara matematis, entropy  $E(S)$  untuk dataset  $S$  dapat dihitung menggunakan rumus berikut:

$$E(S) = - \sum_{i=1}^n p_i \cdot \log_2(p_i) \quad (1)$$

Penjelasan:

$S$  : himpunan kasus

$N$  : jumlah partisi  $S$

$P_i$  : proporsi dari  $S_i$  terhadap  $S$

(-) : Digunakan karena logaritma probabilitas selalu negatif

## 2.6. Menghitung Information Gain (G)

Gain (atau Information Gain) adalah ukuran yang digunakan untuk menentukan efektivitas suatu atribut dalam memisahkan dataset menjadi subset yang lebih homogen (lebih murni). Dengan kata lain, Gain mengukur pengurangan Entropy setelah data dipisahkan berdasarkan suatu atribut tertentu.

$$\text{GAIN}(S, A) = E(S) - \sum_{v \in \text{values}(A)} \frac{|S_v|}{|S|} \cdot E(S_v) \quad (2)$$

Penjelasan:

S : himpunan data awal

A : atribut yang di gunakan untuk membagi data

Values: nilai-nilai unik dari atribut A

|Sv| : subset data S yang memiliki nilai v pada atribut A

$\frac{|S_v|}{|S|}$  : proporsi data dalam subset Sv terhadap total S

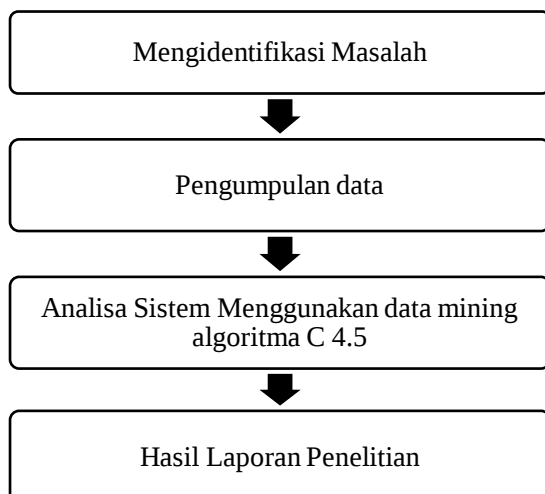
E(Sv) : entropy dari subset Sv

## 2.7. Decision Tree

Decision Tree merupakan algoritma yang termasuk dalam supervised learning. Dikatakan supervised learning karena pada proses pengelompokkan algoritma ini menggunakan data awal untuk membentuk aturan keputusan. Sehingga, data yang telah dikumpulkan harus sudah memiliki label kelompok.[16]

## 3. METODE PENELITIAN

Untuk memastikan penelitian berjalan dengan terencana dan sistematis, peneliti menyusun kerangka sebagai dasar dalam melaksanakan proses penelitian, yang digambarkan melalui flowchart pada Gambar 1 berikut ini.



Gambar 1. Flow Chat Penelitian

Berikut adalah penjelasan untuk setiap langkah yang terdapat pada diagram tersebut:

### 3.1. Mengidentifikasi Masalah

Tahap ini bertujuan untuk memahami dan merumuskan masalah yang akan diselesaikan dalam penelitian atau proyek. Proses ini melibatkan identifikasi kebutuhan, hambatan, atau tantangan yang ada, baik dari aspek teknis maupun non-teknis. Dalam penelitian data mining, identifikasi masalah mencakup pemahaman tentang data, tujuan analisis, dan hasil yang diharapkan.

### 3.2. Pengumpulan Data

Tahap ini melibatkan pengumpulan data yang relevan untuk menyelesaikan masalah yang telah diidentifikasi. Data dapat dikumpulkan dari berbagai sumber seperti survei, database, atau dokumen lain. Penting untuk memastikan bahwa data yang diperoleh berkualitas, lengkap, dan relevan untuk analisis lebih lanjut. Dalam data mining, data yang terkumpul biasanya diolah terlebih dahulu agar siap digunakan untuk algoritma.[17]

### 3.3. Analisa Sistem Menggunakan Data Mining Algoritma C 4.5

Pada tahap ini, dilakukan analisis menggunakan algoritma C4.5 (sebuah algoritma pembentukan pohon keputusan) untuk mendapatkan pola atau insight dari data. Analisis ini melibatkan proses pelatihan model dengan data yang telah dikumpulkan, validasi hasil, dan evaluasi kinerja algoritma terhadap data tersebut. Outputnya biasanya berupa pohon keputusan yang dapat digunakan untuk membuat prediksi atau klasifikasi.

### 3.4. Hasil Laporan Penelitian

Setelah proses analisis selesai, hasil penelitian dirangkum dalam sebuah laporan. Laporan ini mencakup semua tahapan yang telah dilakukan, mulai dari identifikasi masalah, pengumpulan data, hingga hasil analisis. Laporan juga menyajikan temuan, interpretasi pola yang dihasilkan oleh algoritma, dan rekomendasi berdasarkan hasil analisis.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Pengolahan Data

Berikut adalah data yang digunakan dalam penelitian ini:

Tabel 1. Data Penjualan Sembako

| No | Jenis Barang                | Harga     | Kuantitas |     |     | Harga Total |           |           |
|----|-----------------------------|-----------|-----------|-----|-----|-------------|-----------|-----------|
|    |                             |           | Okt       | Nov | Des | Oktober     | November  | Desember  |
| 1  | Tepung Terigu 1 kg          | Rp11.700  | 32        | 31  | 24  | 374.000     | 362.700   | 280.800   |
| 2  | Sunco Minyak Goreng 1 Liter | Rp25.900  | 27        | 20  | 26  | 699.300     | 518.000   | 673.400   |
| 3  | Sania Gula Pasir 1 kg       | Rp14.300  | 75        | 68  | 45  | 1.072.500   | 972.400   | 643.500   |
| 4  | Garam Meja 250 gr           | Rp4.600   | 17        | 10  | 15  | 78.200      | 46.000    | 69.000    |
| 5  | Beras Raos 20 kg            | Rp318.100 | 10        | 8   | 15  | 3.181.000   | 2.544.800 | 4.771.500 |

| No | Jenis Barang                      | Harga    | Kuantitas |     |     | Harga Total |           |           |
|----|-----------------------------------|----------|-----------|-----|-----|-------------|-----------|-----------|
|    |                                   |          | Okt       | Nov | Des | Oktober     | November  | Desember  |
| 6  | Kecap Bango 135 ml                | Rp9.000  | 20        | 13  | 11  | 180.000     | 117.000   | 99.000    |
| 7  | ABC Saus Sambal 335 ml            | Rp19.700 | 15        | 23  | 28  | 295.500     | 453.100   | 551.600   |
| 8  | Indomie Goreng Instant 85 gr      | Rp3.500  | 68        | 53  | 70  | 238.000     | 185.500   | 245.000   |
| 9  | Gaga Sarden 155 gr                | Rp11.300 | 25        | 33  | 34  | 282.500     | 372.900   | 384.200   |
| 10 | Mie Telur Cap Ayam 200 gr         | Rp5.000  | 16        | 10  | 17  | 80.000      | 50.000    | 85.000    |
| 11 | Indomilk Susu Kental Manis 370 gr | Rp11.500 | 23        | 12  | 24  | 264.500     | 138.000   | 276.000   |
| 12 | Gas Elpiji 3 kg                   | Rp22.000 | 52        | 23  | 32  | 1.144.000   | 506.000   | 704.000   |
| 13 | Nuvo Sabun Mandi Cair 450 ml      | Rp25.100 | 8         | 10  | 9   | 200.800     | 251.000   | 225.900   |
| 14 | Sprite Botol 1 liter              | Rp15.000 | 25        | 15  | 34  | 375.000     | 225.000   | 510.000   |
| 15 | Sariwangi Teh Celup isi 25        | Rp7.200  | 29        | 24  | 50  | 208.800     | 172.800   | 360.000   |
| 16 | Jazz One Deterjen 850 gr          | Rp23.400 | 13        | 24  | 43  | 304.200     | 561.600   | 1.006.200 |
| 17 | Boneto Susu Bubuk 800 gr          | Rp45.000 | 12        | 32  | 29  | 540.000     | 1.440.000 | 1.305.000 |
| 18 | Kopi Kapal Api 380 gr             | Rp34.500 | 20        | 43  | 57  | 690.000     | 1.483.500 | 1.966.500 |
| 19 | Ekonomi Sabun Cuci Piring 400 ml  | Rp16.000 | 17        | 95  | 79  | 272.000     | 1.520.000 | 1.264.000 |
| 20 | Blue Band Serbaguna 200 gr        | Rp10.000 | 12        | 37  | 65  | 120.000     | 370.000   | 650.000   |

#### 4.2. Transformasi Data

Berikut adalah hasil transformasi data penjualan yang akan diolah:

Tabel 2. Hasil Pengolahan Data

| No | Jenis Barang                      | Harga | Kuantitas | Keterangan  |
|----|-----------------------------------|-------|-----------|-------------|
| 1  | Tepung Terigu 1 kg                | Murah | Banyak    | Laris       |
| 2  | Sunco Minyak Goreng 1 Liter       | Mahal | Banyak    | Laris       |
| 3  | Sania Gula Pasir 1 kg             | Murah | Banyak    | Laris       |
| 4  | Garam Meja 250 gr                 | Murah | Sedikit   | Tidak Laris |
| 5  | Beras Raos 20 kg                  | Mahal | Sedikit   | Tidak Laris |
| 6  | Kecap Bango 135 ml                | Murah | Sedikit   | Tidak Laris |
| 7  | ABC Saus Sambal 335 ml            | Murah | Banyak    | Laris       |
| 8  | Indomie Goreng Instant 85 gr      | Murah | Banyak    | Laris       |
| 9  | Gaga Sarden 155 gr                | Murah | Banyak    | Laris       |
| 10 | Mie Telur Cap Ayam 200 gr         | Murah | Sedikit   | Tidak Laris |
| 11 | Indomilk Susu Kental Manis 370 gr | Murah | Banyak    | Laris       |
| 12 | Gas Elpiji 3 kg                   | Mahal | Sedikit   | Tidak Laris |
| 13 | Nuvo Sabun Mandi Cair 450 ml      | Mahal | Banyak    | Laris       |
| 14 | Sprite Botol 1 liter              | Murah | Banyak    | Laris       |
| 15 | Sariwangi Teh Celup isi 25        | Murah | Banyak    | Laris       |
| 16 | Jazz One Deterjen 850 gr          | Murah | Banyak    | Laris       |
| 17 | Boneto Susu Bubuk 800 gr          | Mahal | Banyak    | Laris       |
| 18 | Kopi Kapal Api 380 gr             | Mahal | Banyak    | Laris       |
| 19 | Ekonomi Sabun Cuci Piring 400 ml  | Murah | Banyak    | Laris       |
| 20 | Blue Band Serbaguna 200 gr        | Murah | Banyak    | Laris       |

#### 4.3. Perhitungan Algoritma C4.5

##### 4.3.1. Mencari Nilai Entropy

1. Entropy Total:

$$\text{Entropy Total} = \left(-\frac{15}{20} \times \log_2 \frac{15}{20}\right) \left(-\frac{5}{20} \times \log_2 \frac{5}{20}\right) = 0.81125$$

2. Entropy Kategori Sembako:

$$\text{Entropy (Tepung Terigu 1kg)} \\ \text{Entropy} \left(-\frac{1}{1} \times \log_2 \frac{1}{1}\right) \left(-\frac{0}{1} \times \log_2 \frac{0}{1}\right) = 0$$

$$\text{Entropy (Garam Meja 20g)}$$

$$\text{Entropy} \left(-\frac{0}{1} \times \log_2 \frac{0}{1}\right) \left(-\frac{0}{1} \times \log_2 \frac{0}{1}\right) = 0$$

3. Entropy Harga:

Entropy (Harga Murah)

$$\text{Entropy} \left(-\frac{11}{14} \times \log_2 \frac{11}{14}\right) \left(-\frac{3}{14} \times \log_2 \frac{3}{14}\right) = 0,7347$$

Entropy (Harga Mahal)

$$\text{Entropy} \left(-\frac{4}{6} \times \log_2 \frac{4}{6}\right) \left(-\frac{2}{6} \times \log_2 \frac{2}{6}\right) = 0,91829$$

4. Entropy Kuantitas

Entropy (Quantity Sedikit)

$$\text{Entropy} \left(-\frac{0}{5} \times \log_2 \frac{0}{5}\right) \left(-\frac{5}{5} \times \log_2 \frac{5}{5}\right) = 0$$

Entropy (Quantity Banyak)

$$\text{Entropy} \left( -\frac{15}{15} \times \log_2 \frac{15}{15} \right) \left( -\frac{0}{15} \times \log_2 \frac{0}{15} \right) = 0$$

#### 5. Mencari Nilai Gain

Gain Kategori Sembako:

$$\text{Gain} = 0.81125 - \left( \frac{5}{20} \times 0 + \frac{1}{20} \times 0 + \frac{1}{20} \times 0 + \frac{20}{20} \times 0 \right) = 0.81125$$

#### a. Gain Harga:

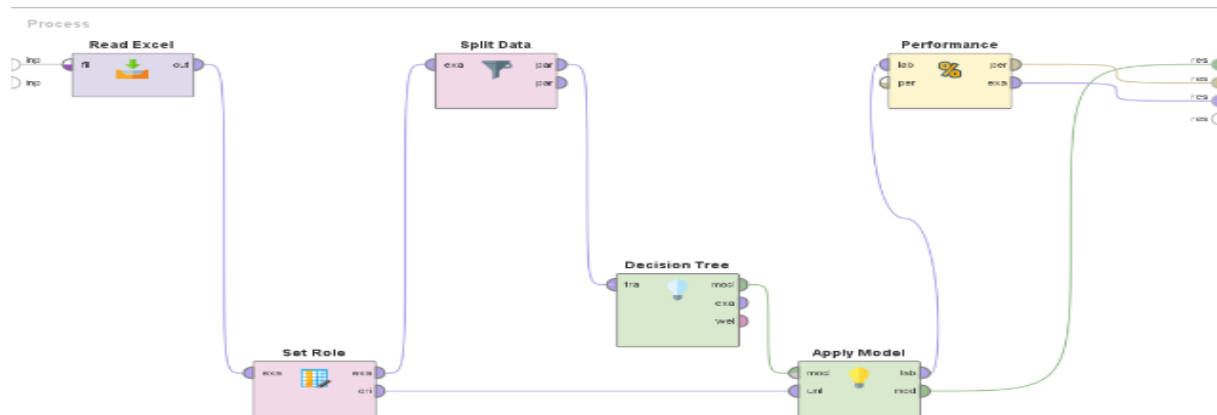
$$\text{Gain} = 0.81125 - \left( \frac{14}{20} \times 0.7347 + \frac{6}{20} \times 0.91829 \right)$$

$$= 0.573447$$

#### b. Gain Kuantitas:

$$\text{Gain} = 0.81125 - \left( \frac{15}{20} \times 0 + \frac{5}{20} \times 0 \right) = 0.81125$$

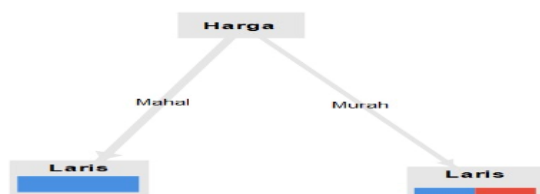
### 4.3.2. Proses dari decision tree



Gambar 2. Process

Gambar 2 menunjukkan alur proses analisis data menggunakan alat seperti RapidMiner. Proses dimulai dengan membaca data dari file Excel melalui blok Read Excel, kemudian data tersebut diproses lebih lanjut dengan blok Set Role untuk menentukan peran atribut, seperti menetapkan variabel target dan fitur prediktor. Setelah itu, data dibagi menjadi dua bagian menggunakan blok Split Data, yang biasanya digunakan untuk membagi data menjadi data latih dan data uji guna memastikan validitas model. Pada langkah berikutnya, model Decision Tree dibangun berdasarkan data latih untuk menghasilkan aturan klasifikasi atau prediksi. Model yang telah dibuat kemudian diterapkan pada data uji menggunakan blok Apply Model guna mendapatkan hasil prediksi. Terakhir, blok Performance digunakan untuk mengevaluasi kinerja model dengan berbagai metrik seperti akurasi, presisi, dan recall. Proses ini secara keseluruhan menunjukkan langkah-langkah dalam membangun, menerapkan, dan mengevaluasi model pembelajaran mesin berbasis pohon keputusan.

### 4.3.3. Pengolahan Data dengan Decision Tree



Gambar 3. Decision Tree

Gambar 3 menunjukkan struktur pohon keputusan yang digunakan untuk menganalisis hubungan antara variabel Harga dan Laris. Pada diagram ini, Harga digunakan sebagai variabel keputusan utama yang menentukan apakah suatu produk akan Laris atau tidak. Jika harga produk Mahal, maka produk tersebut cenderung Laris, seperti ditunjukkan oleh batang biru penuh pada cabang kiri. Sebaliknya, jika harga Murah, maka produk tetap Laris, tetapi dengan distribusi yang lebih bervariasi, sebagaimana ditunjukkan oleh kombinasi batang biru dan merah pada cabang kanan. Warna biru kemungkinan merepresentasikan jumlah produk yang benar-benar laris, sementara merah menunjukkan beberapa kasus di mana produk tidak laris meskipun harganya murah. Model ini mencerminkan bahwa faktor harga memiliki pengaruh yang kompleks terhadap penjualan dan tidak selalu harga murah berarti lebih laris dibanding harga mahal.

## 5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa penerapan data mining dengan algoritma C4.5 dapat digunakan secara efektif untuk mengklasifikasikan data penjualan sembako terlaris. Algoritma ini mampu menghasilkan pohon keputusan yang membantu dalam mengidentifikasi pola penjualan berdasarkan variabel-variabel tertentu, seperti kategori produk, jumlah penjualan, waktu penjualan, dan lokasi. Hasil dari klasifikasi ini memberikan gambaran produk mana yang paling laris, sehingga dapat membantu pelaku usaha dalam pengambilan keputusan strategis, seperti

pengelolaan stok, promosi, dan penentuan harga. Implementasi algoritma C4.5 juga terbukti memiliki tingkat akurasi yang cukup baik dalam proses klasifikasi data penjualan.

Hasil analisis data mining menggunakan metode C4.5 sebaiknya diintegrasikan ke dalam sistem informasi atau aplikasi yang mudah digunakan, sehingga pelaku usaha dapat mengakses data dan laporan secara real-time. Selain itu, perhatian khusus harus diberikan pada keamanan dan privasi data pelanggan agar penggunaan teknologi ini sesuai dengan regulasi yang berlaku. Pemanfaatan teknologi data mining perlu dikelola dengan baik, mulai dari pengumpulan hingga penyimpanan, agar kualitasnya terjamin.

#### DAFTAR PUSTAKA

- [1] J. I. Sianturi, G. Taufik, B. O. Lubis, and D. Laraswati, "Technology Acceptance Model Untuk Analisis Kepuasan Pengguna Terhadap Dompot Digital Dana Di Kota Bekasi," *JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 6, pp. 12794–12800, 2024.*
- [2] Fatawa Imam Al Muftin and Fendi Hidayat, "Sistem Informasi Penjualan," *Zo. Komput. Progr. Stud. Sist. Inf. Univ. Batam*, vol. 13, no. 3, 2024, doi: 10.37776/zkomp.v13i3.1461.
- [3] P. W. Sari, G. Giwangga, N. D. Fitriyanti, A. A. Putra, and H. Irawan, "Implementasi Algoritma C4.5 Untuk Klasifikasi Data Insentif Karyawan PT. Adhi Cakra Utama Mulia," *J. Komput. Antart.*, vol. 2, no. 3, pp. 95–101, 2024, doi: 10.70052/jka.v2i3.318.
- [4] S. Dwiasnati, W. Gunawan, R. R. Oprasto, B. O. Lubis, and B. Santoso, *Algoritma dan Pemrograman Implementasi Menggunakan Python*. Bandar Lampung: CV. Keranjang Teknologi Media, 2023.
- [5] D. Jatmika and G. Taufik, "Penerapan Algoritma C4.5 Untuk Klasifikasi Keberhasilan Pengiriman Barang," *Invotek Polbeng*, vol. 8, no. 1, pp. 12–26, 2021.
- [6] A. Salim and B. O. Lubis, "Sistem Pendukung Keputusan Pemilihan Jenis Tablet Gaming dengan Menggunakan Analytical Hierarchy Process," *Pros. Semin. Nas. Energi Telekomun. dan Otomasi*, no. ISBN: 978-602-74127-4-3, pp. 1–9, 2017.
- [7] D. M. Musa *et al.*, "Penerapan Data Mining Untuk Klasifikasi Data Penjualan Pakan Ternak Terlaris Dengan Algoritma C4.5," *J. Teknol. Inform. dan Komput.*, vol. 10, no. 1, pp. 168–182, 2024, doi: 10.37012/jtik.v10i1.1985.
- [8] B. O. Lubis, D. Oscar, F. W. Fibriany, B. Santoso, Jefa, and A. Rusman, "Classification of tomato leaf disease and combination extraction features using K-NN algorithm," *AIP Conf. Proc.*, vol. 2714, no. November 2021, 2023, doi: 10.1063/5.0128539.
- [9] R. R. Adhitya, Wina Witanti, and Rezki Yuniarti, "Perbandingan Metode Cart Dan Naïve Bayes Untuk Klasifikasi Customer Churn," *INFOTECH J.*, vol. 9, no. 2, pp. 307–318, 2023, doi: 10.31949/infotech.v9i2.5641.
- [10] L. Bachtiar and M. Mahradianur, "Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai," *J. Inform.*, vol. 10, no. 1, pp. 28–36, 2023, doi: 10.31294/inf.v10i1.15115.
- [11] C. I. Ifaru *et al.*, "Penentuan Sparepart Genset Paling Sering Digunakan Pada Operator Indosat Ooredoo Hutchison Dengan Algoritma C4.5," *J. Teknol. Sist. Inf.*, vol. 5, no. 1, pp. 300–309, 2024, doi: 10.35957/jtsi.v5i1.7481.
- [12] T. Tukino, "Penerapan Algoritma C4.5 Untuk Memprediksi Keuntungan Pada PT SMOE Indonesia," *J. Sist. Inf. Bisnis*, vol. 9, no. 1, p. 39, 2019, doi: 10.21456/vol9iss1pp39-46.
- [13] P. R. Sihombing and A. M. Arsani, "Comparison of Machine Learning Methods in Classifying Poverty in Indonesia in 2018," *J. Tek. Inform.*, vol. 2, no. 1, pp. 51–56, 2021, doi: 10.20884/1.jutif.2021.2.1.52.
- [14] S. Faisal, "Klasifikasi Data Mining Menggunakan Algoritma C4.5 Terhadap Kepuasan Pelanggan Sewa Kamera Cikarang," *Techno Xplore J. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2019, doi: 10.36805/technoxplore.v4i1.541.
- [15] S. Setiawansyah, "Penerapan Metode Entropy dan Grey Relational Analysis dalam Evaluasi Kinerja Karyawan," *J. Data Sci. Inf. Syst.*, vol. 2, no. 1, pp. 29–39, 2024.
- [16] A. I. Rizmayanti, N. Hidayati, F. S. Nugraha, and W. Gata, "Penerapan Data Mining Untuk Memprediksi Kompetensi Siswa Menggunakan Metode Decision Tree (Studi Kasus Smk Multicomp Depok)," *Swabumi*, vol. 9, no. 1, pp. 9–18, 2021, doi: 10.31294/swabumi.v9i1.8363.
- [17] Sodik, A. M. B. Aji, B. O. Lubis, and M. Susanti, "Evaluasi Usability Website LPPI Dengan Pendekatan Metode Website Usability Evaluation Tool (WEBUSE)," *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 7, no. 2, pp. 269–279, 2023, doi: 10.52362/jisamar.v7i2.1044.