

PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI STOK BAHAN MINUMAN DI CAFE SEMANIS

Aldila Nur Aziziah¹, Ade Irma Purnamasari², Irfan Ali³

¹⁻² Teknik Informatika, STMIK IKMI Cirebon

³ Rekayasa Perangkat Lunak, STMIK IKMI Cirebon

Jl. Perjuangan No. 10 B Majasem Kec. Kesambi Kota Cirebon, Tlp. 0231) 490480 - 490481

aldilanura0@gmail.com

ABSTRAK

Dalam era populernya minuman kopi, coffee shop menjadi bisnis yang diminati, termasuk Cafe Semanis yang berhasil menarik banyak pelanggan. Namun, kendala dalam manajemen stok bahan masih menggunakan media kertas yang menyebabkan penumpukan berkas dan penumpukan berkasnya tidak dijadikan analisa *prediksi* untuk mempersiapkan stok dimasa yang akan datang, hal tersebut dapat mengancam tingkat keberhasilan penjualan. Penelitian ini bertujuan memprediksi stok bahan yang perlu ditambahkan untuk meningkatkan penjualan dengan menggunakan *data mining* dan *Algoritma C4.5* melalui *Rapidminer*. Dalam penelitian ini, Cafe Semanis yang belum memiliki sistem pendukung keputusan diharapkan dapat merancang strategi bisnis yang lebih efektif. Hasil penelitian menunjukkan tingkat akurasi *Algoritma C4.5* mencapai 97.60%, memberikan bukti bahwa algoritma ini dapat digunakan untuk *prediksi* stok bahan minuman. Implikasi dari penelitian ini memberikan panduan bagi pemilik cafe dalam meningkatkan efisiensi manajemen stok, mendukung keputusan strategis, dan pada gilirannya, meningkatkan tingkat keberhasilan penjualan produk minuman.

Kata kunci : Data Mining, Algoritma C4.5, Prediksi, Rapidminer

1. PENDAHULUAN

Usaha coffee shop saat ini menjadi salah satu usaha yang paling banyak diminati di Bandung. Hal ini ditandai dengan maraknya coffee shop dimana-mana. Masalah pengendalian persediaan adalah salah satu hal terpenting yang harus dihadapi. Tanpa persediaan, perusahaan berisiko tidak dapat memenuhi permintaan pelanggan secara tepat waktu. [1]

Di Cafe Semanis agar tetap ramai pembeli, persediaan stok bahan minuman perlu dijaga setiap harinya, ketersediaan stok bahan minuman yang tidak dikelola dengan baik akan berdampak pada coffee shop itu sendiri. Data stok bahan minuman sebelumnya di cafe semanis di teliti untuk menghasilkan pengetahuan baru melalui metode data mining. Memprediksi stok bahan minuman di cafe semanis menggunakan algoritma C4.5 untuk mengetahui stok bahan mana yang harus stok dan tidak stok. Algoritma C4.5 adalah membentuk pohon keputusan untuk menghasilkan sebuah keputusan.

Penelitian terdahulu yang dilakukan oleh Thosca Penerapan Algoritma C4.5 Dalam Memprediksi Pembelian Rokok Elektrik Pada Toko Silvape. Metode yang digunakan adalah Algoritma C4.5. Hasil yang didapatkan cukup tinggi dengan nilai akurasi sebesar 97,22%. [2] Tujuan dari penelitian ini adalah mengetahui nilai akurasi yang dihasilkan dari Algoritma C4.5, menentukan atribut dan melakukan prediksi untuk mengetahui stok bahan mana yang harus stok dan tidak stok.

2. TINJAUAN PUSTAKA

2.1. Stok

Stok merupakan persediaan barang yang disimpan dan akan dikeluarkan untuk dijual ataupun

untuk membantu proses produksi dalam sebuah perusahaan. [3]

2.2. Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma yang digunakan dalam data mining untuk membangun model pohon keputusan. Algoritma C4.5 menggunakan konsep pembelajaran mesin untuk menghasilkan model prediksi berdasarkan data yang diberikan. [4] Secara umum, rumus akurasi Algoritma C4.5 adalah:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

TP = True Positive TN = True Negative
FP = False Positive FN = False Negative

2.3. Pohon Keputusan

Pohon keputusan adalah metode pengambilan keputusan yang menyusun setiap opsi menjadi bentuk yang bercabang. Dengan begitu membuat metode ini disebut pohon keputusan. Karena bentuknya seperti pohon dan mempunyai banyak cabang dibagian ranting maupun bagian akarnya. [5]

2.4. Akurasi

Akurasi adalah ukuran yang menentukan tingkat kemiripan antara hasil pengukuran dengan nilai yang sebenarnya diukur. Dalam bidang pengukuran, akurasi lebih dikhususkan pada ketidakpastian pengukuran dari alat ukur. Akurasi ditetapkan untuk mengetahui seberapa besar tingkat kesalahan pengukuran yang dapat terjadi pada suatu alat ukur. Pernyataan terhadap akurasi dalam bentuk persentase dengan skala penuh berdasarkan jenis alat ukur yang digunakan. [6]

2.5. Prediksi

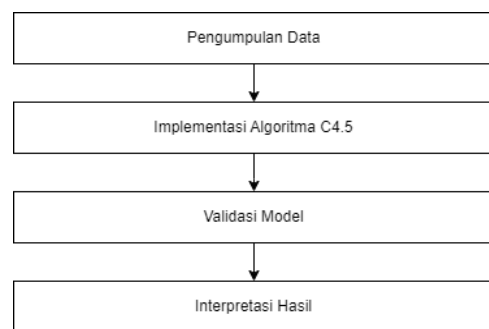
Prediksi termasuk kedalam kelompok data mining, prediksi adalah awal dari suatu proses pengambilan suatu keputusan. Prediksi dapat dicoba dengan cara pengambilan data sebelumnya serta memproyeksikan ke masa yang akan datang. [7]

2.6. Rapidminer

RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner pilihan perangkat lunak sebagai *tools* data mining dengan pengambilan informasi. [8] RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik.

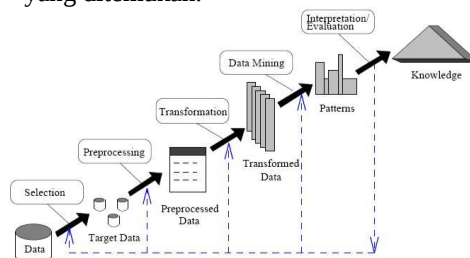
3. METODE PENELITIAN

Penelitian ini menggunakan metode analisis data menggunakan proses KDD dan algoritma C4.5. Seperti ditunjukkan pada Gambar 1 dan 2.



Gambar 1. Tahapan Metode Penelitian Algoritma C4.5

- Pengumpulan Data**
Mengumpulkan data stok bahan minuman 1 bulan terakhir dari Cafe Semanis.
- Implementasi Algoritma C4.5**
Dataset yang telah disiapkan terapkan algoritma C4.5 atau masukan ke rapidminer.
- Validasi Model**
Evaluasi performa model dengan menggunakan data uji.
- Interpretasi Hasil**
Analisis hasil prediksi dan interpretasikan pola yang ditemukan.



Gambar 2. Proses KDD

3.1. Selection

Selection didefinisikan sebagai proses di mana data yang relevan dengan analisis diputuskan dan diambil dari pengumpulan data. Pada tahap ini, data yang digunakan sebagai bahan penelitian adalah data yang diambil dari bulan oktober 2023 lalu dipilih atribut sesuai yang dibutuhkan.

3.2. Preprocessing

Preprocessing adalah sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Saat sebelum proses prediksi, diperlukan proses preprocessing yang mencakup antara lain membuang duplikasi data, menghapus data yang kosong, dan memperbaiki kesalahan pada data.[9]

3.3. Transformation

Pada tahap ini merupakan proses transformasi pada data yang sudah melalui seleksi untuk di proses ke data mining. Bertujuan untuk memudahkan koordinasi data yang diproses oleh algoritma dan alat yang digunakan dalam penelitian.

3.4. Data Mining

Data mining merupakan proses mencari pola ataupun data menarik dalam informasi terpilih dengan memakai metode ataupun tata cara tertentu. Algoritma yang digunakan dalam penelitian ini adalah Algoritma C4.5 yang termasuk dalam algoritma pohon keputusan / prediksi. [10]

3.5. Evaluation

Pada tahap evaluasi digunakan untuk mengevaluasi hasil dari algoritma C4.5 untuk mengetahui kesimpulan dari proses data mining. Nilai performance yang digunakan adalah Performance Vector.

3.6. Knowledge

Knowledge adalah proses paling akhir dari proses KDD, data-data yang sudah diproses divisualisasikan agar lebih mudah dipahami oleh pengguna dan diharapkan bisa diambil tindakan berdasarkan analisis.

4. HASIL DAN PEMBAHASAN

4.1 Data Selection

Pada tahap ini dilakukan data *selection* bertujuan untuk memilih data yang akan digunakan dalam *tools* RapidMiner, tahapan seleksi dilakukan di *Microsoft Excel*. Pada proses penelitian menggunakan data stok bahan di bulan Oktober 2023. Hasil dari seleksi data stok bahan diketahui terdapat 6 atribut yaitu atribut nama bahan, satuan, stok awal, bahan keluar, stok akhir dan keterangan. Bisa dilihat pada tabel 1 sebagai berikut :

Tabel 1. Data Selection

Nama Bahan	Satuan	Stok Awal	Barang Keluar	Stok Akhir	Keterangan
THAI TEA	pak	2	1	1	stok
GREEN TEA	pak	1	0	1	stok
SKM	kaleng	14	3	11	tidak stok
TARO	gram	750	0	750	tidak stok
FRESH MILK	mili	17.000	10.350	6650	stok
LEMON TEA	gram	1.700	100	1600	tidak stok
FN	kaleng	9	1	8	stok
BEANS ESPRESSO	gram	250	10	240	stok
CREAMER	gram	2.400	400	2000	tidak stok
GULA PUTIH	gram	2.000	500	1500	tidak stok
CRÈME BRULEE	mili	650	30	620	stok
HAZELNUT	mili	500	40	460	stok
CARAMEL	mili	1.250	0	1250	tidak stok
COKLAT	gram	900	90	810	tidak stok
STRAWBERRY	botol	7	0	7	tidak stok
RED VELVET	gram	1.000	180	820	tidak stok
MILO	gram	1.500	175	1325	tidak stok
CUP MINUMAN	pcs	400	40	360	tidak stok
KLEPON	gram	805	0	805	tidak stok
CHARCOAL	gram	400	0	400	stok
MANGO SYRUP	botol	3	0	3	stok
LYCHEE SYRUP	botol	3	0	3	stok
SODA	pcs	23	0	23	tidak stok
GULA AREN LIQUID	jerigen	1.500	600	900	tidak stok
PEPPERMINT	mili	730	30	700	stok
LEMON	gram	1.500	300	1200	tidak stok
...	...	...	...	...	...
SODA	pcs	45	5	40	stok
GULA AREN LIQUID	jerigen	5.500	100	5400	tidak stok
PEPPERMINT	mili	600	50	550	stok
LEMON	gram	800	300	500	tidak stok
V60	gram	100	100	0	stok

4.2 Data Preprocessing

Dalam tahap ini dilakukan Preprocessing yaitu proses Cleaning data yang duplikat atau kosong sebelum ke proses selanjutnya, karena data yang peneliti gunakan tidak ada data yang duplikat atau hilang maka tidak dilakukan penghapusan.

4.3 Data Transformation

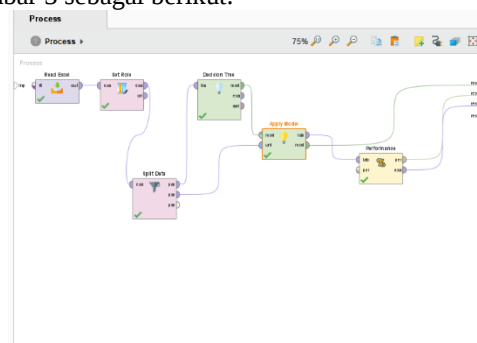
Pada tahap ini dilakukan proses transformasi pada data yang sudah melalui seleksi untuk di proses ke data mining. Bertujuan untuk memudahkan koordinasi data yang diproses oleh algoritma dan tools yang digunakan dalam penelitian yaitu rapidminer. Bisa dilihat pada Gambar 2 sebagai berikut :

	Nama Bahan	Satuan	Stok Awal	Barang Keluar	Stok Akhir	Keterangan
1	THAI TEA	pak	2	1	1	stok
2	GREEN TEA	pak	1	0	1	stok
3	SKM	kaleng	14	3	11	tidak stok
4	TARO	gram	750	0	750	tidak stok
5	FRESH MILK	mili	17000	10350	6650	stok
6	LEMON TEA	gram	1700	100	1600	tidak stok
7	FN	kaleng	9	1	8	stok
8	BEANS ESPRESSO	gram	250	10	240	stok
9	CREAMER	gram	2400	400	2000	tidak stok
10	GULA PUTIH	gram	2000	500	1500	tidak stok
11	CRÈME BRULEE	mili	650	30	620	stok
12	HAZELNUT	mili	500	40	460	stok

Gambar 3. Transformation

4.4 Data Mining

Proses data mining yang diterapkan dalam penelitian ini yaitu prediksi menggunakan algoritma C4.5. Pada penelitian ini akan dimuat pemodelan pohon keputusan dan akurasi prediksi dari 27 nama stok bahan. Desain pemodelan bisa dilihat dari gambar 3 sebagai berikut:



Gambar 4. Proses di RapidMiner

Selain perhitungan melalui aplikasi RapidMiner, dalam penelitian ini, dilakukan percobaan Algoritma C4.5 dengan melakukan perbandingan dari data 6:4 7:3 8:2. Hasil terbaik akan digunakan dan dipilih,

Data dengan perbandingan 6:4 terdapat Truepositive (TP) 162, True negative (TN) 8, False positive (FP) 1, False negative (FN) 0.

$$(1) \quad accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{162+8}{162+8+1+0} = 96.11\%$$

Data dengan perbandingan 7:3 terdapat True positive (TP) 120, True negative (TN) 1, False positive (FP) 0, False negative (FN) 1.

$$(2) \quad accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{120+1}{120+1+0+1} = 97.60\%$$

Data dengan perbandingan 8:2 terdapat True positive (TP) 82, True negative (TN) 77, False positive (FP) 8, False negative (FN) 0.

$$(3) \quad accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{82+77}{82+77+8+0} = 95.21\%$$

Tabel 2. Hasil perbandingan akurasi prediksi

No	Training	Test	Accuracy	Precision	Recall
1	60	40	96.11%	93.64%	98.78%
2	70	30	97.60%	96.06%	99.19%
3	80	20	95.21%	91.11%	100.00%

The screenshot shows the PerformanceView window in RapidMiner. It displays a confusion matrix for a classification task. The matrix is as follows:

	True Class	Predicted Class	Count
True Class: true stok	Predicted: true stok	Count	122
True Class: true stok	Predicted: false stok	Count	5
True Class: false stok	Predicted: true stok	Count	0
True Class: false stok	Predicted: false stok	Count	122
True Class: true stok	Predicted: true stok	Count	0
True Class: true stok	Predicted: false stok	Count	0
True Class: false stok	Predicted: true stok	Count	0
True Class: false stok	Predicted: false stok	Count	0
True Class: true stok	Predicted: true stok	Count	0
True Class: true stok	Predicted: false stok	Count	0
True Class: false stok	Predicted: true stok	Count	0
True Class: false stok	Predicted: false stok	Count	0

The overall accuracy is 97.60%.

Gambar 5. Hasil akurasi nilai paling tinggi di RapidMiner

4.5 Evaluation

Hasil evaluasi dari 3 rasio perbandingan menghasilkan akurasi data dengan perbandingan 6:4 adalah 96.11%, akurasi data dengan perbandingan 7:3 adalah 97.60% dan akurasi data dengan perbandingan 8:2 adalah 95.21. Hasil yang optimal terdapat pada 7:3 dengan nilai akurasi 97.60% sehingga 97.60% dianggap paling optimal pada tingkat akurasi dataset.

4.6 Knowledge

Pada tahapan yang telah dilakukan bahwa sesuai prediksi data stok bahan minuman di Cafe Semanis menggunakan algoritma C4.5 dengan proses tahapan KDD. Berdasarkan hasil perbandingan data, nilai 97.60% adalah paling tinggi dan akan digunakan.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil pengujian menggunakan algoritma C4.5 terhadap stok bahan minuman pada Cafe Semanis dapat disimpulkan, penerapan metode C4.5 menggunakan data stok bahan minuman dapat digunakan untuk memprediksi stok bahan mana yang harus stok dan tidak stok. Menggunakan Algoritma C4.5 dan dihitung oleh *RapidMiner* dan manual menghasilkan akurasi yang cukup tinggi. Pegujian dilakukan dengan menghitung nilai *Confusion Matrix*, maka nilai akurasi didapat sebesar 97.60%, sehingga Algoritma C4.5 sangat efektif dalam memprediksi stok bahan minuman untuk dimasa mendatang. Hasil penelitian ini, penulis menyadari adanya beberapa

kekurangan dalam penelitian ini. Maka dari itu peneliti menyarankan hal-hal berikut: Perlu adanya penelitian lebih lanjut dengan melakukan pengujian dengan metode lain seperti Naïve Bayes, Regresi Linear dan lain sebagainya agar memperoleh akurasi yang paling tinggi. Pada penelitian selanjutnya bisa menggunakan data yang lebih banyak agar menghasilkan nilai yang lebih akurat.

DAFTAR PUSTAKA

- [1] J. Teknika, J. Fitriana, and J. Triloka, "Teknika 17 (2): 435-445 Prediksi Stok Bahan Baku Minuman Pada Cafe Kiyo Menggunakan Machine Learning," *Ijccs*, vol. x, No.x, no. x, pp. 1-5, 2023.
- [2] M. G. A. Thosca, "Penerapan Algoritma C4. 5 Dalam Memprediksi Pembelian Rokok Elektrik Pada Toko Silvape," *Log. J. Ilmu Komput. dan Pendidik.*, vol. 1, no. 1, pp. 14-19, 2023, [Online]. Available: <http://ejurnal.seminar-id.com/index.php/bulletinds/article/view/810%0Ahttp://ejurnal.seminar-id.com/index.php/bulletinds/article/download/810/594>
- [3] Murlena and D. Apriana, "Penerapan Data Mining Untuk Memprediksi Ketersediaan Stok dengan beberapa fitur dan informasi yang mudah diakses secara real time diantaranya," vol. 2, no. 1, pp. 19-32, 2022.
- [4] C. Nas, "Data Mining Prediksi Minat Calon Mahasiswa Memilih," vol. 11, pp. 131-145, 2021, doi: 10.34010/jamika.v11i2.5506.
- [5] R. Hidayat, A. S. Sunge, and L. I. Achmad, "PENERAPAN DATA MINING DENGAN ALGORITMA," vol. 14, no. 1, pp. 1-11, 2021.
- [6] N. Fitrya, D. Ginting, S. F. Retnawaty, N. Febriani, Y. Fitri, and S. P. Wirman, "Pentingnya Akurasi Dan Presisi Alat Ukur Dalam Rumah Tangga," *J. Pengabd. UntukMu NegeRI*, vol. 1, no. 2, pp. 60-63, 2017, doi: 10.37859/jpumri.v1i2.237.
- [7] M. Taufik Hidayat, N. Suarna, and N. Rahaningsih, "Implementasi Algoritma Naïve Bayes Untuk Prediksi Persediaan Barang Pt. Dilmoni Citra Mebel Indonesia," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 693-699, 2023, doi: 10.36040/jati.v7i1.6310.
- [8] Rian Pratama, B. Huda, E. Novalia, and H. Kabir, "Perbandingan Algoritma C4.5 dan Naïve Bayes dalam Menentukan Persediaan Stok," *Metik J.*, vol. 6, no. 2, pp. 115-122, 2022, doi: 10.47002/metik.v6i2.379.
- [9] A. Wijaya and A. M. Bakti, "Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Memprediksi Penjualan Alat – Alat Telekomunikasi Dan IT (," *Jupiter*, vol. 15, pp. 597-606, 2023.
- [10] M. D. Wahyudi, "Penerapan data mining dengan algoritma c4.5 dalam prediksi penjualan buku," vol. 1, no. 1, pp. 1-6, 2023.