

## Prediksi Penjualan Pada Toko YTI Menggunakan Metode Decision Tree

M. Chabib Nurroziqin<sup>1\*</sup>, Retno Wardhani<sup>2</sup>, M. Rosidi Zamroni<sup>2</sup>

<sup>1,2,3</sup>Teknik Informatika/Fakultas Sains dan Teknologi/Universitas Islam Lamongan

Email: [muhammadchabibnurroziqin@gmail.com](mailto:muhammadchabibnurroziqin@gmail.com)

Email: [retzno.teknik@unisla.ac.id](mailto:retzno.teknik@unisla.ac.id)

Email: [rosidizamroni@unisla.ac.id](mailto:rosidizamroni@unisla.ac.id)

\*) Corresponding Author

---

### ABSTRACT

*This study aims to develop a sales prediction system using the Decision Tree C4.5 algorithm to assist YTI Store in identifying product demand. The research employed a quantitative approach with experimental methods. The dataset consists of sales transactions from YTI Store covering six months. Data preprocessing was performed to ensure consistency and completeness, followed by model training using the C4.5 algorithm. The system was implemented with features such as data input, preprocessing, decision tree visualization, and prediction reporting. Evaluation results using a confusion matrix yielded an accuracy of 84%, supported by acceptable precision and recall values. The findings indicate that sales prediction using C4.5 is effective in identifying product sales patterns, which can help store managers optimize inventory decisions, marketing strategies, and profit margins.*

**Keywords:** sales prediction; decision tree; algoritma C4.5; classification.

### ABSTRAK

*Penelitian ini bertujuan untuk mengembangkan sistem prediksi penjualan dengan menggunakan algoritma Decision Tree C4.5 untuk membantu YTI Store dalam mengidentifikasi permintaan produk. Penelitian menggunakan pendekatan kuantitatif dengan metode eksperimen. Dataset yang digunakan berupa transaksi penjualan YTI Store selama enam bulan. Data dilakukan pra-pemrosesan untuk memastikan konsistensi dan kelengkapan, kemudian dilatih menggunakan algoritma C4.5. Sistem diimplementasikan dengan fitur input data, preprocessing, visualisasi pohon keputusan, dan pelaporan hasil prediksi. Evaluasi menggunakan confusion matrix menghasilkan akurasi sebesar 84% dengan nilai precision dan recall yang memadai. Hasil penelitian menunjukkan bahwa prediksi penjualan menggunakan C4.5 efektif dalam mengidentifikasi pola penjualan produk, sehingga dapat membantu manajer toko dalam mengoptimalkan keputusan persediaan, strategi pemasaran, dan margin keuntungan.*

**Kata kunci:** prediksi penjualan; pohon keputusan; algoritma C4.5; klasifikasi.

---

## A. PENDAHULUAN

Persaingan bisnis ritel semakin ketat seiring dengan meningkatnya kebutuhan konsumen. Toko-toko modern maupun tradisional harus mampu melakukan pengelolaan persediaan dengan tepat agar tidak terjadi penumpukan stok maupun kekosongan barang. YTI Store sebagai salah satu toko ritel menghadapi tantangan dalam memprediksi produk apa saja yang akan laku terjual. Keputusan yang hanya berdasarkan intuisi pemilik berisiko menimbulkan inefisiensi, baik berupa overstock maupun lost sales.

Data mining menjadi solusi untuk mengolah data transaksi menjadi informasi yang bermanfaat. Salah satu teknik yang banyak digunakan adalah algoritma *Decision Tree*, khususnya C4.5. Algoritma ini mampu menghasilkan aturan klasifikasi yang mudah dipahami serta divisualisasikan dalam bentuk pohon keputusan. Berbagai penelitian sebelumnya mendukung efektivitas C4.5, seperti penelitian [1], yang mencapai akurasi 92,11% dalam prediksi penjualan produk elektronik, dan [2] yang menggunakannya pada prediksi kualitas benih dengan hasil memuaskan.[3]

Fenomena ini menunjukkan pentingnya penerapan metode prediksi penjualan yang efektif, seperti Decision Tree, untuk membantu pengelola toko membuat keputusan berbasis data. Metode Decision Tree dikenal sebagai salah satu teknik analisis kuantitatif yang mampu mengidentifikasi pola-pola kompleks dari informasi data penjualan sebelumnya untuk memprediksi tren penjualan di masa mendatang untuk hasil akurasi yang tinggi. Dengan penggunaan metode ini, pengelola toko dapat mengoptimalkan pengelolaan stok, meningkatkan efisiensi operasional, dan meningkatkan kepuasan pelanggan[3].

Berdasarkan penelitian terdahulu, penelitian ini difokuskan untuk mengembangkan sistem prediksi penjualan di YTI Store dengan algoritma C4.5. Tujuannya adalah menghasilkan model yang mampu mengklasifikasikan status produk laku atau tidak laku dengan tingkat akurasi yang baik, sehingga dapat digunakan sebagai dasar pengambilan keputusan bisnis.

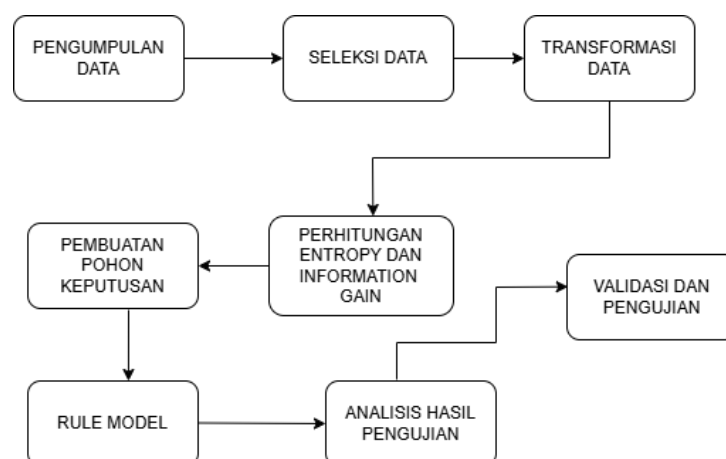
Hipotesis penelitian adalah penerapan algoritma C4.5 mampu meningkatkan ketepatan prediksi penjualan di YTI Store dengan akurasi minimal 80%.

## B. METODE

Metodologi penelitian ini menjelaskan tahapan sistematis yang digunakan dalam merancang, mengimplementasikan, dan menguji aplikasi prediksi penjualan pada Toko YTI dengan algoritma Decision Tree C4.5. Tahapan penelitian meliputi:

### 1. DIAGRAM ALIR PENELITIAN

Penelitian ini diawali dengan pengumpulan data, dilanjutkan dengan preprocessing, perhitungan entropy dan information gain, pembentukan pohon keputusan, implementasi sistem, hingga evaluasi akurasi model.



Gambar 1 Diagram Alir Sistem

Gambar 1 menunjukkan alur proses prediksi penjualan produk menggunakan algoritma Decision Tree, dimulai dari pengumpulan dan seleksi data, dilanjutkan dengan transformasi data agar sesuai untuk analisis. Selanjutnya dilakukan perhitungan entropy dan information gain

untuk menentukan atribut terbaik sebagai dasar pembuatan pohon keputusan. Pohon yang terbentuk kemudian diubah menjadi rule model yang digunakan untuk memprediksi status penjualan produk. Hasil prediksi diuji dan divalidasi melalui pengujian akurasi untuk menilai keandalan sistem dalam menghasilkan prediksi yang tepat.

## 2. DATASET PENELITIAN

Dataset diperoleh dari transaksi penjualan Toko YTI selama 6 bulan, dengan total 500 records. Atribut yang digunakan meliputi nama produk, diskon ya atau tidak, stok, jumlah penjualan, dan status laku atau tidak laku.

Tabel 1 Data Produk

No	Nama Produk	Diskon	Stok	Penjualan	Status
1	Greenfields Fresh Milk 1L	Tidak	99	70	Laku
2	Roma Biskuit Kelapa	Tidak	31	3	Tidak Laku
3	Indomilk Full Cream 1L	Ya	78	21	Tidak Laku
4	So Klin Pewangi 900ml	Tidak	87	26	Tidak Laku
5	Nabati Wafer Keju	Ya	27	20	Laku
6	Yakult 5 pcs	Tidak	92	8	Tidak Laku
7	Sariwangi Teh Celup	Tidak	73	37	Laku
8	Yakult 5 pcs	Ya	84	71	Laku
9	Frisian Flag Kental Manis	Tidak	66	1	Tidak Laku
10	So Klin Pewangi 900ml	Tidak	38	6	Tidak Laku
...	.....	.....	....	.....	.....
500	Rinso Deterjen Cair 800ml	Tidak	27	4	Tidak Laku

## 3. ALGORITMA C45

Algoritma C4.5 adalah algoritma yang diterapkan untuk membentuk pohon keputusan (*decision tree*) dalam proses pengambilan keputusan. Algoritma ini merupakan salah satu bentuk dari algoritma induksi pohon keputusan yang berasal dari ID3 (Iterative Dichotomiser 3), yang dikembangkan oleh J. Ross Quinlan. Dalam proses algoritma ID3, data masukan terdiri dari sampel pelatihan, label pelatihan, dan atribut. C4.5[4] sendiri merupakan versi lanjutan dari ID3 dengan sejumlah penyempurnaan, seperti kemampuan menangani data yang hilang (*missing value*), mengelola data kontinu, dan melakukan pemangkasan (*pruning*). Pemilihan atribut sebagai akar dalam algoritma ini didasarkan pada nilai gain tertinggi dari seluruh atribut yang tersedia[5]

Algoritma C4.5 merupakan metode klasifikasi berbasis pohon keputusan yang dibangun melalui tahapan pemilihan atribut utama sebagai titik awal, dilanjutkan dengan pembentukan cabang sesuai nilai atribut. Setiap kasus kemudian dikelompokkan berdasarkan cabang hingga semua data dalam cabang berada pada kelas yang sama. Proses ini menghasilkan pohon keputusan lengkap yang pada akhirnya diterjemahkan menjadi aturan (*rule*) untuk klasifikasi data[5].

Data mining adalah proses menganalisis data dengan memanfaatkan perangkat lunak untuk menemukan pola dan aturan dalam sebuah dataset. Metode ini memungkinkan pengolahan data berukuran besar menjadi informasi yang lebih bermanfaat sebagai dasar pengambilan keputusan. Salah satu teknik yang banyak digunakan dalam data mining yaitu klasifikasi, yakni proses membangun model atau fungsi yang dapat menggambarkan serta membedakan kelas atau konsep data, sehingga model tersebut bisa dipakai untuk memprediksi label kelas pada data baru yang belum diketahui sebelumnya.[6]

Sementara itu, perhitungan nilai entropi dapat dilihat pada persamaan 2 berikut. Dimana:

$$Entropy (S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Keterangan dari persamaan 2 :

S : himpunan kasus

A : fitur

N : jumlah partisi S

pi : proporsi dari Si terhadap S

Secara garis besar, tahapan algoritma C4.5 dalam membangun pohon keputusan menurut [7] adalah sebagai berikut :

Memilih atribut sebagai akar.

Membuat cabang pada tiap-tiap nilai.

Membagi kasus dalam cabang.

Ulangi proses pada setiap cabang hingga semua kasus pada cabang mempunyai kelas yang sama.

Algoritma C4.5 digunakan untuk membangun *Decision Tree* berdasarkan perhitungan *Entropy* dan *Information Gain*. Atribut dengan nilai gain ratio tertinggi dipilih sebagai akar pohon.[8]

Rumus Information Gain:

$$Entropy (S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Perhitungan entropy dan gain pada atribut diskon, stok, dan penjualan menghasilkan nilai tertinggi pada atribut penjualan, sehingga atribut ini dipilih sebagai akar pohon [9].

Tabel 2 Perhitungan *Entropy* dan *Gain*

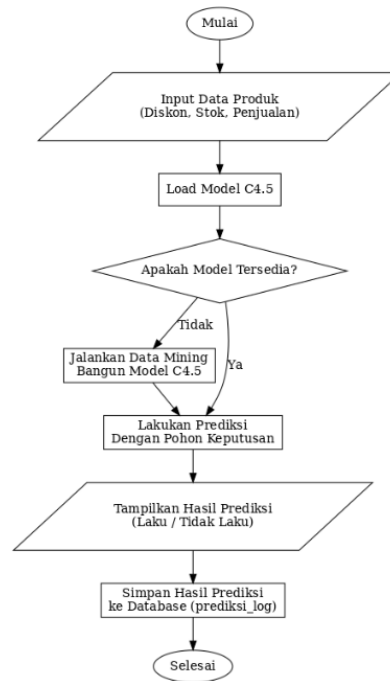
Atribut	Nilai Entropy Subset	Nilai Gain	Keterangan
Diskon	Entropy (Ya) = 0,81 Entropy (Tidak) = 0,997	0,09	Rendah kontribusi
Stok	Entropy (Tinggi) = 0,72 Entropy (Rendah) = 0,999	0,14	Sedang kontribusi
Penjualan	Entropy (Tinggi) = 0,48 Entropy (Rendah) = 0,97	0,21	Tertinggi, dipilih sebagai akar pohon

#### 4. PERANCANGAN SISTEM

Sistem dibangun dengan pendekatan data mining berbasis web menggunakan framework CodeIgniter, MySQL sebagai basis data, serta XAMPP sebagai server lokal.

Untuk menggambarkan alur proses kerja sistem, digunakan flowchart berikut.[10]

Gambar 2 menunjukkan alur proses prediksi dengan Algoritma C4.5, dimulai dari start, lalu dilanjutkan dengan proses data mining. Setelah itu, sistem akan melakukan perhitungan menggunakan Algoritma C4.5, termasuk perhitungan Entropy dan Gain, hingga akhirnya menghasilkan tampilan pohon keputusan.



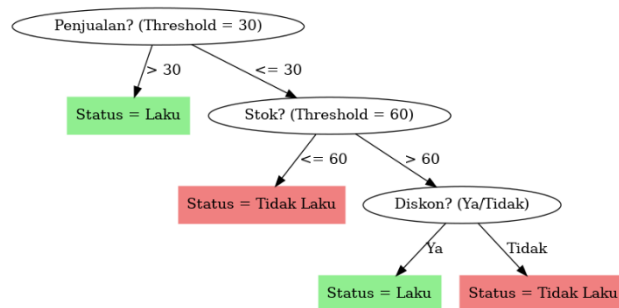
Gambar 2 Flowchart Sistem Prediksi

## C. HASIL DAN PEMBAHASAN

### 1. HASIL PEMBENTUKAN POHON KEPUTUSAN

Berdasarkan perhitungan Entropy dan Information Gain pada atribut Diskon, Stok, dan Penjualan, diperoleh bahwa atribut Penjualan memiliki nilai Gain tertinggi (0,21). Dengan demikian, atribut ini dipilih sebagai node akar pada pohon keputusan.

Berikut adalah ilustrasi visual dari pohon keputusan yang terbentuk:



Gambar 3 Pohon Keputusan Hasil Prediksi

Gambar 3 Pohon Keputusan menunjukkan alur klasifikasi status penjualan berdasarkan atribut penjualan, stok, dan diskon.

Model prediksi penjualan berbasis Decision Tree telah dibangun dan diuji pada data sampel. Hasilnya mampu memprediksi status penjualan laku atau tidak laku meskipun akurasi masih terbatas. Dengan pengembangan lanjutan dan data lebih besar, sistem berpotensi menjadi alat bantu manajemen stok yang efektif.

## 2. HASIL PREDIKSI PENJUALAN

Proses prediksi dilakukan terhadap data uji (*testing set*) yang berasal dari transaksi Toko YTI, menggunakan 70 Data Uji. Hasil prediksi status penjualan dapat dilihat pada tabel berikut:

Tabel 3 Hasil Prediksi

No	Produk	Diskon	Stok	Penjualan	Aktual	Prediksi	Hasil
1	Pepsi Botol 1L	Tidak	73	58	L	L	Ya
2	Good Day Mocacino	Ya	79	15	L	L	Ya
3	Baygon Aerosol 600ml	Tidak	82	27	TL	TL	Ya
4	Rinso Deterjen Cair 800ml	Ya	92	20	L	L	Ya
5	Pantene Shampo 180ml	Ya	91	83	L	L	Ya
6	Indomie Goreng Original	Tidak	49	48	L	L	Ya
7	Filma Minyak Goreng 1L	Ya	94	3	L	L	Ya
8	Sariwangi Teh Celup	Tidak	69	4	TL	TL	Ya
9	Le Minerale 1500ml	Tidak	35	20	L	TL	Tidak
10	Kapal Api Kopi 165gr	Ya	66	33	L	L	Ya
11	Kapal Api Kopi 165gr	Tidak	92	56	L	L	Ya
12	Nabati Wafer Keju	Ya	59	6	Tidak Laku	Tidak Laku	Ya
13	Good Day Mocacino	Ya	55	50	Laku	Laku	Ya
14	Sunsilk Black Shine 170ml	Tidak	89	59	Laku	Laku	Ya
15	Nabati Wafer Keju	Tidak	22	18	Laku	Tidak Laku	Tidak
16	Good Day Mocacino	Ya	55	44	Laku	Laku	Ya
17	Good Day Mocacino	Ya	27	14	Tidak Laku	Tidak Laku	Ya
...	.....	.....	....	...	.....	.....	.....
70	Rinso Deterjen Cair 800ml	Tidak	27	4	Tidak Laku	Tidak Laku	Ya

Tabel 3 menyajikan hasil prediksi status penjualan menggunakan model Decision Tree berdasarkan atribut diskon, stok, dan penjualan. Sebanyak 70 data uji digunakan. Prediksi sistem dibandingkan dengan data aktual untuk mengukur performa model dari keseluruhan data uji, sebanyak 59 prediksi dinyatakan benar.

$$Accuracy = \frac{59}{70} \times 100\% = 84\%.$$

## D. KESIMPULAN

Berdasarkan hasil penelitian dan implementasi sistem prediksi penjualan dengan algoritma Decision Tree C4.5, diperoleh beberapa kesimpulan sebagai berikut:

### 1. Penerapan metode *Decision Tree* untuk prediksi penjualan

Metode Decision Tree berhasil diterapkan melalui tahapan pengolahan dataset, pembagian data menjadi training dan testing, pembentukan pohon keputusan, hingga menghasilkan aturan klasifikasi (rules). Aturan ini digunakan untuk menentukan apakah suatu produk masuk kategori laku atau tidak laku. Dengan demikian, metode ini terbukti mampu diaplikasikan untuk memprediksi penjualan di toko ritel.

### 2. Akurasi model prediksi

Hasil evaluasi pada 70 data uji menunjukkan bahwa model C4.5 mampu memprediksi dengan akurasi sekitar 84% 59 prediksi benar dari 70 data uji. Nilai ini menunjukkan bahwa model cukup andal digunakan untuk membantu pengambilan keputusan dalam manajemen persediaan dan strategi promosi toko

## E. DAFTAR PUSTAKA

- [1] M. Leonardi, R. Emilda, I. Katrin, dan A. Yulianto, “Prediksi Penjualan Produk Rokok Pada PT. Indomarco Prismatama Menggunakan Algoritma C4.5,” *P*, vol. 23, no. 2, Okt 2021, doi: 10.31294/p.v23i2.11151.
- [2] F. Faisal, H. Dhika, dan H. Veris, “Penerapan Algoritma Decision Tree Dalam Penjualan Handphone,” *JRKT*, vol. 1, no. 04, Des 2021, doi: 10.30998/jrkt.v1i04.6157.
- [3] R. Beay dan F. M. Sarimole, “Application of Decision Tree Method for Sales Prediction at PT. Cipta Naga Semesta (Mayora Group) North Jakarta for 2023,” *ijsecs*, vol. 4, no. 3, hlm. 943–952, Des 2024, doi: 10.35870/ijsecs.v4i3.2999.
- [4] “Supervised learning: From theory to applications,” dalam *Artificial Intelligence and Machine Learning for EDGE Computing*, Elsevier, 2022, hlm. 23–32. doi: 10.1016/b978-0-12-824054-0.00026-5.
- [5] M. K. Al Fatach dan A. Wibowo, “PERBANDINGAN METODE ALGORITMA C4.5 DAN NAIVE BAYES UNTUK MEMPREDIKSI PENJUALAN KOSMETIK PADA TOKO JELITA,” *Mnemonic*, vol. 7, no. 2, hlm. 220–225, Sep 2024, doi: 10.36040/mnemonic.v7i2.10730.
- [6] E. Saputri, “Teknik dan aplikasi data mining di Indonesia: tinjauan literatur satu dekade (2015-2024),” vol. 04, no. 02.
- [7] A. H. Nasrullah, “IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS,” *JIKOM*, vol. 7, no. 2, hlm. 45–51, Sep 2021, doi: 10.35329/jiik.v7i2.203.
- [8] N. Febiani, Abd. C. Fauzan, dan M. M. Huda, “IMPLEMENTASI ALGORITMA DECISION TREE C4.5 DENGAN IMPROVISASI MEAN DAN MEDIAN PADA DATASET NUMERIK,” *Tekinkom*, vol. 5, no. 1, hlm. 105, Jun 2022, doi: 10.37600/tekinkom.v5i1.435.
- [9] D. M. Musa *dkk.*, “Penerapan Data Mining Untuk Klasifikasi Data Penjualan Pakan Ternak Terlaris Dengan Algoritma C4.5,” *JTIK*, vol. 10, no. 1, hlm. 172–186, Mar 2024, doi: 10.37012/jtik.v10i1.1985.
- [10] F. R. Hariri dan C. Mashuri, “Sistem Informasi Peramalan Penjualan dengan Menerapkan Metode Double Exponential Smoothing Berbasis Web,” *generation*, vol. 6, no. 1, hlm. 68–77, Jan 2022, doi: 10.29407/gj.v6i1.16204.