

Implementasi Data Mining Dalam Klasifikasi Produk Laris Dan Tidak Laris Dengan Algoritma C4.5

Implementation Of Data Mining In Classification Of Best-Selling And Non-Selling Products With The C4.5 Algorithm

Nurul Syahfitri Sitorus^{a,1}, Dedi Leman^{b,2}

^a Jl. KL, Yos Sudarso Km. 6,5 No. 3A, Tanjung Mulia, Kec.Medan Deli, Kota Medan, 20241, Indonesia

^b Jl. KL, Yos Sudarso Km. 6,5 No. 3A, Tanjung Mulia, Kec.Medan Deli, Kota Medan, 20241, Indonesia
nurulsyahfitri078@gmail.com¹, dedileman01@gmail.com²

ABSTRAK

Data mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan. Dilakukannya penelitian disebuah perusahaan distributor yang bergerak dibidang otomotif seperti spare part motor atau komponen pengganti untuk kendaraan. Selama penelitian, penulis mendapatkan suatu masalah dimana sering terjadinya tidak ada stok barang disaat pembeli ingin membeli suatu produk tertentu dikarenakan perusahaan tidak menyetok produk tersebut dengan jumlah banyak yang padahal produk tersebut banyak diminati. Metode yang digunakan dalam penelitian ini ialah algoritma data mining C4.5 dengan klasifikasi beberapa kriteria meliputi Merk, Tingkatan, Harga dengan jumlah sebanyak 695 data. Dari hasil perhitungan tersebut metode data mining C4.5 dapat disimpulkan memiliki tingkat keakuratan yang cukup bagus dimana dalam penelitian ini memiliki akurasi 69,86%, presisi 64,76%, dan recall 87,05%. Dimana hal ini mampu untuk mengatasi masalah tingginya jumlah stok barang yang memiliki tingkat penjualan yang rendah.

Kata Kunci : Data Mining, algoritma C4.5, Klasifikasi Produk, akurasi, presisi, recall

ABSTRACT

Data mining is the process of obtaining useful information from a large database that needs to be extracted to become new information and can help in decision making. The research was conducted at a distributor company engaged in the automotive sector such as motorcycle spare parts or replacement components for vehicles. During the research, the author found a problem where there was often no stock of goods when buyers wanted to buy a particular product because the company did not stock the product in large quantities even though the product was in great demand. The method used in this study is the C4.5 data mining method with a classification of several criteria including Brand, Level, Price with a total of 695 data. From the results of these calculations, the C4.5 data mining method can be concluded to have a fairly good level of accuracy where in this study it had an accuracy of 69.86%, a precision of 64.76%, and a recall of 87.05%. Where this is able to overcome the problem of the high number of stocks of goods that have low sales levels.

Keywords: Data Mining, C4.5 Method, Product Classification, accuracy, precision, recall

Info Artikel :

Disubmit: 11 Januari 2025

Direview: 15 Februari 2025

Diterima : 12 Maret 2025

Copyright © 2025 - Journal UPU. All rights reserved.

1. PENDAHULUAN

Di era digital dan pesatnya teknologi saat ini mendorong banyak perusahaan untuk selalu meningkatkan sistem dalam perusahaannya untuk menjaga stabilitas penjualan. Dengan adanya sistem yang baik maka akan sangat membantu perusahaan dalam menjalankan bisnisnya, salah satunya sistem pengolahan data. Dengan adanya sistem pengolahan data yang baik akan menciptakan suatu informasi

penting yang bermanfaat dan bisa dijadikan sebagai sumber pengambilan keputusan dan juga menciptakan strategi baru dalam meningkatkan penjualan perusahaan demi dapat mempertahankan konsumen dan memperoleh konsumen baru, dimana konsumen ini memiliki peran yang sangat penting yaitu sebagai tanda meningkatnya penjualan atau pendapatan untuk suatu perusahaan.

Dalam penelitian, didapatkan suatu masalah dimana sering terjadinya tidak ada stok barang disaat pembeli ingin membeli suatu produk tertentu dikarenakan perusahaan tidak menyetok produk tersebut dengan jumlah banyak yang padahal produk tersebut banyak diminati. Dan disisi lain, perusahaan melakukan penyetokkan stok dengan jumlah yang banyak padahal produk tersebut kurang diminati atau memiliki tingkat penjualan yang rendah dimana ini mengakibatkan terjadinya penumpukkan stok barang karna jarang terjual. Hal ini terjadi dikarenakan ada banyaknya jenis barang/produk sehingga perusahaan sulit untuk mengetahui produk mana saja yang memiliki tingkat penjualan yang tinggi.

Klasifikasi produk menggunakan data mining algoritma C4.5 yang merupakan sebuah meted untuk mengolah data yang besar sehingga menghasilkan suatu informasi yang berguna khususnya dalam hal mengambil keputusan. Dalam penelitian ini untuk mengetahui produk mana yang paling banyak terjual dengan menggunakan pembagian beberapa atribut meliputi *Merk*, Kualitas, Harga, dan Stok.

Tujuan dari penelitian ini adalah untuk menciptakan suatu sistem pengolahan data produk yang tadinya dianggap tidak penting menjadi suatu informasi yang penting dan bermanfaat. Berdasarkan yang telah dibahas guna membantu perusahaan dalam menetapkan strategi baru dan mengetahui produk terlaris mereka, dimana hasil dari proses klasifikasi tersebut akan diinput dalam sebuah web sehingga perusahaan dapat dengan mudah melihat data produknya.

Berikut beberapa penelitian terdahulu terkait penelitian yang serupa dilakukan oleh Tabitha Salsabila dan Sulastri dengan judul “Implementasi Algoritma C.45 Untuk Klasifikasi Produk Laris Sepeda Motor Honda Pada CV. Cendana Motor Cepiring” menghasilkan kesimpulan bahwa klasifikasi dengan akurasi sebesar 99% yang menunjukkan bahwa algoritma C4.5 cocok digunakan untuk mengukur perkiraan penjualan sepeda motor Honda terlaris (Salsabilla dan Sulastri 2022).

Kemudian ada penelitian lain yang dilakukan oleh Enjel Lia dan Stefanus Setyo Wibagso dengan judul “Implementasi Algoritma C.45 Untuk Penentuan Produk Terlaris (Studi Kasus : Toko XYZ Palembang)” menghasilkan kesimpulan bahwa berdasarkan hasil pengujian dengan menggunakan 437 data diperoleh nilai *accuracy* sebesar 91,53%, nilai *precision* sebesar 27,77%, dan nilai *recall* sebesar 17,24%. Hal ini menunjukan bahwa sistem memiliki kinerja yang baik dan algoritma C4.5 dapat diterapkan pada kasus klasifikasi penentuan produk terlaris di toko XYZ Palembang (Lia dan Wibagso 2022).

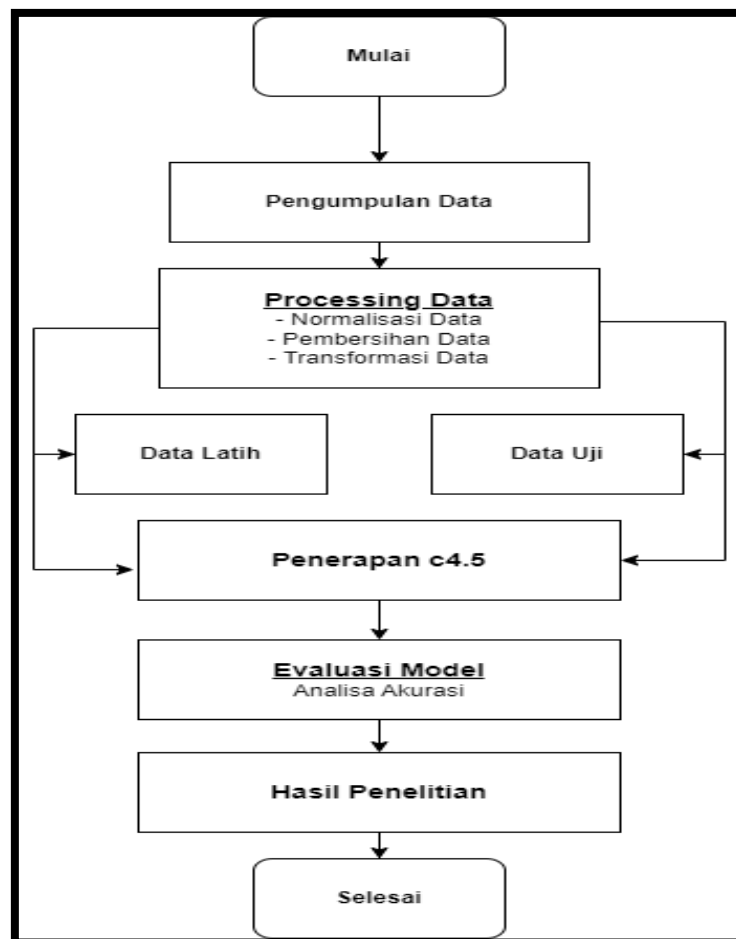
2. METODE

Berikut tahapan penelitian sekaligus tahapan proses data mining yang dilakukan dalam rangka penyelesaian masalah yang akan dibahas.

- 1) Pengumpulan Data: Mengumpulkan data yang relevan untuk analisis.
- 2) Preprocessing Data
 - a. Normalisasi Data: Pastikan data berada dalam skala yang konsisten, terutama jika fitur memiliki unit yang berbeda, Menyebabkan model lebih efisien dan efektif dalam mengidentifikasi pola karena fitur berada dalam skala yang sama.
 - b. Pembersihan Data: Hapus data yang tidak relevan atau noise dari dataset, Data yang bersih dan konsisten meningkatkan akurasi model dan mencegah hasil yang bias
 - c. Transformasi Fitur: Misalnya, jika data terdistribusi secara log-normal, pertimbangkan untuk melakukan transformasi logaritmik, Mengurangi dimensi dapat membantu dalam

menangani kompleksitas dan waktu komputasi yang tinggi, serta meningkatkan performa algoritma.

3. Pembagian Dataset: Membagi dataset menjadi dua bagian: data latih (training) dan data uji (testing).
4. Penerapan Algoritma C4.5: Menggunakan algoritma C4.5 untuk membuat pohon keputusan berdasarkan data latih. Dalam tahap ini, data klasifikasi produk yang paling banyak terjual yang telah diperoleh akan diolah dengan menggunakan algoritma C4.5. Proses analisis dengan metode c45 dimulai dengan penentuan faktor-faktor penilaian kemudian melakukan pemodelan dengan rumus-rumus yang dimiliki oleh algoritma C4.5.
5. Evaluasi Model: mengevaluasi serta menghitung data-data yang akan digunakan untuk perhitungan data mining C4.5 dalam klasifikasi produk laris dan tidak laris
6. Hasil Penelitian: menampilkan output berdasarkan perhitungan dalam kajian yang telah dilakukan.



Gambar 1. *Flowchart* Tahap Penelitian

Data mining adalah proses mengumpulkan dan mengolah data yang cukup populer. Menurut (Suntoro, 2019) data mining adalah proses untuk mendapatkan informasi yang berguna dari basis data yang besar dan perlu diekstraksi agar menjadi informasi baru dan dapat membantu dalam pengambilan keputusan.

Data mining juga dapat di artikan sebagai suatu aktivitas yang mengumpulkan dan menggunakan data historis untuk menemukan pola, keteraturan, dan hubungan dalam kumpulan data yang besar. Data mining juga disebut sebagai serangkaian langkah untuk mempelajari nilai tambah secara manual berupa pengetahuan yang tidak diketahui dari suatu kumpulan data (Salsabilla dan Sulastri 2022). Data mining memiliki beberapa karakteristik sebagai berikut :

- 1) Data mining saling berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- 2) Dalam proses pengolahan datanya, data mining bisa menggunakan data yang kecil hingga besar.
- 3) Data mining sangat berguna untuk membuat keputusan yang kritis terutama dalam strategi.

Dan juga, *Decision Tree* (pohon keputusan) merupakan representasi sederhana dari teknik klasifikasi untuk sejumlah kelas terbatas, dimana simpul internal maupun simpul akar ditandai dengan nama atribut, rusuk-rusuknya ditandai dengan nilai atribut yang mungkin dan simpul daun ditandai dengan kelas-kelas yang berbeda. Pohon keputusan memiliki beberapa kekurangan dan kelebihan, yaitu

- 1) Kelebihan
 - a. Daerah pengambilan keputusan dapat diubah sehingga menjadi sederhana.
 - b. Dapat menghilangkan perhitungan tidak penting karena dalam proses pengujian hanya berdasarkan kriteria yang diperlukan saja.
 - c. Proses pemilihan fitur dari internal node memiliki karakteristik perbedaan antara satu kriteria dan yang lainnya.
 - d. Dapat menghindar sehingga munculnya permasalahan dalam menggunakan kriteria dengan sejumlah kecil pada node internal tanpa mengurangi kualitas keputusan yang dihasilkan.
- e. Kekurangan
 - f. Dapat menjadi *overlap* jika hasil keputusan dan kriteria yang digunakan jumlahnya sangat besar yang berakibat bertambahnya waktu untuk pengambilan keputusan dan jumlah memori yang dibutuhkan semakin tinggi.
 - g. Pengumpulan jumlah kesalahan setiap pohon keputusan besar.
 - h. Proses mendesai pohon keputusan yang cukup sulit.
 - i. Kualitas keputusan yang diambil sangat tergantung dengan bagaimana pohon tersebut di desain

3. HASIL DAN PEMBAHASAN

Algoritma C4.5 merupakan salah satu algoritma dalam data mining untuk membangun model pohon keputusan (*Decision Tree*), secara esensial, metode ini berfokus pada memahami struktur data dan mengklasifikasikannya kedalam kelompok yang sesuai. Fungsi utama dari algoritma ini adalah digunakan untuk klasifikasi dan prediksi.

Algoritma C4.5 dan pohon keputusan merupakan dua model yang tak terpisahkan, karena untuk membangun sebuah pohon keputusan dibutuhkan algoritma C4.5. Adapun langkah-langkah dari proses algoritma C4.5 ini meliputi :

- 1) Pembentukan pohon awal, melibatkan pemilihan atribut yang dianggap *paling informatif*.
- 2) Pembagian Data, setiap cabang pohon mempresentasikan nilai tertentu dari atribut yang dipilih.
- 3) Perhitungan *gain ratio*, mengukur seberapa baik suatu atribut untuk dapat memisahkan data menjadi kelas-kelas yang berbeda dalam memandu proses pengambilan keputusan.

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5, yaitu.

- 1) Menyiapkan data training, dimana hal ini biasanya dari history yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
- 2) Penentuan akar pohon dengan menghitung gain dari semua atribut. Sebelumnya dilakukan perhitungan *entropy*, dengan rumus sebagai berikut.

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

Rumus 1. Perhitungan *entropy*

Keterangan :

S = Himpunan Kasus

N = Jumlah partisi S

P_i = Proporsi dari S_i terhadap S

- 3) Hitung nilai *gain* menggunakan rumus sebagai berikut.

$$\text{Gain (S,A)} = \text{Entropy(S)} - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy(S}_i)$$

Rumus 2. Perhitungan *Gain*

Keterangan :

S = Himpunan Kasus

A = Atribut

N = Jumlah partisi atribut A

|S_i| = Jumlah kasus pada partisi ke-i

|S| = Jumlah kasus dalam S

- 1) Menghitung nilai *Gain Ratio* untuk masing-masing atribut, dimana yang memiliki *Gain Ratio* tertinggi dipilih menjadi akar (*root*) dan atribut dengan *Gain Ratio* terendah dari akar (*root*) dipilih menjadi cabang (*branches*).
- 2) Ulangi langkah ke 2 hingga semua record terpartisi dengan sempurna.
- 3) Proses partisi pohon keputusan (*decision tree*) akan berhenti pada saat berikut:
 - a. Semua record dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam record yang dipartisi lagi.
 - c. Tidak ada record di dalam cabang yang kosong.

Berikut cara algoritma C4.5 dalam membangun pohon keputusan, sebagai berikut :

- 1) Melakukan pemilihan atribut yang akan digunakan sebagai akar
- 2) Membuat sebuah cabang untuk setiap nilai
- 3) Melakukan pembagian kasus dalam sebuah cabang
- 4) Mengulangi proses setiap cabang hingga semua kasus pada cabang memiliki kelas yang sama.

Dilakukan juga analisis masalah, pengenalan terhadap masalah atau inventarisasi masalah yang akan diteliti. Permasalahan yang ditemukan kemudian akan dianalisa, langkah dalam proses analisa masalah adalah langkah untuk memahami masalah yang telah ditentukan. Dengan menganalisa permasalahan yang telah ditentukan tersebut, maka diharapkan masalah tersebut dapat dipahami dengan baik. Dalam proses analisis masalah juga telah menemukan faktor-faktor penilaian apa saja yang akan digunakan klasifikasi produk laris dan tidak laris.

Berdasarkan analisis masalah diatas, penulis mendapatkan hasil dari analisa untuk inputan sistem klasifikasi produk laris dan tidak laris menggunakan algoritma C4.5 yang dapat dilihat sebagai berikut.

- 1) Data Produk, digunakan sebagai inputan pada aplikasi untuk perincian produk-produk data penjualan.
- 2) Data Penjualan, digunakan sebagai inputan pada aplikasi yang akan dirancang dan untuk perhitungan metode yang akan digunakan yaitu data mining C4.5.

- 3) Algoritma C4.5, digunakan sebagai metode yang dapat menyelesaikan masalah yang di angkat dalam kajian ini.

Pengumpulan data, melakukan pengumpulan data yang akan digunakan dalam proses pengkajian klasifikasi produk laris dan tidak laris.

LAPORAN BARANG MASUK DAN KELUAR TAHUN 2023										
NO	TANGGAL	Nama Barang	Stok	Barang	Harga Beli	Barang	Harga	Stok Akhir	TOTAL	KETERANGAN
SI105673	03 April 2023	KNZ CAKRAM ASSY+PIRINGAN RXK OLD/RXS	2891	2771	Rp 4.200	234	Rp 5.700	5428	Rp 22.797.600	
		KNZ CAKRAM ASSY+PIRINGAN SATRIA FU 150 REAR	18	14	Rp 416.000	23	Rp 417.500	9	Rp 3.744.000	
		KNZ CAKRAM ASSY+PIRINGAN SMASH NEW	8	4	Rp 454.000	5	Rp 455.500	7	Rp 3.178.000	
		KNZ CAKRAM ASSY+PIRINGAN SUPRA X	240	236	Rp 12.200	40	Rp 5.700	436	Rp 500.000	
SI105674	03 April 2023	KNZ CAKRAM ASSY+PIRINGAN SUPRA X 125 (REAR)	1420	1416	Rp 10.200	156	Rp 5.700	2680	Rp 1.638.000	
		KNZ CAKRAM ASSY+PIRINGAN TIGER NEW REAR/MEGA PRO NEW REAR	1030	897	Rp 11.200	178	Rp 5.700	1749	Rp 2.047.000	
		KNZ BUSI KARISMA CPR6EA	6960	3870	Rp 11.200	573	Rp 5.700	10257	Rp 6.589.500	
		KNZ BOHLAM T10 BENING	6100	6096	Rp 200	423	Rp 5.700	11773	Rp 211.500	
SI105679	03 April 2023	KNZ BENTENGAN KOLING SHOGUN	79	75	Rp 28.000	123	Rp 29.500	31	Rp 868.000	
SI105697	03 April 2023	KNZ BENTENGAN KOPLING FORCE1	177	173	Rp 28.000	23	Rp 29.500	327	Rp 9.156.000	
		KNZ BENTENGAN KOPLING JUP MX	130	126	Rp 28.000	23	Rp 29.500	233	Rp 6.524.000	
		KNZ BENTENGAN KOPLING JUP	1	-3	Rp 31.000	0	Rp 32.500	-2	Rp -	

Gambar 2. Data Penjualan

Processing data, dilakukan proses pembersihan data-data dengan kriteria tertentu yang menjadi data perhitungan yang digunakan dalam klasifikasi produk laris dan tidak laris.

No	nama	Kualitas	merk	harga	stok	kelas asli
1	KNZ Block Vario	ori	knz	Rp260.000	60	tersedia
2	KNZ Block Seher -ring Xeon	ori	knz	Rp260.000	60	tersedia
3	KNZ BLOCK SEHER+RING BEAT	ori	knz	Rp260.000	60	tersedia
4	KNZ BLOCK SEHER+RING BEAT F1	ori	knz	Rp260.000	60	tersedia
5	KNZ BLOCK SEHER+RING BEAT POP	ori	knz	Rp260.000	60	tersedia
6	KNZ BLOCK SEHER+RING BLADE/REVO ABS	ori	knz	Rp260.000	60	tersedia
7	KNZ BLOCK SEHER+RING CB150R	ori	knz	Rp260.000	60	tersedia
8	KNZ BLOCK SEHER+RING CBR 150 NEW LED/GTR(K56)	ori	knz	Rp260.000	60	tersedia
9	KNZ BLOCK SEHER+RING F1ZR	ori	knz	Rp260.000	60	tersedia
10	KNZ BLOCK SEHER+RING JUP MX KING	ori	knz	Rp260.000	60	tersedia
11	KNZ BLOCK SEHER+RING JUP Z/ VEGA R NEW	ori	knz	Rp260.000	60	tersedia
12	KNZ BLOCK SEHER+RING JUPITER MX	ori	knz	Rp260.000	60	tersedia
13	KNZ BLOCK SEHER+RING KHARISMA	ori	knz	Rp260.000	60	tersedia
14	KNZ BLOCK SEHER+RING KLX 150 S	ori	knz	Rp260.000	60	tersedia
15	KNZ BLOCK SEHER+RING MEGA PRO NEW 2011	ori	knz	Rp260.000	60	tersedia
16	KNZ BLOCK SEHER+RING MIO	ori	knz	Rp260.000	60	tersedia
17	KNZ BLOCK SEHER+RING MIO J	ori	knz	Rp260.000	60	tersedia
18	KNZ BLOCK SEHER+RING MIO M3	ori	knz	Rp260.000	60	tersedia

Gambar 3. Hasil processing data

Deskripsi dataset, dataset yang digunakan dalam penelitian ini diperoleh dari perusahaan tempat penelitian yang dapat penulis lampirkan sebagai berikut :

Tabel 1. Tabel dataset

	nama	kualitas/tingkatan	merk	harga	stok	kelas asli
1	KNZ Block Vario	Ori	knz	3E+05	60	laris
2	KNZ Block Seher -ring Xeon	Ori	knz	3E+05	60	laris
3	KNZ BLOCK SEHER+RING BEAT	Ori	knz	3E+05	60	laris
4	KNZ BLOCK SEHER+RING BEAT F1	Ori	knz	3E+05	60	laris
5	KNZ BLOCK SEHER+RING BEAT POP	Ori	knz	3E+05	60	laris
6	KNZ BLOCK SEHER+RING BLADE/REVO ABS	Ori	knz	3E+05	60	laris
7	KNZ BLOCK SEHER+RING CB150R	Ori	knz	3E+05	60	laris
8	KNZ BLOCK SEHER+RING CBR 150 NEW LED/GTR(K56)	Ori	knz	3E+05	60	laris
9	KNZ BLOCK SEHER+RING F1ZR	Ori	knz	3E+05	60	laris
10	KNZ BLOCK SEHER+RING JUP MX KING	Ori	knz	3E+05	60	Tidaklaris
11	KNZ BLOCK SEHER+RING JUP Z/ VEGA R NEW	Ori	knz	3E+05	60	laris
12	KNZ BLOCK SEHER+RING JUPITER MX	Ori	knz	3E+05	60	laris
13	KNZ BLOCK SEHER+RING KHARISMA	Ori	knz	3E+05	60	laris
14	KNZ BLOCK SEHER+RING KLX 150 S	Ori	knz	3E+05	60	Tidaklaris
Cont.	KNZ BLOCK SEHER+RING MEGA PRO NEW 2011	Ori	knz	3E+05	60	tersedia

Transformasi data, ini dilakukan proses penyesuaian data yang dibutuhkan untuk perhitungan dalam penelitian ini, yaitu sebagai berikut :

- 1) Merk = terbagi menjadi merk KNZ, AHM, YGP
- 2) Tingkatan = berdasarkan kualitas meliputi ORI, KW 1, KW 2
- 3) Harga = <250.000 dan >250.000
- 4) Stok = > 65 dan <65 jumlah stok
- 5) Label Class = Laris dan Tidak Laris

Penerapan algoritma C4.5, menggunakan algoritma C4.5 untuk membuat pohon keputusan berdasarkan data. Entropy total adalah nilai informasi yang menyatakan ukuran ketidakpastian dari atribut dari suatu kumpulan objek data dengan menghitung jumlah kasus Laris (329) dan kasus Tidak laris (366) dimana 695 adalah jumlah kasus keseluruhan.

Perhitungan Nilai Entropy menggunakan rumus sebagai berikut :

Dimana cara penyelesaian nya sebagai berikut :

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

$$\begin{aligned}\text{Entropy (total)} &= ((-329/695) * \log_2 (329/695) + (-366/695) * \log_2 (366/695)) \\ &= ((-329/695) * (-1.0789) + (-366/695) * (-0.925169)) \\ &= ((-0.473381) * (-1.0789) + (-0,526619) * (-0.925169)) \\ &= 0.510743 + 0.487211 \\ &= 0.997954574\end{aligned}$$

Bahwa Nilai Entropy Total dari keseluruhan kasus tersebut yaitu 0.997954574

a. Entropy Atribut

1) Kategori **Merk**

Untuk atribut Kategori Pelanggan terdiri dari 3 nilai yaitu KNZ, AHM dan YGP dimana nilai Entropy masing-masing sebagai berikut :

a. Merk = KNZ

(Jumlah Kasus = 422, Kasus Laris = 202, Kasus Tidak laris = 240).

$$\begin{aligned}\text{Entropy (S)} &= \sum_{i=1}^n -p_i * \log_2 p_i \\ &= ((-202/422) * \log_2 (202/422) + (-240/422) * \log_2 (240/422)) \\ &= ((-0.457014) * -1.12969) + (-0.542986) * -0.881012)) \\ &= 0.516284 + 0.478378 \\ &= 0.994661694\end{aligned}$$

Bahwa Nilai Entropy Atribut KNZ tersebut yaitu : 0.994661694

b. Merk = AHM

(Jumlah Kasus = 103, Kasus Laris = 51, Kasus Tidak laris = 52).

$$\begin{aligned}\text{Entropy (S)} &= \sum_{i=1}^n -p_i * \log_2 p_i \\ \text{Entropy (AHM)} &= ((-51/103) * \log_2 (51/103) + (-52/103) * \log_2 (52/103)) \\ &= ((-51/103) * \log_2 (-1.014) + (-52/103) * \log_2 (-0.986061)) \\ &= ((-0.495146) * -1.014) + (-0.504854) * -0.986061)) \\ &= 0.502115 + 0.497817 \\ &= 0.999932005\end{aligned}$$

Bahwa Nilai Entropy Atribut AHM tersebut yaitu : 0.999932005

c. Merk = YGP

(Jumlah Kasus = 150, Kasus Laris = 76, Kasus Tidak laris = 74).

$$\begin{aligned}\text{Entropy (S)} &= \sum_{i=1}^n -p_i * \log_2 p_i \\ \text{Entropy (YGP)} &= ((-76/150) * \log_2 (76/150) + (-74/150) * \log_2 (74/150)) \\ &= ((-76/150) * \log_2 (-0.980891) + (-74/150) * \log_2 (-1.01937)) = ((-0.506) * -0.980891) + (-0.493) * -1.01937)) \\ &= 0.496985 + 0.502887 \\ &= 0.999871757\end{aligned}$$

Bahwa Nilai Entropy Atribut YGP tersebut yaitu : 0.999871757

2) Entropy Atribut Tingkatan Sparepart

Untuk atribut Warna terdiri dari 4 nilai yaitu Ori, KW1, KW2 dan KW3 dimana nilai Entropy masing-masing sebagai berikut :

a. Tingkatan Sparepart = Ori

(Jumlah Kasus = 174, Kasus Laris = 71, Kasus Tidak Laris = 103)

$$\begin{aligned}\text{Entropy (Ori)} &= ((-71/174) * \log_2 (71/174) + (-103/174) * \log_2 (103/174)) \\ &= ((-71/174) * \log_2 (0.408046) + (-103/174) * \log_2 (0.591954)) \\ &= ((-0.408046) * -1.2932) + (-0.591954) * -0.7564)) \\ &= 0.97546\end{aligned}$$

Bahwa Nilai Entropy Atribut Ori tersebut yaitu : 0.97546

b. Tingkatan Sparepart = KW1

(Jumlah Kasus = 172, Kasus Laris = 88, Kasus Tidak Laris = 84)

$$\begin{aligned}\text{Entropy (KW1)} &= ((-88/172) * \log_2 (88/172) + (-84/172) * \log_2 (84/172)) \\ &= ((-88/172) * \log_2 (0.511627) + (-84/172) * \log_2 (0.488372)) \\ &= 0.99961\end{aligned}$$

Bahwa Nilai Entropy Atribut KW1 tersebut yaitu : = 0.99961

c. Tingkatan Sparepart = KW2

(Jumlah Kasus = 349, Kasus Laris= 170, Kasus Tidak Laris = 179)

$$\begin{aligned}\text{Entropy (KW2)} &= ((-170/349) * \log_2 (170/349) + (-179/349) * \log_2 (179/349)) \\ &= ((-170/349) * \log_2 (0.48710) + (-179/349) * \log_2 (0.51289)) \\ &= 0.99952\end{aligned}$$

Bahwa Nilai Entropy Atribut KW2 tersebut yaitu : 0.99952

3) Entropy Atribut Harga

Untuk atribut Kategori harga terdiri dari 2 nilai yaitu mahal dan murah dimana nilai Entropy masing-masing sebagai berikut:

a. Harga = >250.000

(Jumlah Kasus = 571, Kasus Laris = 269, Kasus Tidak Laris = 302)

$$\begin{aligned}\text{Entropy (Jelas)} &= ((-269/571) * \log_2 (269/571) + (-302/571) * \log_2 (302/571)) \\ &= ((-269/571) * \log_2 (0.47110) + (-302/571) * \log_2 (0.52889)) \\ &= 0.997589305\end{aligned}$$

Bahwa Nilai Entropy Atribut untuk kategori Harga Mahal tersebut yaitu: 0.997589305

b. Harga = <250.000

(Jumlah Kasus = 124, Kasus laris = 60, Kasus Tidak Laris= 64)

$$\begin{aligned}\text{Entropy (Murah)} &= ((-60/124) * \log_2 (60/124) + (-64/124) * \log_2 (64/124)) \\ &= ((-60/124) * \log_2 (0.48387) + (-64/124) * \log_2 (0.51612)) \\ &= 0.999249248\end{aligned}$$

Bahwa Nilai Entropy Atribut untuk kategori harga murah tersebut yaitu : 0.999249248

4) Entropy Atribut Stok

Untuk atribut Kategori stok terdiri dari 2 nilai yaitu ada dan tidak dimana nilai Entropy masing-masing sebagai berikut:

a. Stok = >65

(Jumlah Kasus = 136, Kasus Laris = 77, Kasus Tidak Laris= 59)

$$\begin{aligned}\text{Entropy (ada)} &= ((-77/136) * \log_2 (77/136) + (-59/136) * \log_2 (59/136)) \\ &= ((-77/136) * \log_2 (0.56617) + (-59/136) * \log_2 (0.41176)) \\ &= 0.98733\end{aligned}$$

Bahwa Nilai Entropy Atribut stok ada tersebut yaitu : 0.98733

b. Stok= <65

(Jumlah Kasus = 559, Kasus Laris =252 50, Kasus Tidak Laris = 307 67)

$$\begin{aligned}\text{Entropy (Tidak)} &= ((-252/559)*\log_2 (252/559) + (-307/559)*\log_2 (307/559)) \\ &= ((-252/559)*\log_2 (0.45080)+(-307/559)*\log_2 (0.5491949)) \\ &= 0.99301\end{aligned}$$

Bahwa Nilai Entropy untuk Atribut stok tidak ada tersebut yaitu : 0.99301

b. Nilai Total Gain

1) Nilai Gain Kategori Merek

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|Merk_i|}{|Total|} * \text{Entropy}(\text{Kategori Merk}_i)$$

$$\begin{aligned} \text{Gain}(\text{Total}, \text{Outlook}) &= 0.997954574 - ((442/695 * 0.994661694) - (103/695 * 0.999932005) - (150/695 * 0.999871757)) \\ &= 0.997954574 - 0.632576 - 0.148191 - 0.218 \\ &= 0.365378574 - 0.366191 \\ &= 0.001387339 \end{aligned}$$

Bahwa Nilai Total Gain untuk Merk tersebut yaitu : 0.001387339

2) Nilai Gain Tingkatan Sparepart

$$\begin{aligned} \text{Gain}(S,A) &= \text{Entropy}(S) - \sum_{i=1}^n \frac{|Tingkatan_i|}{|Total|} * \text{Entropy}(\text{Tingkatan}_i) \\ &= 0.997954574 - ((174/695 * 0.975463037) - (172/695 * 0.99961) - (349/695 * 0.99951))) \\ &= 0.00444 \end{aligned}$$

Bahwa Nilai Total Gain untuk Tingkatan Sparepart tersebut yaitu : 0.00444

3) Nilai Gain Harga

$$\begin{aligned} \text{Gain}(S,A) &= \text{Entropy}(S) - \sum_{i=1}^n \frac{|Harga_i|}{|Total|} * \text{Entropy}(\text{Harga}_i) \\ &= 0.997954574 - ((571/695 * 0.997589305) - (124/695 * 0.999249248)) \\ &= 0.00007 \end{aligned}$$

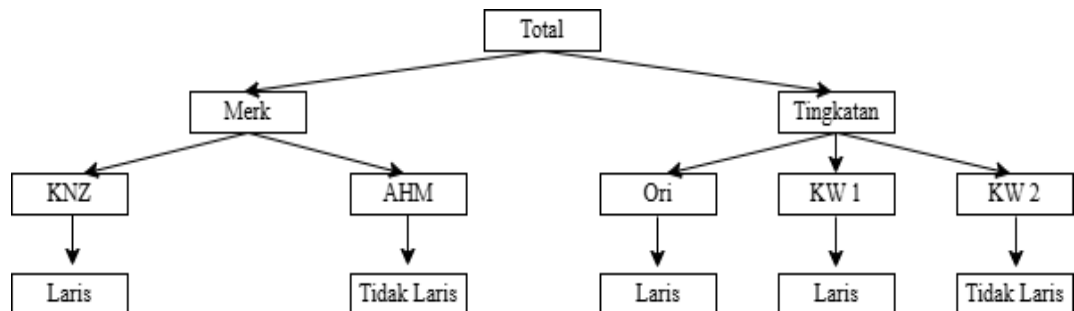
Bahwa Nilai Total Gain untuk Tingkatan Sparepart tersebut yaitu : 0.00007

4) Nilai Gain Stok

$$\begin{aligned} \text{Gain}(S,A) &= \text{Entropy}(S) - \sum_{i=1}^n \frac{|Stok_i|}{|Total|} * \text{Entropy}(\text{Garansi}_i) \\ &= 0.997954574 - ((136/695 * 0.98733) - (559/695 * 0.99301)) \\ &= 0.00606 \end{aligned}$$

Bahwa Nilai Total Gain untuk Durasi tersebut yaitu : 0.00606

Dengan hasil perhitungan tersebut diperoleh Pohon Keputusan adalah sebagai berikut :



Gambar 4. Pohon Keputusan

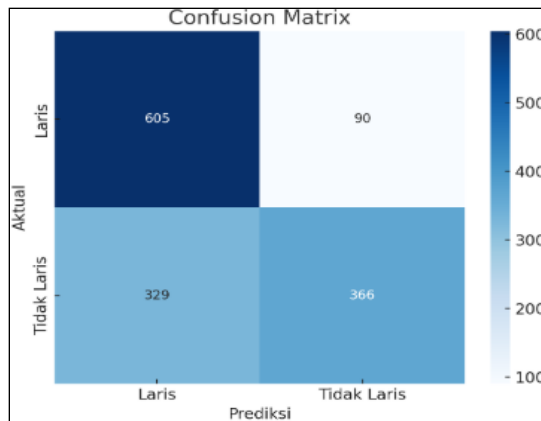
Evaluasi model, menggunakan data uji untuk mengevaluasi kinerja model, biasanya menggunakan metrik seperti akurasi, presisi, dan recall. Jika model tidak memadai, kembali ke langkah Penerapan Algoritma C4.5 untuk penyesuaian atau optimasi. Dalam tahap ini, data klasifikasi produk yang paling banyak terjual yang telah diperoleh akan diolah dengan menggunakan c4.5. Proses analisis dengan algoritma c4.5 dimulai dengan cek data, hingga tahap akhir.

Tabel 2. Perhitungan Nilai

Node	Atribut	Value	Jlh Kasus	Laris	Tidak Laris	Entropy	Gain
1	Total		695	329	366	0.99795	
	Merk						0.001387339
		KNZ	442	202	240	0.99466	
		AHM	103	51	52	0.99993	
		YGP	150	76	74	0.99987	
	Tingkatan						0.00444
		ORI	174	71	103	0.97546	
		KW1	172	88	84	0.99961	
		KW2	349	170	179	0.99952	
	Harga						0.00007
		>250.000	571	269	302	0.99759	
		<250.000	124	60	64	0.99925	
	Stok						0.00606
		>65	136	77	59	0.98733	
		<65	559	252	307	0.99301	

Hasil penelitian, menganalisis dan menginterpretasikan hasil prediksi. Analisis hasil pengujian dilakukan dengan melakukan perhitungan secara manual dengan Confusion Matrix. Perhitungan menggunakan model confusion matrix. Berikut ini merupakan hasil dari perhitungan confusion matrix pada algoritma C4.5:

Aktual / Prediksi	Laris	Tidak Laris
Laris	605 (TP)	90 (FN)
Tidak Laris	329 (FP)	366 (TN)



Gambar 5. Confusion Matrix

Untuk menghitung akurasi, presisi, dan recall, kita memerlukan informasi mengenai hasil klasifikasi: yaitu True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN).

Perhitungan:

- $$\begin{aligned} \text{Akurasi} &= (TP + TN) / (TP + TN + FP + FN) \\ &= (605 + 366) / (605 + 366 + 329 + 90) \\ &= 971 / 1390 \approx 0.6986 \text{ atau } 69.86\% \end{aligned}$$
- $$\begin{aligned} \text{Presisi} &= TP / (TP + FP) \\ &= 605 / (605 + 329) \\ &= 605 / 934 \approx 0.6476 \text{ atau } 64.76\% \end{aligned}$$
- $$\begin{aligned} \text{Recall} &= TP / (TP + FN) \\ &= 605 / (605 + 90) \\ &= 605 / 695 \approx 0.8705 \text{ atau } 87.05\% \end{aligned}$$

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dalam membuat dan mengklasifikasikan produk laris dan tidak laris menggunakan algoritma C4.5, maka dapat ditarik kesimpulan, sebagai berikut :

- 1) Sistem memudahkan dalam mengetahui dan identifikasi produk mana saja yang laris dan tidak laris serta memiliki akurasi penjualan yang tinggi.
- 2) Dengan menerapkan algoritma C4.5 dalam proses klasifikasi produk yang paling banyak terjual, membantu untuk lebih baik dalam mengelola data penjualan, sehingga meminimalisir terjadinya penumpukkan stok barang alias pembelian berlebihan pada barang yang sedikit penjualannya.
- 3) Dalam penelitian ini, dapat disimpulkan bahwa algoritma C4.5 memiliki tingkat akurasi sebesar 69,86%, presisi 64,76% dan recall 87,05% dalam mengklasifikasi produk laris dan tidak laris.

Dengan penelitian yang telah dilakukan ini, diharapkan pada peneliti selanjutnya dapat melakukan pengembangan sistem yang tidak hanya untuk klasifikasi produk laris dan tidak laris saja tapi juga dengan penambahan informasi penting lain yang lebih detail dan akurat.

UCAPAN TERIMA KASIH

Dengan selesainya penyusunan Jurnal ini, Penulis mengucapkan terima kasih yang sedalam-dalamnya untuk seluruh pihak yang membantu atas bimbingan, masukan, dan dukungannya dalam menyelesaikan penelitian ini.

REFERENSI

- [1] Aprianti, Chella, Muhammad Faishal, dan Yuyun Umaidah. 2021. "Penerapan Data Mining Untuk Klasifikasi Penjualan Baju Muslim Dimasa Pandemi Covid-19 Menggunakan Metode Algoritma C4.5 (Studi Kasus: Garaya Collection)." *Jurnal Ilmiah Wahana Pendidikan* 8(1):104–12. doi: 10.5281/zenodo.5816231.
- [2] Desena Damanik, Stevan, dan Muhammad Ihsan Jambak. 2023. "KLIK: Kajian Ilmiah Informatika dan Komputer Klasifikasi Customer Churn pada Telekomunikasi Industri Untuk Retensi Pelanggan Menggunakan Algoritma C4.5." *Media Online* 3(6):1303–9. doi: 10.30865/klik.v3i6.829.
- [3] Asep Hardiyanto Nugroho, dan Toyib Rohimi. 2020. "Perancangan Aplikasi Sistem Pengolahan." *Jutis* 8(1):17749231–5527063.
- [4] Farrah Meirisah, dan Tata Sutabri. 2023. "Analisa Kasus Kejahatan Siber Dengan Menggunakan Visualisasi Data." *Jurnal Informatika Teknologi dan Sains* 5(1):32–37. doi: 10.51401/jinteks.v5i1.2211.
- [5] Kukuh Wahyudi, Adjie, Noor Azizah, dan Heru Saputro. 2022. "Data Mining Klasifikasi Kepribadian Siswa Smp Negeri 5 Jepara Menggunakan Metode Decision Tree Algoritma C4.5." *Journal of Information System and Computer* 2(2):8–13. doi: 10.34001/jister.v2i2.392.
- [6] Lia, Enjel, dan Stefanus Setyo Wibagso. 2022. "Implementasi Algoritma C4 . 5 Untuk Penentuan Produk." 3(2):243–55.
- [7] Maydianto, dan Muhammat Rasid Ridho. 2021. "Rancang Bangun Sistem Informasi Point of Sale Dengan Framework Codeigniter Pada Cv Powershop." *Jurnal Comasie* 02:50–59.
- [8] Perguruan, Aliansi, Tinggi Aperti, dan Menggunakan Algoritma C. 2020. "Jurnal Teknologia Klasifikasi Untuk Memprediksi Pembayaran Kartu Kredit Macet Jurnal Teknologia." 3(1):91–101.
- [9] Putra Eshardiansyah, Raka, Nina Sulistiyowati, Mohamad Jajuli, Teknik Informatika, Universitas HS Singaperbangsa Karawang Jl Ronggowaluyo Kel Puseurjaya Kec Telukjambe Timur Kab Karawang, dan Prov Jawa. 2021. "Algoritma C4.5 Untuk Klasifikasi Jenis Kekerasan pada Anak (Kasus DP3A Kabupaten Karawang)." *Jurnal Sains Komputer & Informatika (J-SAKTI* 5(September):687–96.
- [10] Randa, Dimas Dwi, Yeviki Maisyah Putra, dan Refli Noviardi. 2022. "Perancangan Sistem Informasi Penjualan Berbasis Online Menggunakan Pemrograman PHP Dan Pemodelan UML (Studi Kasus Koningshoes Padang)." *Jrti* 7(4):756–62.
- [11] Salsabilla, Tabitha, dan Sulastri Sulastri. 2022. "Implementasi Algoritma C4.5 Untuk Klasifikasi Produk Laris Sepeda Motor Honda Pada Cv Cendana Motor Cepiring." *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab* 7(2):164–71. doi: 10.36341/rabit.v7i2.2489.
- [12] Supriatmaja, Gede Agus, Putu Mas Yuda Pratama I, Komang Mahendra, Kadek Dwika Darma Widyaputra, Jaya Deva, dan Gede Surya Mahendra. 2022. "Sistem Informasi Perpustakaan Menggunakan Framework Bootstrap Dengan PHP Native dan Database MySQL Berbasis Web Pada SMP Negeri 2 Dawan." *Jurnal Teknologi Ilmu Komputer* 1(1):7–15. doi: 10.56854/jtik.v1i1.30.
- [13] B. S. Sulastio, H. Anggono, and A. D. Putra, "Sistem Informasi Geografis Untuk Menentukan Lokasi Rawan Macet Di Jam Kerja Pada Kota Bandarlampung Pada Berbasis Android," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 104–111, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSl>
- [14] Nahla Naufan, "Perancangan Aplikasi Agenda Harian Guru Piket Pada Smk Taruna Bhakti Berbasis Android," *J. Inform. Kaputama*, vol. 8, no. 1, pp. 8–13, 2024, doi: 10.59697/jik.v8i1.242.

- [15] A. Juansyah, "Pembangunan aplikasi child tracker berbasis assisted-global positioning system (a-gps) dengan platform android," *J. Ilm. Komput. dan Inform.*, vol. 1, no. 1, pp. 1–8, 2015.