

KLASIFIKASI HASIL PENJUALAN PAKAIAN MENGGUNAKAN ALGORITMA C4.5 PADA TOKO PAKAIAN RIA BUSANA

Fajar Firmansyah¹, Nana Suarna², Willy Prihartono³

Teknik Informatika, STMIK IKMI Cirebon

Rekayasa Perangkat Lunak, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135

Ffajaff1@gmail.com

ABSTRAK

Industri fashion saat ini mengalami peningkatan permintaan yang pesat, dengan konsumen semakin memperhatikan tren pakaian dan memilih berbagai merek untuk memenuhi kebutuhan mereka. Namun, salah satu masalah yang sering dihadapi oleh bisnis pakaian adalah kesulitan dalam mengelola stok, terutama dalam menentukan produk yang laris dan tidak laris di pasar, yang sering menyebabkan overstock atau kekurangan persediaan. Penelitian ini bertujuan untuk menganalisis perilaku pembelian konsumen terhadap produk fashion di RIA Busana menggunakan metode algoritma C4.5. Algoritma ini diterapkan untuk mengidentifikasi faktor-faktor yang memengaruhi keputusan pembelian konsumen berdasarkan data historis yang mencakup transaksi pembelian, preferensi konsumen, dan karakteristik demografi. Metode ini menghasilkan pohon keputusan yang dapat memberikan wawasan mengenai pola pembelian, dengan kategori produk yang terbagi menjadi 'best seller' dan 'non-best seller'. Tujuan penelitian ini adalah untuk membantu mengoptimalkan pengelolaan stok di gudang, menghindari penumpukan stok produk yang tidak laku, dan memastikan kestabilan stok pakaian yang laris. Data dikumpulkan dari transaksi penjualan dan digunakan untuk mengklasifikasikan produk berdasarkan algoritma C4.5, yang diimplementasikan melalui alat RapidMiner. Hasil penelitian menunjukkan bahwa algoritma C4.5 efektif dalam menentukan keakuratan klasifikasi produk terlaris dan tidak laris, sehingga membantu meminimalkan masalah overstock dan kekurangan stok di gudang, yang dapat meningkatkan efisiensi operasional dan daya saing bisnis.

Kata kunci : *Prediksi, Pakaian, Algoritma C4.5, Klasifikasi, Jumlah stok, Jumlah Penjualan.*

1. PENDAHULUAN

Industri pakaian dan gaya hidup berkembang pesat seiring dengan semakin kompleksnya pasar dan perubahan preferensi konsumen yang terus berubah. Konsumen kini semakin selektif dalam memilih produk berdasarkan merek, harga, dan tren terkini. Hal ini menambah tantangan bagi perusahaan pakaian dalam memprediksi penjualan secara akurat guna mengoptimalkan produksi dan mengelola persediaan dengan lebih baik. Prediksi yang tidak tepat dapat menyebabkan overstock atau kekurangan stok, yang berdampak pada efisiensi biaya dan kepuasan pelanggan. Oleh karena itu, memiliki sistem prediksi penjualan yang akurat sangat penting untuk menjaga daya saing dan efisiensi operasional.

Untuk mengatasi permasalahan ini, penerapan teknik analisis data, khususnya algoritma C4.5, menjadi sangat relevan dalam mengekstraksi wawasan dari data penjualan yang kompleks. C4.5 merupakan algoritma pembelajaran yang mampu mengidentifikasi pola dan membangun pohon keputusan untuk memprediksi penjualan berdasarkan variabel-variabel seperti jenis pakaian, harga, musim, dan profil konsumen. Dengan menggunakan algoritma ini, perusahaan dapat menentukan produk yang berpotensi laris dan mengelola stok lebih efektif. Penelitian ini bertujuan untuk menerapkan C4.5 dalam menganalisis perilaku pembelian konsumen di Toko Ria Busana, dengan harapan dapat membantu perusahaan dalam

merancang strategi pemasaran yang lebih tepat dan mengoptimalkan pengelolaan penjualan.

2. TINJAUAN PUSTAKA

Algoritma C4.5 merupakan metode data mining yang efektif untuk membuat pohon keputusan berdasarkan atribut yang memiliki nilai informasi tertinggi, seperti harga, jenis barang, dan waktu pembelian. Dalam industri pakaian, prediksi penjualan menjadi krusial untuk mengelola stok, produksi, dan strategi pemasaran, terutama karena pengaruh berbagai faktor seperti musim dan preferensi konsumen. Penelitian sebelumnya menunjukkan bahwa C4.5 mampu mengungkap pola penjualan yang kompleks dan memberikan hasil prediksi yang akurat. Dengan menerapkan algoritma ini pada data penjualan Toko Ria Busana, penelitian ini bertujuan memberikan wawasan strategis untuk meningkatkan efisiensi dan pengambilan keputusan dalam pengelolaan bisnis pakaian.

2.1. Hasil Literature Review

Metode literature review yang digunakan dalam penelitian ini adalah dengan cara mereview pada artikel-artikel yang terkait dengan penerapan algoritma C4.5 untuk prediksi hasil penjualan pakaian. Langkah-langkah dari literature review meliputi 4 tahapan, yaitu (1) formulasi permasalahan, (2) pencarian literature, (3) evaluasi data, (4) analisis dan interpretasi[1]

Penelitian menurut Husna membahas tentang Untuk mengklasifikasikan himpunan data ke dalam kelas-kelas yang berbeda, proses klasifikasi mengidentifikasi atribut yang identik. Penelitian ini bertujuan untuk mengetahui model Decision Tree menggunakan Algoritma C4.5 dan keakuratan algoritma untuk melakukan klasifikasi terhadap penjualan hijab dari beberapa merek hijab. Model Decision Tree didapat menggunakan Algoritma C4.5 dengan akar pertama adalah atribut Harga, dengan nilai akurasi 87%. [2]

Penelitian menurut Tukino. membahas Perencanaan adalah standar pelaksanaan proyek yang mencakup dokumen, spesifikasi teknis, jadwal, dan anggaran, dan digunakan sebagai acuan untuk melakukan pekerjaan. Hasil proyek konstruksi sangat dipengaruhi oleh perencanaan yang salah, investigasi realisasi proyek yang salah, keahlian manajemen proyek yang tidak memadai, dan kurangnya penyedia jasa profesional. PT SMOE Indonesia bergerak dalam bidang jasa konsultasi konstruksi. [3]

Penelitian menurut Amrullah. membahas beberapa bisnis untuk mendorong karyawannya untuk mencapai target penjualan mereka. Untuk memenuhi kebutuhan berbagai pelanggan, banyak strategi diperlukan. Suamzu Boutique, sebuah usaha dagang (U.D.) swasta yang bergerak dalam penjualan pakaian, telah mengalami masalah dengan penjualan bulan berikutnya karena modalnya terbatas. Metode trend moment, teknik peramalan, digunakan untuk menentukan jumlah penjualan bulan berikutnya di toko Boutique Suamzu. [4]

Penelitian menurut Dewi. membahas Karena UD Andar menjual berbagai jenis barang, penerapan Data Mining sangat penting. Bisnis ini menjual bahan makanan dan minuman, jamu bubuk, kantong plastik, dan makanan beku, yang sangat disukai pelanggan. Dilihat dari banyaknya permintaan konsumen, ternyata ada produk yang paling laris dan produk yang tidak laris. Oleh karena itu, untuk memudahkan pelaku usaha dagang dalam merencanakan penyedia saham, diperlukan prediksi penjualan produk yang paling laris berdasarkan data satu tahun terakhir. [5]

Penelitian menurut Abdillah. membahas yang lebih berfokus pada pengajaran nilai-nilai karakter tanpa mempertimbangkan latar belakang peserta didik, sebagian besar penelitian EDM masih menggunakan faktor kognitif. Oleh karena itu, metode yang dibutuhkan untuk mengidentifikasi siswa yang potensial dan memproyeksikan keberhasilan mereka dalam sistem pendidikan karakter. Dalam penelitian ini, peneliti menggunakan data mahasiswa Unires Yogyakarta. [6]

Penelitian menurut Bachtiar & Mahradianur membahas Bantuan Langsung Tunai (BLT) adalah bantuan yang diberikan oleh pemerintah kepada individu atau keluarga yang kurang mampu dan rentan terhadap kemiskinan dan memenuhi persyaratan program. Penelitian ini bertujuan untuk memberikan masukan tentang apakah keluarga tersebut masih

berhak menerima atau tidak bantuan BLT dalam kasus di mana keluarga tersebut masih memenuhi syarat untuk mendapatkan bantuan pemerintah [7]

Penelitian menurut Wardhana membahas tentang Supermarket adalah pusat perbelanjaan yang menawarkan berbagai kebutuhan sehari-hari kepada pelanggannya. Banyak konsumen pergi ke supermarket untuk membeli apa yang mereka butuhkan. Supermarket semakin berkembang, yang menghasilkan kompetisi pasar yang ketat. Supermarket menawarkan berbagai macam barang dengan berbagai merek, cabang, dan demografi pelanggan. Tujuan penelitian ini adalah untuk menemukan indeks penjualan dari dataset yang mengandung data supermarket yang dilengkapi dengan berbagai informasi terkait supermarket. [8]

Penelitian menurut Aprianti. membahas tentang Contoh paling nyata adalah penurunan perputaran pendapatan UMKM. konsekuensi dari hal ini membutuhkan pendekatan khusus untuk meningkatkan penjualan selama pandemi. Toko Garaya Collection menjual pakaian muslim, tetapi dari berbagai jenis pakaian yang dijual belum tentu semuanya laku atau tidak laku. Ini karena data yang tidak terstruktur dengan baik tentang penjualan, pembelian, barang masuk, dan pengeluaran barang tak terduga tidak digunakan untuk menyusun strategi penjualan. Oleh karena itu, klasifikasi data penjualan stok harus dilakukan menggunakan metode data mining algoritma C4.5 [9]

Penelitian Menurut Irawan B. H. dengan judul penelitian prediksi laba penjualan menggunakan metode algoritma c4.5 pada pt. basunjaya nastari Untuk tetap bertahan dalam persaingan yang ketat saat ini, diperlukan teknik perhitungan laba penjualan yang efektif dan tepat karena banyak data penjualan dipengaruhi oleh era komputerisasi saat ini. masalah PT. Basunjaya Nastari sebagai perusahaan yang bergerak di bidang penyedia jasa dan jual beli bahan kimia harus menentukan keuntungan penjualan untuk pelanggan yang memenuhi persyaratan kualitas dan kuantitas, sehingga perusahaan dapat memprioritaskan pelanggan tersebut untuk memenuhi permintaan yang dibutuhkan. Untuk memecahkan masalah tersebut, Decision Tree, teknik klasifikasi dan prediksi data mining, digunakan. Teknik ini banyak digunakan karena dapat menyajikan data prediksi penjualan lebih cepat. Evaluasi hasil klasifikasi dalam prediksi keuntungan menunjukkan tingkat akurasi 81,71% untuk data pelatihan dan tingkat akurasi 85,83% untuk data pengujian [10]

2.2. Pengertian Data Mining

Data mining adalah proses mengidentifikasi pola, pengetahuan, dan wawasan berharga yang tersembunyi dalam kumpulan data yang besar dan kompleks. Proses ini melibatkan eksplorasi data untuk menemukan informasi yang relevan dengan tujuan mengenali pola-pola yang dapat dimanfaatkan dalam pengambilan keputusan, membuat prediksi, atau

mengungkap pengetahuan baru. Penambahan data menggunakan berbagai metode dan algoritma, seperti clustering, klasifikasi, regresi, asosiasi, dan visualisasi data, untuk menemukan pola yang sulit terlihat secara langsung oleh manusia. Tujuan utama dari data mining adalah mengubah data mentah menjadi informasi yang bernilai dan dapat diterapkan dalam berbagai konteks.

2.3. Pengertian Algoritma C4.5

Algoritma C4.5 adalah metode pembelajaran mesin berbasis pohon keputusan yang dirancang untuk tugas klasifikasi. Algoritma ini dikembangkan oleh Ross Quinlan sebagai peningkatan dari algoritma ID3. Tujuannya adalah membangun pohon keputusan yang dapat digunakan untuk memprediksi atau menentukan kelas data baru berdasarkan pola dalam data pelatihan. C4.5 bekerja dengan memilih atribut terbaik sebagai pemisah untuk setiap simpul pada pohon keputusan menggunakan konsep *entropy* dan *information gain*. Selain itu, algoritma ini dapat menangani atribut kontinu, nilai data yang hilang (*missing values*), serta melakukan *pruning* pada pohon untuk mencegah overfitting.

2.4. Tahapan Kerja Algoritma C4.5

2.4.1. Menghitung Entropy

Entropy digunakan untuk mengukur tingkat ketidakpastian atau keragaman data. Rumusnya adalah

$$Entropy(s) = -i = 1 \sum_{i=1}^n P_i \log_2 (P_i)$$

2.4.2. Menghitung Informasi Grain

Information Gain digunakan untuk menentukan atribut terbaik dengan mengukur pengurangan entropy setelah data dibagi berdasarkan atribut tertentu. Rumusnya:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

2.4.3. Menghitung Ratio

Untuk mengatasi bias pada Information Gain, algoritma ini menghitung Gain Ratio, yang didefinisikan sebagai:

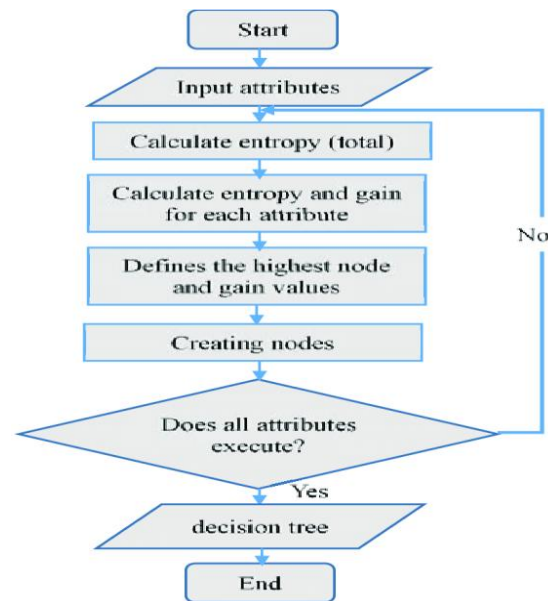
$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)}$$

dengan $SplitInfo(S, A) = - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \log_2 \left(\frac{|S_v|}{|S|} \right)$ dihitung sebagai:

$$SplitInfo(S, A) = - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \log_2 \left(\frac{|S_v|}{|S|} \right)$$

2.5. Flowchart Algoritma C4.5

Langkah-langkah utama dalam algoritma ini dapat dijelaskan sebagai berikut:



Gambar 1. Flowchart algoritma C4.5

- Mulai
Proses dimulai dengan memasukkan data yang akan digunakan untuk membangun pohon keputusan.
- Hitung Entropy Dataset
Hitung nilai entropy untuk mengukur tingkat ketidakpastian dalam dataset.
- Hitung Information Gain untuk Setiap Atribut
Perkirakan seberapa banyak ketidakpastian akan berkurang jika dataset dibagi menggunakan atribut tertentu.
- Pilih Atribut dengan Gain Ratio Tertinggi
Dari semua atribut yang dihitung, pilih yang memiliki Gain Ratio tertinggi sebagai pemisah data.
- Pecah Dataset Berdasarkan Atribut Terpilih
Data dibagi menjadi subset sesuai dengan nilai-nilai atribut yang telah dipilih.
- Periksa Kondisi Berhenti
 - Jika semua data dalam subset memiliki kelas yang sama, maka buat simpul daun (*leaf node*).
 - Jika tidak, ulangi proses untuk subset data tersebut.
- Selesai
Pohon keputusan dianggap selesai jika setiap cabang menghasilkan *leaf node*.

3. METODE PENELITIAN

Dalam menjalankan sebuah penelitian, penting untuk menggunakan metode penelitian yang tepat agar proses penelitian dapat berjalan dengan lancar dan efisien. Metode penelitian berfungsi sebagai panduan atau kerangka kerja yang mengarahkan pelaksanaan penelitian. Dengan mengikuti panduan tersebut, penelitian akan dilaksanakan secara terstruktur dan menghasilkan outcome yang optimal.

3.1. Tahapan Metode Penelitian



Gambar 2. Tahapan Metode Penelitian

Tabel 1 . Deskripsi Metode Penelitian

Tahapan	Aktivitas	Deskripsi Aktivitas
Pengumpulan Data	Mengumpulkan Data	Tahapan Pengumpulan data yang tepat yaitu dengan mempertimbangkan penggunaannya berdasarkan jenis data dan sumbernya. Data yang objektif dan relevan dengan pokok permasalahan penelitian yang merupakan indikator keberhasilan suatu penelitian.
Pengelolaan data	Mengelola Data	Tahapan ini menjelaskan tahap awal dalam data mining. Data terkumpul dan dinyatakan lengkap, data tersebut kemudian di seleksi atas dasar reliabilitas dan validitas. Pengelolaan data juga harus memperhatikan bentuk rumusan masalah dan tujuan penelitian.
Identifikasi Masalah	Mengelompokan Permasalahan	Tahap identifikasi masalah adalah mengelompokan permasalahan yang di peroleh sesuai dengan tujuan penelitian.
Pemodelan	Memilih Metode yang akan di gunakan	Tahapan ini Menjelaskan tentang pemilihan dan pemodelan metode yang digunakan dengan metode algoritma C4.5 pada peneliltian. Penjelasan ini juga mengenai pengaturan untuk mendukung proses dan pemilihan nilai dari atribut – atribut serta tahapan proses pemodelan dengan algoritma C4.5 yang digunakan.
Pengujian model	Pengujian Metode yang sudah di pilih	Tahapan ini Menjelaskan mengenai evaluasi dan validitas hasil penerapan metode algoritma C4.5 pada penelitian yang dilakukan.
Evaluasi Dan Hasil	Evaluasi Hasil	Tahapan terakhir ini menjelaskan mengenai evaluasi dan validasi hasil penerpan metode algoritma C4.5 pada penelitian yang dilakukan.

3.2. Teknik Pengumpulan Data

Pada tahap pengumpulan data, dilakukan melalui beberapa metode, yaitu:

- Observasi
- Survei

Dalam penelitian ini, pengumpulan data dilakukan dengan turun langsung ke lapangan untuk memperoleh informasi yang akurat dan relevan terkait dengan masalah yang sedang diteliti. Setelah data

terkumpul, langkah berikutnya adalah melakukan analisis data.

3.3. Sumber Data

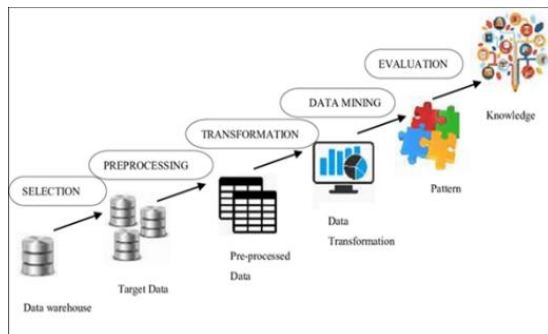
Data yang di pakai dari penelitian ini adalah data primer, yang dikumpulkan secara langsung oleh pengumpul data dalam penelitian ini, diperoleh dari **Toko Ria Busana** melalui wawancara dan observasi.

Tabel 2. Sumber Data

No	Tanggal	Jenis pakaian	Stok	Harga	Terjual	sis	Total	Status Penjualan
1	01-Feb-23	celana	100	Rp130.000	35	65	Rp4.550.000	TERJUAL
2	01-Feb-23	baju	100	Rp75.000	32	68	Rp2.400.000	TERJUAL
3	01-Feb-23	sandal	100	Rp45.000	29	71	Rp1.305.000	TERJUAL
4	01-Feb-23	sepatu	100	Rp95.000	10	90	Rp950.000	KURANG TERJUAL
5	02-Feb-23	celana	100	Rp130.000	39	61	Rp5.070.000	TERJUAL
6	02-Feb-23	baju	100	Rp75.000	75	25	Rp5.625.000	TERJUAL
7	02-Feb-23	sandal	100	Rp45.000	35	65	Rp1.575.000	TERJUAL
8	02-Feb-23	sepatu	100	Rp95.000	25	75	Rp2.375.000	KURANG TERJUAL
9	03-Feb-23	celana	100	Rp130.000	46	54	Rp5.980.000	TERJUAL
10	03-Feb-23	baju	100	Rp75.000	69	31	Rp5.175.000	KURANG TERJUAL
....					...	..		
111	28-Mar-23	sandal	100	Rp45.000	32	68	Rp1.440.000	TERJUAL
112	28-Mar-23	sepatu	100	Rp95.000	25	75	Rp2.375.000	KURANG TERJUAL

3.4. Teknik Analisis Data

Untuk menganalisis data dalam penerapan data mining, penelitian ini menggunakan tahapan dalam proses Knowledge Discovery in Databases (KDD), untuk mengetahui tahapan Tahapan KDD



Gambar 3. Tahapan Metode KDD

Knowledge Discovery in Databases (KDD) adalah proses untuk mengekstraksi pengetahuan dari data. Berikut adalah tahapan utama dalam KDD:

3.5. Seleksi Data

Untuk membaca dataset dalam bentuk file excel, menggunakan operator Read Excel. Data yang digunakan pada penelitian ini ialah sebanyak 250 record dataset.

3.6. Pemilihan Data

Proses cleansing atau pembersihan data yang missing atau memiliki nilai yang tidak konsisten pada langkah preprocessing. Sebelum melakukan proses ini, dilakukan analisa terlebih dahulu apakah dataset memiliki nilai missing atau tidak serta memiliki data yang konsisten atau tidak.

3.7. Transformasi

Proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Pada langkah transformation dilakukan untuk data bertipe polynominal menjadi numerik agar data dapat diolah berdasarkan jarak. Untuk mengubah data bertipe polynominal menjadi numerik menggunakan operator Nominal to Numerical.

3.8. Data Mining

Data mining merupakan proses penemuan pola atau informasi yang menarik dari data yang telah dipilih, menggunakan berbagai teknik atau metode. Terdapat beragam teknik, metode, atau algoritma dalam data mining, dan pemilihan yang tepat sangat tergantung pada tujuan serta proses *Knowledge Discovery in Databases* (KDD) secara keseluruhan.

3.9. Implementasi/Evaluasi

Penerjemah alur yang dihasilkan dari data mining, alur informasi yang dihasilkan dari proses perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Proses ini merupakan proses yang mencakup pemeriksaan apakah alur atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada pada sebelumnya

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

Hasil penelitian yang disajikan dalam pembahasan ini menjelaskan bagaimana menganalisis dataset penjualan pakaian dengan menggunakan Algoritma C4.5. Penelitian ini dilakukan melalui tahapan penerapan machine learning yang menggunakan perangkat lunak RapidMiner Studio.

4.2. Lokasi Penelitian

Penelitian ini dilakukan di Ria Busana yang terletak di Jl. Raya Laswi No.272, Majalaya, Kec. Majalaya, Kabupaten Bandung, Jawa Barat 40382.

4.3. Data Selection

Data yang digunakan dalam penelitian ini adalah data penjualan Pakaian pada Bulan September - November 2023. Kemudian data tersebut diseleksi dan akan digunakan untuk diolah menggunakan aplikasi/tools Rapidminer. Tahapan seleksi dilakukan di Microsoft Excel dengan atribut yang digunakan diantaranya No, Tanggal, Jenis Pakaian, Qty, Harga, Jumlah Terjual, sisa, Total, dan Status Penjualan. Hasil seleksi data menghasilkan 364 data

Tabel 3. Data Penjualan

No	Tanggal	Jenis pakaian	Stok	Harga	Terjual	sisa	Total	Status Penjualan
1	01-Feb-23	celana	100	Rp130.000	35	65	Rp4.550.000	TERJUAL
2	01-Feb-23	baju	100	Rp75.000	32	68	Rp2.400.000	TERJUAL
3	01-Feb-23	sandal	100	Rp45.000	29	71	Rp1.305.000	TERJUAL
4	01-Feb-23	sepatu	100	Rp95.000	10	90	Rp950.000	KURANG TERJUAL
5	02-Feb-23	celana	100	Rp130.000	39	61	Rp5.070.000	TERJUAL
6	02-Feb-23	baju	100	Rp75.000	75	25	Rp5.625.000	TERJUAL
7	02-Feb-23	sandal	100	Rp45.000	35	65	Rp1.575.000	TERJUAL
8	02-Feb-23	sepatu	100	Rp95.000	25	75	Rp2.375.000	KURANG TERJUAL
9	03-Feb-23	celana	100	Rp130.000	46	54	Rp5.980.000	TERJUAL
10	03-Feb-23	baju	100	Rp75.000	69	31	Rp5.175.000	KURANG TERJUAL
....					...	..		
111	28-Mar-23	sandal	100	Rp45.000	32	68	Rp1.440.000	TERJUAL
112	28-Mar-23	sepatu	100	Rp95.000	25	75	Rp2.375.000	KURANG TERJUAL

4.4. Data preprocessing

Pada tahap ini dilakukan preprocessing data yaitu proses cleaning atau pembersihan data. Proses penghapusan data atribut yang tidak diperlukan seperti Harga, Terjual, Sisa, Total dan Status Penjualan. Sedangkan atribut yang akan digunakan adalah No, Tanggal, Jenis Pakaian dan Stock

Tabel 4. Data Preprocessing

No	Tanggal	Jenis pakaian	Stok
1	01-Feb-23	celana	100
2	01-Feb-23	baju	100
3	01-Feb-23	sandal	100
4	01-Feb-23	sepatu	100
5	02-Feb-23	celana	100
6	02-Feb-23	baju	100
7	02-Feb-23	sandal	100
8	02-Feb-23	sepatu	100
...	...	...	...
109	28-Mar-23	celana	100
110	28-Mar-23	baju	100
111	28-Mar-23	sandal	100
112	28-Mar-23	sepatu	100

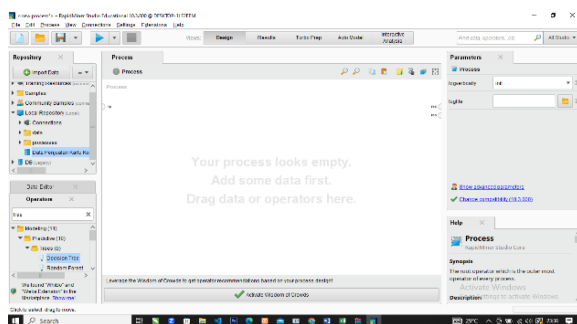
4.5. Data Mining

Teknik analisis data yang digunakan dalam penelitian ini adalah Perhitungan Data Menggunakan Algoritma C4.5 dengan Rapidminer. Analisis yang dilakukan yaitu melakukan proses analisis dengan menggunakan tahapan Algoritma C4.5 sampai dengan membentuk akhir dari pohon keputusan dengan software rapidminer.

Berikut pengaplikasian metode C4.5 dalam aplikasi/tools Rapidminer Diantaranya sebagai berikut:

4.6. Tampilan Halaman Utama Rapidminer

Pada gambar dibawah adalah tampilan halaman utama Rapidminer ketika dibuka



Gambar 4. Halaman Utama

RapidMiner adalah perangkat lunak yang digunakan untuk pengolahan data mining, khususnya dalam analisis teks dan pola dari kumpulan data besar. Software ini menggabungkan teknik statistik, kecerdasan buatan, dan basis data untuk memperoleh informasi berkualitas tinggi dari teks yang diolah. RapidMiner mendukung berbagai teknik penambangan data dan pembelajaran mesin, termasuk proses ETL (Ekstrak, Transformasi, Muat),

prapemrosesan data, visualisasi, pemodelan, dan evaluasi. Proses penambangan data di RapidMiner ditulis dalam XML dan terdiri dari operator bersarang yang dapat dibuat melalui antarmuka pengguna grafis (GUI), sementara presentasi ditulis dalam bahasa pemrograman Java.

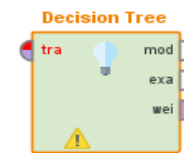
4.7. Proses input data dengan Format excel



Gambar 5. Retrive Data

Pada gambar diatas menjelaskan proses penginputan data. Data yang diinput bisa berbagai macam format seperti xls, Excel dan masih banyak lainnya. Akan tetapi file yang akan diuji di aplikasi Rapidminer ini menggunakan data dengan format Excel yang sudah disiapkan

4.8. Decision Tree (Pohon Keputusan)



Gambar 6. Decision Tree

Pada gambar di atas Operator Decision Tree menghasilkan model pohon keputusan, yang dapat digunakan untuk klasifikasi dan regresi.

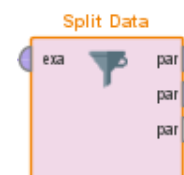
4.9. Set Role



Gambar 7. Set role

Set Role menjelaskan bahwa setiap atribut dalam dataset memiliki **peran** yang menentukan bagaimana atribut tersebut akan digunakan atau ditangani oleh algoritme atau operator dalam proses analisis data.

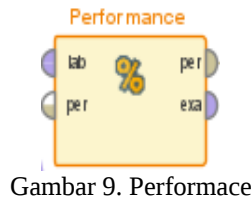
4.10. Split Data



Gambar 8. Split Data

Operator ini menghasilkan jumlah subset yang diinginkan dari ExampleSet yang diberikan. ExampleSet dipartisi menjadi subset sesuai dengan ukuran relatif yang ditentukan. Operator Split Data mengambil ExampleSet sebagai inputnya dan mengirimkan subset dari ExampleSet tersebut melalui port outputnya.

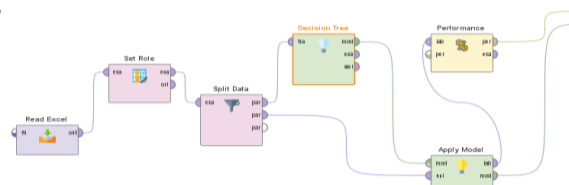
4.11. Performace



Gambar 9. Performace

Berbeda dengan operator evaluasi kinerja lainnya seperti operator Kinerja (Klasifikasi), operator Kinerja (Klasifikasi Binominal) atau operator Kinerja (Regresi), operator ini dapat digunakan untuk semua jenis tugas pembelajaran. Operator ini secara otomatis menentukan jenis tugas pembelajaran dan menghitung kriteria yang paling umum untuk jenis tersebut.

4.12. Apply Model



Gambar 10. Apply Model

Apply Model adalah tahap akhir di mana model data mining C4.5 diterapkan berdasarkan langkah-langkah sebelumnya, seperti membaca data dari Excel, menetapkan peran atribut, dan membangun pohon keputusan. Setelah seluruh proses dilakukan, hasil akhirnya adalah pohon keputusan yang dihasilkan dari data yang telah dianalisis.

Pada gambar 4 di bagian atas di operator Retrieve data akan yang akan dimasukkan seperti gambar 10

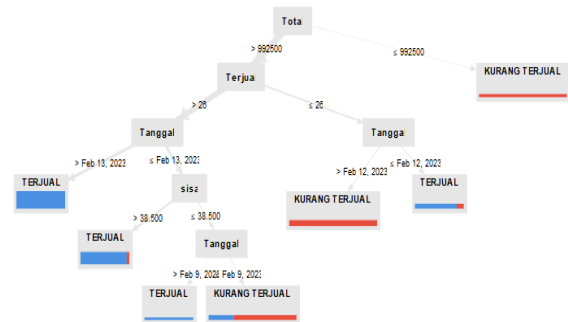
No	Status Peng...	Tanggal	Jenis pakaian	Stok	Harga	Terjual	sisa	Total
1	TERJUAL	Feb 1, 2023	celana	100	130000	35	65	4550000
2	TERJUAL	Feb 1, 2023	baju	100	75000	32	68	2400000
4	KURANG TE...	Feb 1, 2023	sepatu	100	95000	10	90	950000
6	TERJUAL	Feb 2, 2023	baju	100	75000	75	25	5625000
7	TERJUAL	Feb 2, 2023	sandal	100	45000	35	65	1575000
8	KURANG TE...	Feb 2, 2023	sepatu	100	95000	25	75	2375000
9	TERJUAL	Feb 3, 2023	celana	100	130000	46	54	5980000
10	KURANG TE...	Feb 3, 2023	baju	100	75000	69	31	5175000
12	TERJUAL	Feb 3, 2023	sepatu	100	95000	21	79	1995000
13	TERJUAL	Feb 4, 2023	celana	100	130000	27	73	3510000
14	KURANG TE...	Feb 4, 2023	baju	100	75000	83	17	6225000
16	TERJUAL	Feb 4, 2023	sepatu	100	95000	13	87	1235000
19	TERJUAL	Feb 5, 2023	sandal	100	45000	24	76	1080000
20	TERJUAL	Feb 5, 2023	sepatu	100	95000	29	71	2755000
21	TERJUAL	Feb 6, 2023	celana	100	130000	44	56	5720000

Gambar 11. Data yang dimasukkan retrieve data

4.13. Pembahasan

Di bawah ini adalah penjelasan detail setiap langkah dalam membangun pohon keputusan dengan menggunakan algoritma C4.5

a. Pengaplikasian Pohon Keputusan metode C4.5 dalam aplikasi/tools Rapidminer Diantaranya sebagai berikut.



Gambar 12. Pengaplikasian Pohon Keputusan

b. Data Description Pohon keputusan metode C4.5 dalam aplikasi/tools Rapidminer Diantaranya sebagai berikut :

4.14. Tree

Total > 992500
 | Terjual > 26
 | | Tanggla > Feb 13, 2023: TERJUAL {TERJUAL=29, KURANG TERJUAL=0}
 | | Tanggla ≤ Feb 13, 2023
 | | | sisa > 38.500: TERJUAL {TERJUAL=19, KURANG TERJUAL=1}
 | | | sisa ≤ 38.500
 | | | | Tanggla > Feb 9, 2023: TERJUAL {TERJUAL=3, KURANG TERJUAL=0}
 | | | | Tanggla ≤ Feb 9, 2023: KURANG TERJUAL {TERJUAL=2, KURANG TERJUAL=5}
 | Terjual ≤ 26
 | | Tanggla > Feb 12, 2023: KURANG TERJUAL {TERJUAL=0, KURANG TERJUAL=9}
 | | Tanggla ≤ Feb 12, 2023: TERJUAL {TERJUAL=6, KURANG TERJUAL=1}
 Total ≤ 992500: KURANG TERJUAL {TERJUAL=0, KURANG TERJUAL=4}

c. Sebuah eksperimen, dengan metrik *accuracy*, *precision*, *recall*, dan *F1 score* sebagai indikator kinerja model sebagai berikut :

accuracy: 87.88%

	true TERJUAL	true KURANG TERJUAL	class precision
pred. TERJUAL	24	3	88.89%
pred. KURANG TERJUAL	1	5	83.33%
class recall	96.00%	62.50%	

Gambar 13. Indikator kerja model

Penjelasan dibawah ini adalah data dengan ratio 7:3 terdapat True positive (TP) 24, True negative (TN) 5, False positive (FP) 3, False negative (FN) 1, Maka dapat di hitung nilai *accuracy*, *presision*, *recall*, dan *f1 score*

- a. Accuracy Menggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar .

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

$$accuracy = \frac{24 + 5}{24 + 5 + 3 + 1}$$

$$accuracy = 87.88 \%$$

- b. Recall menggambarkan keberhasilan model dalam menentukan kembali sebuah informasi sebagai berikut :

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{24}{24 + 1}$$

$$Recall = 96.00 \%$$

- c. Presicion Menggambarkan akurasi anatara data yang di minta dengan hasil prediksi yang di berikan oleh model sebagai berikut :

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{24}{24 + 3}$$

$$Precision = 88.89 \%$$

- d. F1 score menggambarkan perbandingan rata – rata precision dan recall yang di bobotkan. Accuracy tepat kita gunakan sebagai acuan performasi algoritma jika dataset kita memiliki jumlah data false negatif dan false posiif yang sangat mendekati (symmetric). Namun jika jumlahnya tidak mendekati, maka sebaliknya kita menggunakan F1 Score sebagai acuan

$$F1\ Score = \frac{2 \times recall \times precision}{recall + precision}$$

$$F1\ Score = \frac{2 \times 0.96 \times 0.8889}{0.96 + 0.8889}$$

$$F1\ Score = 92.30 \%$$

- e. Maka dapat hasisl nilai *accuracy*, *presision*, *recall*, dan *f1 score* adalah.

- Akurasi Total adalah 87.88 %
- Presisi untuk kelas TL adalah 88.89 % dan Presisi untuk kelas L adalah 83.33 %.
- Presisi untuk kelas TL adalah 96.00 % dan Presisi untuk kelas L adalah 62.50 %.
- F1 score rata rata adalah sekitar 92,30 %

5. KESIMPULAN DAN SARAN

Berdasarkan hasil pengujian menggunakan algoritma C4.5 terhadap status penjualan pakaian di TOKO RIA BUSANA, dapat disimpulkan bahwa penerapan metode algoritma C4.5 dengan menggunakan data status penjualan sebelumnya efektif untuk mengklasifikasikan status penjualan, yaitu memisahkan produk yang terjual dan kurang terjual. Hasil pengujian menunjukkan bahwa algoritma C4.5 di Toko Ria memiliki nilai akurasi 87,88%, presisi 88,89%, recall 96,00%, dan F1 score 92,30%, yang menunjukkan bahwa algoritma ini sangat efektif dalam menghasilkan hasil yang akurat dan memberikan wawasan yang berguna untuk pengelolaan bisnis yang lebih baik. Berdasarkan hasil ini, algoritma C4.5 dapat digunakan dengan baik untuk mengklasifikasikan penjualan masa depan. Sebagai saran, penelitian selanjutnya disarankan untuk mencoba metode lain seperti Naïve Bayes, Regresi Linear, atau teknik lainnya guna meningkatkan tingkat akurasi. Selain itu, penggunaan data yang lebih besar dan lebih spesifik juga dianjurkan untuk menghasilkan prediksi yang lebih akurat.

DAFTAR PUSTAKA

- [1] Y. Indah, "Prediksi Tingkat Kepuasan Pelayanan Online Menggunakan Metode Algoritma C4.5," *J. Inform. Ekon. Bisnis*, 2022, doi: 10.37034/inf.v4i2.99.
- [2] F. Husna, H. Rahman, and J. Juhari, "Implementasi Data Mining Menggunakan Algoritma C4.5 pada Klasifikasi Penjualan Hijab," *J. Ris. Mhs. Mat.*, vol. 2, no. 2, 2022, doi: 10.18860/jrmm.v2i2.14891.
- [3] T. Tukino, "Penerapan Algoritma C4.5 Untuk Memprediksi Keuntungan Pada PT SMOE Indonesia," *J. Sist. Inf. BISNIS*, vol. 9, no. 1, 2019, doi: 10.21456/vol9iss1pp39-46.
- [4] A. Amrullah, E. Affandi, W. Riansyah, and S. Sobirin, "Peramalan Penjualan Bulanan menggunakan metode Trend Moment pada Toko Suamzu Boutique," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 19, no. 2, 2020, doi: 10.53513/jis.v19i2.2423.
- [5] S. P. Dewi, N. Nurwati, and E. Rahayu, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor," *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, 2022, doi: 10.47065/bits.v3i4.1408.
- [6] M. A. Abdillah, A. Setyanto, and S. Sudarmawan, "Implementasi Decision Tree Algoritma C4.5 Untuk Memprediksi Kesuksesan Pendidikan Karakter," *Respati*, vol. 15, no. 2, 2020, doi: 10.35842/jtir.v15i2.349.
- [7] L. Bachtiar and M. Mahradianur, "Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai," *J. Inform.*, vol. 10, no. 1, 2023, doi: 10.31294/inf.v10i1.15115.

- [8] A. W. Wardhana, E. Patimah, A. I. Shafarindu, Y. M. Siahaan, B. V. Haekal, and D. S. Prasvita, "Klasifikasi Data Penjualan pada Supermarket dengan Metode Decision Tree," *Senamika*, vol. 2, no. 1, 2021.
- [9] C. Aprianti, M. Faishal, and Y. Umaidah, "Penerapan Data Mining Untuk Klasifikasi Penjualan Baju Muslim Dimasa Pandemi Covid-19 Menggunakan Metode Algoritma C4.5 (Studi Kasus: Garaya Collection)," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 1, 2021.
- [10] B. H. Irawan, "PREDIKSI LABA PENJUALAN MENGGUNAKAN METODE ALGORITMA C4.5 PADA PT. BASUNJAYA NASTARI," *Educic - Sci. J. Informatics Educ.*, vol. 6, no. 2, 2020, doi: 10.21107/educic.v6i2.6344.