

PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA C4.5 UNTUK MEMPREDIKSI PERSEDIAAN STOK BERAS

Smansal Hartawan Totonafu Gulo✉, Olva Andrian Naibaho, Dewi S. Br Rajagukguk,
Gracetia Sari Madu Sihite, Indra M. Sarkis S.

Fakultas Ilmu Komputer, Universitas Methodist Indonesia, Medan, Indonesia

Email: smansalgulo06@gmail.com

ABSTRACT

This research aims to apply data mining using the C4.5 algorithm to predict rice stock supplies at Grosir Media Jaya located in Medan City. Data was obtained through direct interviews with wholesale owners and transaction records for four weeks, which were then processed using Microsoft Office Excel. The data used includes the attributes Type of Rice, Quantity Purchased Per Week, Remaining Quantity Per Week, Category, and Stock Requirements. After data preprocessing, the C4.5 algorithm is applied to build a decision tree model. The analysis results show that the variable "Amount Remaining Per Week" is the main factor in predicting stock availability. The accuracy of the resulting model is 100%, indicating the effectiveness of the C4.5 algorithm in accurately predicting rice supplies. This research makes a significant contribution to stock management in wholesalers by optimizing prediction systems to avoid stock shortages or excesses.

Keyword: Data Mining, Algoritma C4.5, Rice Stock Prediction, Decision Trees, Stock Management.

ABSTRAK

Penelitian ini bertujuan untuk menerapkan data mining dengan menggunakan algoritma C4.5 dalam memprediksi persediaan stok beras di Grosir Media Jaya yang terletak di Kota Medan. Data diperoleh melalui wawancara langsung dengan pemilik grosir dan catatan transaksi selama empat minggu, yang kemudian diolah menggunakan Microsoft Office Excel. Data yang digunakan meliputi atribut Jenis Beras, Jumlah Terbeli Per Minggu, Jumlah Tersisa Per Minggu, Kategori, dan Keterangan Stok. Setelah preprocessing data, algoritma C4.5 diterapkan untuk membangun model pohon keputusan. Hasil analisis menunjukkan bahwa variabel "Jumlah Tersisa Per Minggu" merupakan faktor utama dalam memprediksi ketersediaan stok. Akurasi dari model yang dihasilkan adalah 96.88%, mengindikasikan efektivitas algoritma C4.5 dalam memprediksi persediaan stok beras secara akurat. Penelitian ini memberikan kontribusi signifikan terhadap pengelolaan stok di grosir dengan mengoptimalkan sistem prediksi untuk menghindari kekurangan atau kelebihan stok.

Kata Kunci: Data Mining, Algoritma C4.5, Prediksi Stok Beras, Pohon Keputusan, Pengelolaan Stok.

PENDAHULUAN

Beras adalah kebutuhan pokok bagi masyarakat Indonesia dan memiliki peran vital dalam stabilitas ekonomi serta kesejahteraan sosial. Bagi para pedagang, memastikan ketersediaan stok beras sangat penting untuk mendapatkan keuntungan. Oleh karena itu, pemantauan dan pengelolaan stok beras sangatlah penting. Harga beras terbagi menjadi tiga kategori murah, normal, dan mahal yang dipengaruhi oleh berbagai faktor seperti preferensi konsumen, permintaan pasar, harga, ukuran karung beras, dan tren pasar. Penjualan bertujuan untuk menemukan, mempengaruhi, dan mengarahkan pembeli sehingga mereka dapat menyesuaikan kebutuhan dengan produk yang ditawarkan serta mencapai kesepakatan harga yang adil dan menguntungkan bagi kedua belah pihak. Dalam dunia bisnis, penjualan merupakan aspek paling

penting dalam operasional perusahaan. Semakin tinggi tingkat penjualan, semakin besar keuntungan yang diperoleh, memungkinkan perusahaan untuk bertahan dalam persaingan dan mengembangkan usahanya. Tantangan umum yang dihadapi perusahaan adalah bagaimana memprediksi penjualan produk di masa depan dengan akurat berdasarkan data penjualan sebelumnya.

Media Jaya adalah grosir yang bergerak khusus dalam penjualan beras. Saat ini, Media Jaya mengelola stok berdasarkan beras yang habis tanpa mengetahui secara pasti kapan stok tersebut akan habis. Akibatnya, terkadang stok beras bisa kurang atau berlebihan. Kekurangan stok dapat menyebabkan kerugian karena pelanggan tidak dapat membeli beras yang mereka inginkan, sedangkan kelebihan stok dapat mengakibatkan penumpukan yang tidak efisien dan

potensi pemborosan. Untuk mengoptimalkan keuntungan dan efisiensi, sangat penting bagi Media Jaya untuk memiliki sistem prediksi yang akurat dalam mengelola persediaan beras. Dengan demikian, mereka dapat memastikan ketersediaan stok yang tepat dan menghindari kerugian akibat ketidakseimbangan stok.

Data mining adalah salah satu tahap dalam proses Knowledge Discovery in Database (KDD). Melalui data mining, kita dapat melakukan klasifikasi, prediksi, estimasi, dan memperoleh informasi berharga lainnya dari kumpulan data yang besar. Klasifikasi dalam data mining bisa dilakukan dengan menggunakan algoritma C4.5. Algoritma C4.5 menghasilkan sebuah pohon keputusan yang mudah dipahami dan dimengerti (Salim & Azis, 2018).

Penelitian serupa juga telah dilakukan oleh Akbar Wijaya dan Ahmad Mutatkin Bakti (2023) dengan judul “Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Memprediksi Penjualan Alat – Alat Telekomunikasi Dan IT (Information Technology) Pada PT Julian Adiputra Utama”. Penelitian ini bertujuan untuk memprediksi penyediaan stok yang tepat di masa mendatang. Langkah pertama dalam penelitian ini adalah memilih atribut-atribut yang akan diproses dalam data mining menggunakan alat RapidMiner dan diolah dengan algoritma C4.5. Diharapkan penelitian ini dapat menghasilkan peningkatan penyediaan stok yang memadai sehingga tidak terjadi penumpukan stok. Hasil akhir penelitian ini menunjukkan akurasi sebesar 72.31% (Wijaya & Bakti, 2023).

Selain itu, Aldila Nur Aziziah, Ade Irma Purnamasari, dan Irfan Ali (2024) dengan judul "Penerapan Algoritma C4.5 Untuk Prediksi Stok Bahan Minuman di Cafe Semanis." Penelitian ini memanfaatkan algoritma C4.5 untuk memprediksi stok bahan minuman di Cafe Semanis dengan tujuan meningkatkan efisiensi manajemen stok. Peneliti menggunakan RapidMiner sebagai alat data mining dan menemukan bahwa algoritma C4.5 mampu menghasilkan akurasi sebesar 97.60%. Hasil ini menunjukkan bahwa algoritma ini sangat efektif dalam memprediksi stok yang diperlukan, sehingga dapat digunakan sebagai panduan dalam pengambilan keputusan bisnis yang lebih baik di Cafe Semanis (Nur Aziziah et al., 2024).

Penelitian lainnya oleh Alam Supriyatna, Arif Harbani, Wahyu Hidayat, dan M. Nuril Afton (2023) dengan judul “Implementasi Prediksi Ketersediaan Tanaman Hias Menggunakan Metode Algoritma C4.5”. Penelitian ini bertujuan untuk mengatasi masalah prediksi ketersediaan tanaman hias yang tidak akurat di beberapa toko tanaman hias, yang berdampak

pada berkurangnya keuntungan bisnis. Dengan menggunakan algoritma C4.5, penelitian ini menganalisis data penjualan dan melakukan pengelompokan untuk menentukan tanaman yang cukup atau kurang, menggunakan variabel seperti jenis tanaman, harga, stok, dan musim. Hasil penelitian menunjukkan bahwa aplikasi prototipe yang dibangun memiliki tingkat kelayakan sangat tinggi, dengan akurasi prediksi sebesar 90,97%, serta kualitas sistem yang dinilai sangat layak berdasarkan hasil uji dengan matrix confusion dan kuisioner PSSUQ (Alam et al., 2024).

Penelitian oleh Suwandi et al. (2022) di PT Aerofood Indonesia, dengan judul “Klasifikasi Barang Menggunakan Algoritma C4.5 Dalam Penentuan Prediksi Stok Barang Pada PT Aerofood Indonesia” sebuah perusahaan layanan kuliner penerbangan yang menyediakan sekitar 80.000 set makanan per hari. Penelitian ini menggunakan algoritma C4.5 untuk memprediksi persediaan barang dengan tujuan menjaga stabilitas stok dan mengurangi kesalahan pencatatan yang masih dilakukan secara manual. Dengan menerapkan algoritma C4.5 menggunakan RapidMiner, penelitian ini berhasil mencapai akurasi prediksi sebesar 100%, menunjukkan efektivitas tinggi algoritma dalam mengklasifikasikan barang yang tersedia atau tidak tersedia, sehingga mendukung efisiensi operasional perusahaan (Gustami & Astuti, 2022).

Urgensi kajian ini terletak pada pentingnya prakiraan stok beras yang akurat untuk membantu pengambil kebijakan dalam mengelola stok bagi pedagang Media Jaya. Dalam konteks ini, algoritma C4.5 yang merupakan algoritma pohon keputusan memberikan pendekatan yang andal untuk mengklasifikasikan beras yang dibutuhkan untuk dipasok berdasarkan berbagai variabel yang mempengaruhi penyedia stok beras. Algoritma C4.5 memiliki keunggulan dalam menangani data dengan atribut kategorikal dan kontinu serta nilai yang hilang, menjadikannya alat yang cocok untuk analisis ini.

Tujuan dari penelitian ini adalah untuk mengkaji penyediaan stok beras yang tepat menggunakan algoritma C4.5 agar persediaan stok beras pada grosir Media Jaya tidak kehabisan saat ada pembeli. Kajian ini dilakukan dengan mengumpulkan data pemasokan beras, volume pemasokan beras tiap bulan, kondisi cuaca, dan faktor ekonomi lainnya. Selanjutnya, preprocessing data dan normalisasi data dilakukan untuk menangani missing value. Algoritma C4.5 kemudian diterapkan pada data yang diolah untuk membuat model klasifikasi yang akurat.

METODE PENELITIAN

Penelitian ini menggunakan metode deskriptif, yaitu dengan melakukan studi kasus melalui aktivitas lapangan, observasi, dan wawancara dengan para pemangku kepentingan. Metode deskriptif adalah metode penelitian yang bertujuan untuk menggambarkan atau melukiskan suatu keadaan dengan menggunakan sampel mentah atau data populasi secara sistematis, faktual, dan akurat (Senubekti & Puspita Dewi, 2022). Metode ini juga dapat diartikan sebagai prosedur pemecahan masalah yang dilakukan melalui representasi subjek atau objek penelitian, seperti individu, institusi, komunitas, dan lain-lain, berdasarkan fakta yang ada saat ini (Rizmayanti et al., 2021).



Gambar 1. Tahapan Prediksi dengan Rapid Miner

Data Mining

Penelitian ini menggunakan metode knowledge discovery in database (KDD), adapun langkah-langkah untuk KDD (Septiani, 2017) adalah:

- a. Data Selection
 Kumpulkan satu set data yang dituju dari sejumlah variabel atau bidang data, lalu pilih data tersebut dan fokuskan pada area di mana penemuan akan dilakukan. Data hasil seleksi ini disimpan dalam file terpisah dari database operasional.
- b. Pre-processing/ Cleaning
 Ini adalah proses penghilangan noise. Bagian dari proses pembersihan ini meliputi penghapusan data duplikat, pengecekan data yang tidak konsisten, serta koreksi kesalahan data.
- c. Transformation
 Metode penggabungan data dapat digunakan dalam proses penambangan data. Transformasi sangat terkait dengan jenis atau model informasi yang dapat diambil dari basis data.
- d. Data Mining
 Pemilihan tugas data mining adalah penentuan tujuan dari metode KDD, seperti fiturisasi, klasifikasi, pengelompokan, dan asosiasi.
- e. Interpretation/ Evolution
 Pada tahap ini, penerjemah model menggunakan data mining untuk menghasilkan informasi tentang model dengan cara yang mudah dipahami. Perlu dievaluasi apakah pola yang ditemukan konsisten dengan fakta yang ada dan relevan dalam konteks aplikatif.

Algoritma C4.5

Algoritma klasifikasi data yang menggunakan pohon keputusan adalah C4.5. Pohon keputusan dalam algoritma C4.5 melibatkan beberapa langkah, termasuk memilih sebuah atribut sebagai akar, membuat cabang untuk setiap nilai, dan membagi kasus ke dalam cabang-cabang tersebut (Dewi et al., 2020).

Langkah – Langkah dalam mencari algoritma C4.5 adalah (Budiman & Ramadina, 2015):

- a. Pemilihan atribut sebagai *root*
- b. Pembuatan cabang untuk setiap nilai.
- c. Pemisahan kasus dalam cabang.
- d. Mengulangi langkah untuk setiap cabang sampai semua kasus dalam cabang memiliki kelas yang sama.

Rumus dalam persamaan 1 digunakan untuk menghitung gain. Dimana: S: himpunan kasus, A: atribut/fitur, N: jumlah partisi atribut A, |Si|: jumlah kasus pada partisi ke-i, dan |S|: jumlah kasus dalam S.

Sementara itu, persamaan 2 digunakan untuk penghitungan entropy. Dimana S himpunan kasus, A: fitur, N: jumlah partisi S, pi: proporsi dari Si terhadap S (Bachtiar & Mahradianur, 2023).

$$Gain(S, A) = Entropy(S)$$

$$= \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \dots\dots (1)$$

$$Entropy(S)$$

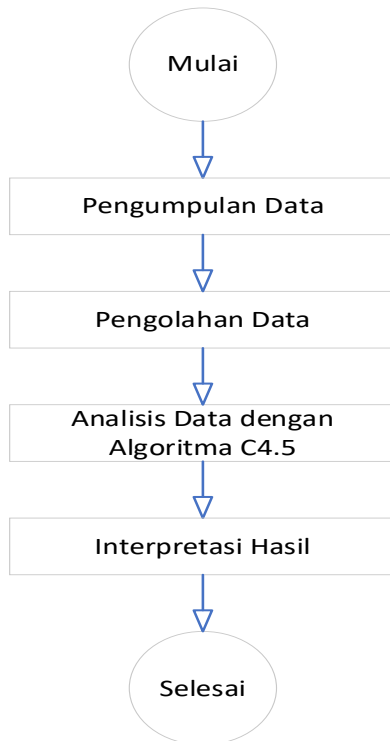
$$= \sum_{i=1}^n - p_i * \log_2 p_i \dots\dots\dots (2)$$

RapidMiner

RapidMiner adalah perangkat lunak ilmu data yang dikembangkan oleh perusahaan yang sama, menyediakan lingkungan terintegrasi untuk pembelajaran mesin, pembelajaran mendalam, penambangan teks, dan analisis prediktif. Program ini mendukung semua fase dalam proses pembelajaran otomatis, termasuk persiapan data, visualisasi hasil, validasi, dan pengoptimalan (Meila Azzahra Sofyan et al., 2023). RapidMiner merupakan bagian dari perangkat lunak pengolahan data yang memungkinkan pembuatan model dari kumpulan data besar dengan menerapkan prinsip-prinsip dan algoritma data mining. Pendekatan ini menggabungkan metode statistika, kecerdasan buatan, dan basis data. RapidMiner memberikan kemudahan bagi pengguna untuk melakukan komputasi data skala besar dengan menggunakan operator-operator yang tersedia (Irnanda et al., 2022).

Tahapan Penelitian

Tahapan dari metode penelitian yang dilakukan peneliti disajikan pada gambar 2 (Andarista & Jananto, 2022).



Gambar 2. Tahapan Penelitian

Berikut merupakan penjelasan gambar 2

1. Pengumpulan Data, dengan teknik melakukan wawancara terhadap pemilik toko grosir beras Media Jaya.
2. Pengolahan Data, Data yang diperoleh masih berupa data mentah yang belum siap untuk diolah. Beberapa data perlu diuji kelayakannya dan dilakukan perbaikan agar data tersebut siap untuk dianalisis.
3. Analisa data, data yang telah diolah lalu dilakukan analisa dengan menggunakan metode Algoritma C4.5.
4. Interpretasi hasil, analisis hasil prediksi dan interpretasikan pola yang didapatkan

Pengumpulan Data

Pengumpulan data adalah kegiatan yang bertujuan untuk mengumpulkan catatan dan informasi yang relevan dengan penelitian (Millah et al., 2023). Dalam penelitian ini, metode yang digunakan meliputi wawancara, observasi, dan studi pustaka. Wawancara dilakukan dengan meminta informasi secara langsung dari pihak grosir Media Jaya untuk memperoleh data yang diperlukan. Observasi dilakukan dengan cara mengamati secara langsung situasi dan aktivitas di

grosir Media Jaya guna mendapatkan informasi yang lebih mendalam dan akurat. Sedangkan studi pustaka melibatkan pengumpulan, penelitian, dan analisis dokumen-dokumen elektronik maupun tertulis, termasuk gambar, buku referensi, jurnal, serta berbagai media lain yang berkaitan dengan objek penelitian (Abdullah et al., 2022).

Data tersebut kemudian dimasukkan dan diolah menggunakan *Microsoft Office Excel*. Proses pengolahan data mencakup hasil wawancara dan transaksi selama empat minggu, dengan fokus pada atribut seperti Jenis Beras, Jumlah Terbeli Per Minggu, Jumlah Tersisa Per Minggu, Kategori, dan Keterangan Stok. Hasil dari proses *preprocessing* data akumulasi adalah sebagai berikut:

Tabel 1. Data Beras Grosir Media Jaya

Jenis Beras	Jumlah Terbeli	Jumlah Tersisa	Harga	Kategori	Keterangan Stok	Keterangan Pasok
Salak 5 kg	120	100	73.000	Normal	Lebih	Tidak Pasok
Salak 10 kg	125	100	146.000	Mahal	Cukup	Pasok
Pandan 5 kg	135	125	72.000	Normal	Kurang	Pasok
Pandan 10 kg	135	100	144.000	Normal	Lebih	Tidak Pasok
Singa 5 kg	125	120	75.000	Normal	Lebih	Tidak Pasok
Singa 10 kg	120	120	150.000	Mahal	Lebih	Tidak Pasok
Mawar 5 kg	125	130	75.000	Normal	Cukup	Pasok
Mawar 10 kg	135	120	150.000	Mahal	Kurang	Pasok
s/d
Mawar 5 kg	120	115	75.000	Normal	Lebih	Tidak Pasok

Variabel Harga dibagi menjadi 3 kategori:

- a. Jika berat 5 kg dan harga dibawah Rp. 70.000 maka tergolong Murah dan Jika berat 10 kg dan harga dibawah Rp. 140.000 maka tergolong Murah.
- b. Jika berat 5 kg dan harga Rp. 70.000 – Rp.75.000 maka tergolong Normal dan Jika berat 10 kg dan harga Rp. 140.000 – Rp. 145.000 maka tergolong Normal.
- c. Jika berat 5 kg dan harga diatas Rp. 75.000 maka tergolong Mahal dan Jika berat 10 kg dan harga diatas Rp. 145.000 maka tergolong Mahal.

Tabel 2. Klasifikasi pada Atribut

No	Atribut	Class
1	Jenis Beras	1 – 5 kg (Kecil)
		6 – 10 kg (Sedang)
		11 – 15 (Besar)
2	Jumlah Terbeli	110 – 120 (Rendah)
		121 – 130 (Sedang)
		131 – 140 (Tinggi)
3	Jumlah Tersisa	100 – 120 (Ada)
		121 – 140 (Tidak Ada)

Tabel 3. Data Set

No	Jenis Beras	Jumlah Terbeli	Jumlah Tersisa	Kategori	Keterangan Stok	Keterangan Pasok
1	Salak 5 kg	Rendah	Ada	Normal	Lebih	Tidak Pasok
2	Salak 10 kg	Sedang	Ada	Mahal	Cukup	Pasok
3	Pandan 5 kg	Tinggi	Tidak Ada	Murah	Kurang	Pasok
4	Pandan 10 kg	Tinggi	Ada	Murah	Lebih	Tidak Pasok
5	Singa 5 kg	Sedang	Ada	Normal	Lebih	Tidak Pasok
...
33	Mawar 5 kg	Rendah	Ada	Normal	Lebih	Tidak Pasok

Tabel 4. Variabel Penentu Keputusan

Variabel	Keterangan
Jenis Beras	Atribute
Jumlah Terbeli	Atribute
Jumlah Tersisa	Atribute
Kategori	Atribute
Keterangan Stok	Atribute
Keterangan Pasok	Label

Tabel 5. Hasil Perhitungan pada Dataset

Total Data	Pasok	Tidak Pasok	Hasil
32	17	15	0.99718

$Entropy(S) = \sum - p_i * \log_2 n \ i=1 \ p_i$
 Menghitung Total : $(-17/32)*\log_2(17/32) + (-15/32)*\log_2(15/32) = 0.99718$

Tabel 6. Nilai Variabel Jenis Beras

Jenis Beras	Keterangan	Jumlah	Total
Kecil (5Kg)	Pasok	8	16
	Tidak Pasok	8	
Sedang (10 Kg)	Pasok	9	16
	Tidak Pasok	7	

Variabel Jenis Beras :

$$Entropy(S) = \sum - p_i * \log_2 n \ i=1 \ p_i$$

$$Entropy(Kecil) = (-8/16)*\log_2(8/16) + (-8/16)*\log_2(8/16) = 1$$

$$Entropy(Sedang) = (-9/16)*\log_2(9/16) + (-7/16)*\log_2(7/16) = 0.98869$$

Gain Jenis Beras :

$$= 0.99718 - ((16/32)*1) + (16-32)*0.98869)) = 0.99152$$

Tabel 7. Nilai Variabel Jumlah Terbeli

Jumlah Terbeli	Keterangan	Jumlah	Total
Rendah	Pasok	0	8
	Tidak Pasok	8	
Sedang	Pasok	9	12
	Tidak Pasok	3	
Tinggi	Pasok	8	12
	Tidak Pasok	4	

Variabel Jumlah Terbeli:

$$Entropy(S) = \sum - p_i * \log_2 n \ i=1 \ p_i$$

$$Entropy(Rendah) = (-0/8)*\log_2(0/8) + (-8/8)*\log_2(8/8) = 0$$

$$Entropy(Sedang) = (-9/12)*\log_2(9/12) + (-3/12)*\log_2(3/12) = 0.81127$$

$$Entropy(Tinggi) = (-8/12)*\log_2(8/12) + (-4/12)*\log_2(4/12) = 0.91829$$

Gain Jumlah Terbeli:

$$= 0.99718 - ((8/32)*0) + (12/32)*0.81127 + (12/32)*0.91829 = 1.64576$$

Tabel 8. Nilai Variabel Jumlah Tersisa

Jumlah Tersisa	Keterangan	Jumlah	Total
Ada	Pasok	11	26
	Tidak Pasok	15	
Tidak Ada	Pasok	6	6
	Tidak Pasok	0	

Variabel Jumlah Tersisa :

$$Entropy(S) = \sum - p_i * \log_2 n \ i=1 \ p_i$$

$$Entropy(Ada) = (-11/26)*\log_2(11/26) + (-15/26)*\log_2(15/26) = 0.98285$$

$$Entropy(Tidak Ada) = (-6/6)*\log_2(6/6) + (-0/6)*\log_2(0/6) = 0$$

Gain Jumlah Stok Masuk:

$$= 0.99718 - ((26/32)*0.98285) + (6/32)*0 = 0.19861$$

Tabel 9. Nilai Variabel Kategori

Kategori	Keterangan	Jumlah	Total
Murah	Pasok	8	12
	Tidak Pasok	4	
Normal	Pasok	2	8
	Tidak Pasok	6	
Mahal	Pasok	7	12
	Tidak Pasok	5	

Variabel Kategori :

$$Entropy (S) = \sum - p_i * \log_2 n \ i=1 \ p_i$$

$$Entropy (Murah) = (-8/12)*\log_2(8/12) + (-4/12)*\log_2(4/12) = 0.91829$$

$$Entropy (Normal) = (-2/8)*\log_2(2/8) + (-6/8)*\log_2(6/8) = 0.81127$$

$$Entropy (Mahal) = (-7/12)*\log_2(7/12) + (-5/12)*\log_2(5/12) = 0.97986$$

Gain Kategori :

$$= 0.99718 - ((12/32)* 0.91829) + (8/32)* 0.81127 + (12/32)* 0.97986$$

$$= 1.22308$$

Tabel 10. Nilai Variabel Keterangan Stok

Keterangan Stok	Keterangan	Jumlah	Total
Lebih	Pasok	0	14
	Tidak Pasok	14	

Cukup	Pasok	9	10
	Tidak Pasok	1	
Kurang	Pasok	8	8
	Tidak Pasok	0	

Variabel Keterangan Stok:

$$Entropy (S) = \sum - p_i * \log_2 n \ i=1 \ p_i$$

$$Entropy (Lebih) = (-0/14)*\log_2(0/14) + (-14/14)*\log_2(14/14) = 0$$

$$Entropy (Cukup) = (-9/10)*\log_2(9/10) + (-1/10)*\log_2(1/10) = 0.46899$$

$$Entropy (Kurang) = (-8/8)*\log_2(8/8) + (-0/8)*\log_2(0/8) = 0$$

Gain Keterangan Stok:

$$= 0.99718 - ((14/32)*0) + (10/32)* 0.46899 + (8/32)*0$$

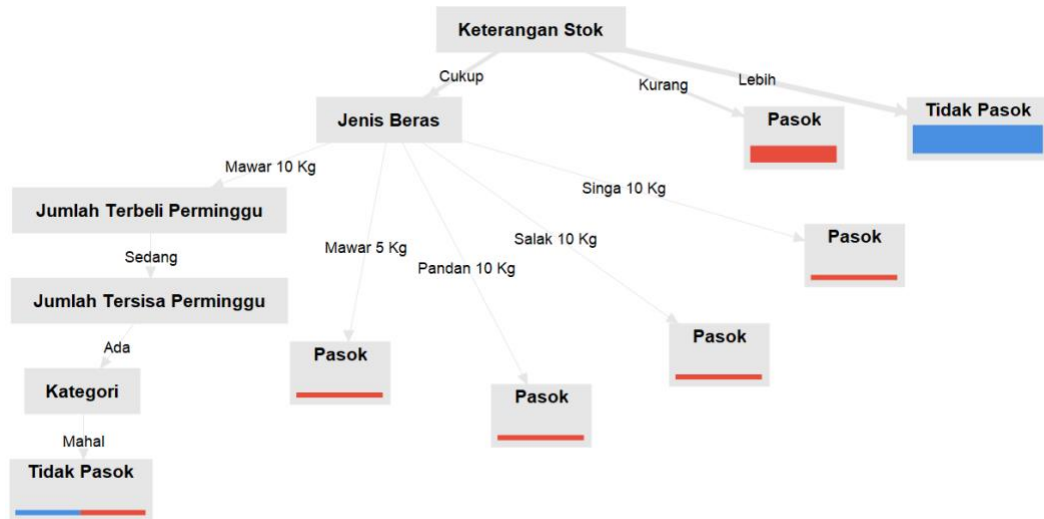
$$= 1.14373$$

Tabel 11. Entropy dan Gain

Variabel	Nilai	Jumlah Data	Pasok	Tidak Pasok	Entropy	Gain
Total		32	17	15	0.99718	
Jenis Beras	Kecil	16	8	8	1	0.99152
	Sedang	16	9	7	0.98869	
Jumlah Terbeli	Rendah	8	0	8	0	1.64576
	Sedang	12	9	3	0.81127	
	Tinggi	12	8	4	0.91829	
Jumlah Tersisa	Ada	26	11	15	0.98285	0.19861
	Tidak Ada	6	6	0	0	
Kategori	Murah	12	8	4	0.91829	1.22308
	Normal	8	2	6	0.81127	
	Mahal	12	7	5	0.97986	
Keterangan Stok	Lebih	14	0	14	0	1.14373
	Cukup	10	9	1	0.46899	
	Kurang	8	8	0	0	

HASIL DAN PEMBAHASAN

Data yang telah dipreproses kemudian diolah menjadi pohon keputusan menggunakan software RapidMiner untuk menentukan stok yang perlu dipasok oleh toko grosir beras. Pohon keputusan yang dihasilkan adalah sebagai berikut:



Gambar 3. Pohon Keputusan Prediksi Kebutuhan Stok

Berdasarkan hasil analisis pohon keputusan (decision tree) untuk klasifikasi faktor ketersediaan stok di grosir Media Jaya, atribut yang paling berpengaruh dalam memprediksi ketersediaan stok adalah variabel "Jumlah Tersisa Per Minggu." Variabel ini ditempatkan sebagai simpul akar, sebagaimana ditunjukkan pada gambar berikut:

Tree

```

Keterangan Stok = Cukup
| Jenis Beras = Mawar 10 Kg
| | Jumlah Terbeli Perminggu = Sedang
| | | Jumlah Tersisa Perminggu = Ada
| | | | Kategori = Mahal: Tidak Pasok (Tidak Pasok=1, Pasok=0)
| Jenis Beras = Mawar 5 Kg: Pasok (Tidak Pasok=0, Pasok=2)
| Jenis Beras = Pandan 10 Kg: Pasok (Tidak Pasok=0, Pasok=2)
| Jenis Beras = Salak 10 Kg: Pasok (Tidak Pasok=0, Pasok=2)
| Jenis Beras = Singa 10 Kg: Pasok (Tidak Pasok=0, Pasok=2)
Keterangan Stok = Kurang: Pasok (Tidak Pasok=0, Pasok=8)
Keterangan Stok = Lebih: Tidak Pasok (Tidak Pasok=14, Pasok=0)
    
```

Gambar 4. Hasil Pohon Keputusan Prediksi Kebutuhan Stok Berupa Teks

Pohon keputusan yang dihasilkan oleh algoritma C4.5 pada gambar 2 untuk memprediksi keputusan pasokan beras berdasarkan kondisi stok dan

- jenis beras dapat dijelaskan sebagai berikut:
1. Jika Keterangan Stok sama dengan Cukup dan Jenis Beras adalah Mawar 10 Kg, maka Tidak Pasok.
 2. Jika Keterangan Stok sama dengan Cukup dan Jenis Beras adalah Mawar 5 Kg, maka Pasok.
 3. Jika Keterangan Stok sama dengan Cukup dan Jenis Beras adalah Pandan 10 Kg, maka Pasok.
 4. Jika Keterangan Stok sama dengan Cukup dan Jenis Beras adalah Salak 10 Kg, maka Pasok.
 5. Jika Keterangan Stok sama dengan Cukup dan Jenis Beras adalah Singa 10 Kg, maka Pasok.
 6. Jika Keterangan Stok sama dengan Kurang, maka Pasok.
 7. Jika Keterangan Stok sama dengan Lebih, maka Tidak Pasok.

Berdasarkan hasil perhitungan, akurasi yang diperoleh adalah 96.88%. Hal ini menunjukkan bahwa metode C4.5 sangat efektif dalam memprediksi penyediaan stok secara akurat. Berikut adalah hasil perhitungan yang dilakukan menggunakan software RapidMiner.

accuracy: 96.88%

	true Tidak Pasok	true Pasok	class precision
pred. Tidak Pasok	15	1	93.75%
pred. Pasok	0	16	100.00%
class recall	100.00%	94.12%	

Gambar 5. Hasil Akurasi Penerapan Algoritma C4.5

KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma C4.5 dalam memprediksi persediaan stok beras di Grosir Media Jaya. Dengan menggunakan data transaksi dan wawancara selama empat minggu, serta

penerapan algoritma C4.5, model yang dihasilkan menunjukkan akurasi 96.88%. Hal ini menegaskan bahwa algoritma C4.5 adalah alat yang efektif untuk memprediksi kebutuhan stok beras secara akurat. Penelitian ini juga menyoroti pentingnya sistem

prediksi yang tepat untuk mengelola persediaan dan meminimalkan kerugian akibat ketidakseimbangan stok. Temuan ini dapat diterapkan untuk meningkatkan efisiensi pengelolaan stok dan profitabilitas grosir.

DISEMINASI

Artikel ini telah diseminasikan pada Seminar Nasional Teknologi Informasi dan Komunikasi (SEMNASTIK) APTIKOM Tahun 2024 yang diselenggarakan oleh Universitas Methodist Indonesia pada tanggal 24-26 Oktober 2024.

DAFTAR PUSTAKA

- Abdullah, R. W., Hartanti, D., Permatasari, H., Septyanto, A. W., & Bagaskara, Y. A. (2022). Penerapan Data Mining untuk Memprediksi Jumlah Produk Terlaris Menggunakan Algoritma Naive Bayes Studi Kasus (Toko Prapti). *Jurnal Ilmiah Informatika Global*, 13(1), 20–27.
<https://doi.org/10.36982/jiig.v13i1.2060>
- Alam, A. S., Harbani, A., Hidayat, W., & Afton, M. N. (2024). Implementasi Prediksi Ketersediaan Tanaman Hias Menggunakan Metode Algoritma C4.5. *Explore*, 14(1), 42–46.
<https://doi.org/10.35200/ex.v14i1.111>
- Andarista, R. R., & Jananto, A. (2022). Penerapan Data Mining Algoritma C4.5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor. *Jurnal Tekno Kompak*, 16(2), 29.
<https://doi.org/10.33365/jtk.v16i2.1525>
- Bachtiar, L., & Mahradianur, M. (2023). Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai. *Jurnal Informatika*, 10(1), 28–36.
<https://doi.org/10.31294/inf.v10i1.15115>
- Budiman, I., & Ramadina, R. (2015). Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi. *Ijccs*, x, No.x(1), 1–5.
- Dewi, K. R., Mauladi, K. F., & Masruroh. (2020). Analisa Algoritma C4.5 untuk Prediksi Penjualan Obat Pertanian di Toko Dewi Sri. *Seminar Nasional Inovasi Teknologi*, 25, 109–114.
- Gustami, E., & Astuti, K. P. (2022). Klasifikasi Barang Menggunakan Algoritma C4. 5 Dalam Penentuan Prediksi Stok Barang Pada Pt Aerofood Indonesia. *Reputasi: Jurnal Rekayasa Perangkat Lunak*, 3(2), 12–18.
- Inranda, K. F., Windarto, A. P., & Damanik, I. S. (2022). Optimasi Particle Swarm Optimization Pada Peningkatan Prediksi dengan Metode Backpropagation Menggunakan Software RapidMiner. *JURIKOM (Jurnal Riset Komputer)*, 9(1), 122.
<https://doi.org/10.30865/jurikom.v9i1.3836>
- Meila Azzahra Sofyan, F., Voutama, A., & Umaidah, Y. (2023). Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Paru-Paru Menggunakan Rapidminer. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(2), 1409–1415.
<https://doi.org/10.36040/jati.v7i2.6810>
- Millah, A. S., Apriyani, Arobiah, D., Febriani, E. S., & Ramdhani, E. (2023). Analisis Data dalam Penelitian Tindakan Kelas. *Jurnal Kreativitas Mahasiswa*, 1(2), 140–153.
- Nur Aziziah, A., Irma Purnamasari, A., & Ali, I. (2024). Penerapan Algoritma C4.5 untuk Prediksi Stok Bahan Minuman di Cafe Semanis. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 292–295.
<https://doi.org/10.36040/jati.v8i1.8347>
- Rizmayanti, A. I., Hidayati, N., Nugraha, F. S., & Gata, W. (2021). Penerapan Data Mining Untuk Memprediksi Kompetensi Siswa Menggunakan Metode Decission Tree (Studi Kasus Smk Multicomp Depok). *Swabumi*, 9(1), 9–18.
- Salim, Y., & Azis, H. (2018). Kinerja Metode C4.5 dalam Penyaluran Bantuan Dana Bencana. *Prosiding SAKTI (Seminar Ilmu Komputer Dan Teknologi Informasi)*, 84–87.
- Senubekti, M. A., & Puspita Dewi, L. A. (2022). Prinsip Klasifikasi Dan Data Mining Dengan Algoritma C4.5. *Nuansa Informatika*, 16(2), 87–93. <https://doi.org/10.25134/nuansa.v16i2.5834>
- Septiani, D. (2017). Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis. *Jurnal Pilar Nusa Mandiri*, 13(1), 76–84.
- Wijaya, A., & Bakti, A. M. (2023). Implementasi Data Mining Menggunakan Algoritma C4. 5 Untuk Memprediksi Penjualan Alat-Alat Telekomunikasi Dan IT (Information Technology) Pada PT Julian Adiputra Utama. *Jupiter*, 15, 597–606.