Jurnal Informatika Ekonomi Bisnis

http://www.infeb.org

2025 Vol. 7 No. 3 Hal: 646-652 e-ISSN: 2714-8491

Prediksi Penjualan Brand di HGVR Store Menggunakan Algoritma C4.5 dan Naïve Bayes

Candra Naya[1]✉, Elkin Rilvani[2]

[1,2]Universitas Pelita Bangsa

candranaya@pelitabangsa.ac.id

Abstract

HGVR Brand is a creative industry engaged in the production and distribution of ready-to-wear clothing established in 2015, which has a reseller network in several major cities in Java. This study aims to analyze the prediction of HGVR Store product sales levels using data mining methods, specifically the C4.5 and Naïve Bayes algorithms, so that it can assist the company in determining marketing strategies and inventory management. The data used in this study consists of 500 sales data collected in June 2019 through observation, interviews, and internal company documentation. The input variables used include the number of orders (PO), quantity, price, and sales status, while the target variable is the classification of sales into "high" and "low" categories. The analysis process is carried out through the stages of data cleaning, transformation, and validation using the split validation technique (70% training data and 30% testing data). The C4.5 algorithm is used to build a decision tree model, while the Naïve Bayes algorithm is used to calculate the classification probability. The test results show that the C4.5 algorithm has a 100% accuracy rate with an excellent classification category based on the ROC curve (AUC = 1.00). Meanwhile, the Naïve Bayes algorithm also produced good classification results, although its accuracy was lower than that of C4.5. The conclusion of this study is that the C4.5 algorithm is more optimal than Naïve Bayes in predicting sales levels at the HGVR Store. These findings are expected to inform decision-making for the HGVR Brand in formulating business strategies.

Keywords: HGVR Brand, Data Mining, Sales Prediction, C4.5 Algorithm, Naïve Bayes

Abstrak

HGVR Brand merupakan industri kreatif yang bergerak di bidang produksi dan distribusi pakaian jadi yang berdiri sejak tahun 2015, yang memiliki jaringan reseller di beberapa kota besar di Pulau Jawa. Penelitian ini bertujuan untuk menganalisis prediksi tingkat penjualan produk HGVR Store dengan menggunakan metode data mining, khususnya algoritma C4.5 dan Naïve Bayes, sehingga dapat membantu perusahaan dalam menentukan strategi pemasaran dan pengelolaan persediaan barang. Data yang digunakan dalam penelitian ini terdiri dari 500 data penjualan yang dikumpulkan pada bulan Juni 2019 melalui observasi, wawancara, serta dokumentasi internal perusahaan. Variabel input yang digunakan meliputi jumlah pemesanan (PO), kuantitas, harga, dan status penjualan, sedangkan variabel target adalah klasifikasi penjualan dengan kategori “tinggi” dan “rendah”. Proses analisis dilakukan dengan tahapan pembersihan data, transformasi, serta validasi menggunakan teknik split validation (70% data training dan 30% data testing). Algoritma C4.5 digunakan untuk membangun model decision tree, sedangkan algoritma Naïve Bayes digunakan untuk menghitung probabilitas klasifikasi. Hasil pengujian menunjukkan bahwa algoritma C4.5 memiliki tingkat akurasi 100% dengan kategori excellent classification berdasarkan kurva ROC (AUC = 1,00). Sementara itu, algoritma Naïve Bayes juga memberikan hasil klasifikasi yang baik, meskipun tingkat akurasinya lebih rendah dibandingkan C4.5. Kesimpulan dari penelitian ini adalah bahwa algoritma C4.5 lebih optimal dibandingkan Naïve Bayes dalam memprediksi tingkat penjualan di HGVR Store. Temuan ini diharapkan dapat menjadi dasar pengambilan keputusan bagi HGVR Brand dalam merumuskan strategi bisnis.

Kata kunci: Kata Kunci: HGVR Brand, Data Mining, Prediksi Penjualan, Algoritma C4.5, Naïve Bayes

1. Pendahuluan

Perkembangan teknologi digital telah membawa perubahan signifikan dalam sektor perdagangan ritel, khususnya di Indonesia. Pertumbuhan retail digital yang pesat memunculkan pola konsumsi baru, di mana konsumen semakin mengandalkan toko online, media sosial, dan jaringan reseller dalam melakukan pembelian produk [1]. Perubahan ini menuntut pelaku usaha, termasuk industri kreatif seperti HGVR Brand, untuk mampu beradaptasi dan mengembangkan strategi bisnis berbasis data agar tetap kompetitif. HGVR Brand yang berdiri sejak tahun 2015, dengan fokus pada produksi dan distribusi pakaian jadi, kini tidak hanya mengandalkan toko fisik (HGVR Store), tetapi juga memperluas jaringan melalui reseller yang tersebar di berbagai kota besar di Pulau Jawa. Dalam era digital ini, data transaksi menjadi salah satu aset terpenting bagi perusahaan [2].

Data yang tercatat dalam proses penjualan mulai dari jumlah pesanan, kuantitas, harga, hingga status penjualan dapat diolah lebih lanjut untuk menghasilkan informasi yang bernilai strategis. Dengan analisis yang tepat, data tersebut dapat membantu perusahaan dalam memprediksi pola penjualan, mengelola persediaan secara lebih efisien, menetapkan harga yang kompetitif, serta menyusun strategi pemasaran yang

Diterima: 31-08-2025x | Revisi: 10-09-2025 | Diterbitkan: 13-09-2025 | doi: 10.37034/infeb.v7i3.1242

lebih tepat sasaran [3]. Namun, pemanfaatan data transaksi dalam pengambilan keputusan tidaklah sederhana, sebab diperlukan metode analisis yang mampu mengolah data dalam jumlah besar secara akurat dan efisien [4]. Salah satu pendekatan yang dapat digunakan adalah data mining, yaitu proses penggalian pola atau pengetahuan tersembunyi dari kumpulan data.

Dalam penelitian ini, metode algoritma C4.5 [5] dan Naïve Bayes [6] dipilih sebagai model prediksi tingkat penjualan di HGVR Store. Algoritma C4.5 mampu membangun model klasifikasi berbentuk decision tree [7] yang mudah dipahami, sedangkan Naïve Bayes memiliki keunggulan dalam menghitung probabilitas klasifikasi secara cepat meskipun dengan asumsi independensi antarvariabel [8]. Kedua metode ini dianggap relevan untuk membandingkan kinerja prediksi penjualan berdasarkan data transaksi yang tersedia. Namun, terdapat beberapa tantangan yang dihadapi dalam penelitian ini. Pertama, kualitas data transaksi sering kali dipengaruhi oleh ketidaklengkapan, duplikasi, atau ketidakkonsistenan pencatatan, sehingga diperlukan proses data preprocessing yang baik. Kedua, dinamika perilaku konsumen dalam industri fashion cenderung fluktuatif mengikuti tren, sehingga model prediksi harus adaptif terhadap perubahan [9].

Ketiga, pemilihan algoritma yang tepat menjadi krusial, mengingat setiap metode memiliki kelebihan dan keterbatasan. Oleh karena itu, penelitian ini tidak hanya berfokus pada membangun model prediksi penjualan, tetapi juga mengevaluasi kinerja algoritma C4.5 dan Naïve Bayes dalam menghasilkan klasifikasi penjualan yang akurat [10]. Dengan latar belakang tersebut, penelitian ini diharapkan dapat memberikan kontribusi nyata bagi HGVR Brand dalam pengambilan keputusan bisnis berbasis data. Selain itu, hasil penelitian ini juga dapat memperkaya kajian akademis mengenai penerapan data mining di bidang manajemen ritel, khususnya dalam konteks prediksi penjualan produk fashion di era digital [11]. Dalam konteks bisnis modern, data transaksi sering disebut sebagai "aset baru" yang setara dengan sumber daya ekonomi lainnya. Namun, data mentah yang hanya berupa catatan penjualan tanpa diolah sebenarnya tidak memiliki nilai strategis. Data tersebut hanya menjadi deretan angka dan informasi administratif yang sulit memberikan gambaran mendalam mengenai pola konsumsi atau potensi pasar [12].

Nilai nyata dari data baru muncul ketika dilakukan analisis menggunakan pendekatan ilmiah dan metode komputasi yang tepat [13] Bagi HGVR Store, data transaksi yang berisi informasi seperti jumlah pesanan, harga, kuantitas, dan status penjualan hanya akan menjadi tumpukan arsip apabila tidak diolah. Tanpa analisis, perusahaan tidak dapat mengetahui produk mana yang berpotensi tinggi terjual, kapan periode puncak penjualan terjadi, maupun strategi harga yang paling efektif [14]. Oleh karena itu, analisis data mining dengan algoritma C4.5 dan Naïve Bayes menjadi penting, karena mampu mengubah data mentah tersebut menjadi pengetahuan yang dapat mendukung pengambilan keputusan manajerial [14]. Dengan kata lain, penelitian ini menegaskan bahwa nilai sesungguhnya dari data tidak terletak pada jumlahnya, melainkan pada sejauh mana data tersebut dapat diolah menjadi informasi prediktif yang bermanfaat. Melalui perbandingan performa algoritma C4.5 dan Naïve Bayes, diharapkan dapat ditemukan model prediksi penjualan yang paling akurat untuk HGVR Store, sehingga data yang semula hanya bersifat pasif dapat diubah menjadi dasar penyusunan strategi bisnis yang aktif dan tepat sasaran.

Dalam dunia bisnis ritel modern [15], keberadaan data dalam jumlah besar tidak lagi menjadi masalah, tetapi justru menjadi tantangan baru. Data transaksi yang tercatat setiap hari pada HGVR Store menyimpan potensi informasi penting terkait pola pembelian konsumen, kecenderungan tren fashion, serta fluktuasi penjualan antarperiode. Namun, potensi tersebut sulit dimanfaatkan apabila hanya dipandang sebagai data mentah yang bersifat statis. Untuk itu, dibutuhkan sebuah pendekatan analitis yang mampu menggali informasi tersembunyi dan mengubahnya menjadi dasar pengambilan keputusan yang tepat. Salah satu solusi yang dapat digunakan adalah data mining, yaitu proses penggalian pola dan hubungan yang tidak terlihat secara langsung dari data dalam jumlah besar.

Dengan menggunakan metode ini, perusahaan dapat memperoleh wawasan yang lebih dalam mengenai karakteristik konsumen maupun prediksi kinerja penjualan di masa mendatang. Data mining tidak hanya berfungsi sebagai alat analisis, tetapi juga sebagai instrumen strategis yang mampu meningkatkan daya saing perusahaan dalam menghadapi dinamika pasar. Dalam konteks penelitian ini, masalah spesifik yang dihadapi HGVR Store adalah kesulitan dalam memprediksi tingkat penjualan secara akurat. Tanpa prediksi yang tepat, perusahaan berisiko mengalami overstock atau stock-out yang dapat menimbulkan kerugian finansial maupun menurunnya kepuasan konsumen. Selain itu, sifat industri fashion [16] yang sangat dipengaruhi oleh tren membuat pola penjualan cenderung fluktuatif dan sulit dipetakan hanya dengan analisis konvensional. Oleh karena itu, algoritma C4.5 dan Naïve Bayes dipilih sebagai pendekatan komparatif, karena keduanya memiliki kemampuan dalam melakukan klasifikasi serta prediksi berbasis data historis.

Dengan membandingkan kedua metode ini, penelitian diharapkan dapat menemukan model yang paling efektif dalam mendukung strategi penjualan HGVR Store [17]. Salah satu teknik penting dalam data mining yang relevan untuk analisis penjualan ritel adalah association rule mining [18]. Teknik ini digunakan untuk menemukan hubungan atau keterkaitan antarproduk dalam data transaksi, sehingga dapat mengungkap pola pembelian konsumen yang tidak terlihat secara langsung. Misalnya, apabila suatu produk sering dibeli bersamaan dengan produk lain,

maka perusahaan dapat memanfaatkan informasi tersebut untuk strategi promosi bundling, penataan produk di toko, hingga rekomendasi penjualan yang lebih tepat sasaran. Dalam konteks HGVR Store, masalah spesifik yang dihadapi adalah bagaimana menemukan pola pembelian konsumen yang dapat meningkatkan efisiensi operasional.

Tanpa adanya pemetaan pola, perusahaan berisiko tidak mampu mengoptimalkan persediaan produk, sehingga terjadi kelebihan stok pada produk yang jarang laku atau kekurangan stok pada produk yang sebenarnya memiliki tingkat permintaan tinggi. Dengan menerapkan association rule mining, HGVR Store dapat mengidentifikasi kombinasi produk yang paling sering dibeli bersamaan serta tren yang muncul dari data transaksi historis. Dengan demikian, tantangan utama penelitian ini bukan hanya sekadar mengklasifikasikan tingkat penjualan menggunakan algoritma C4.5 dan Naïve Bayes, tetapi juga memahami bagaimana pola pembelian terbentuk melalui teknik association rule mining. Hasil analisis ini diharapkan mampu memberikan efisiensi dalam pengelolaan stok, strategi promosi, serta perencanaan penjualan [19], sehingga HGVR Store dapat lebih adaptif dalam menghadapi dinamika pasar fashion yang cepat berubah.

Dalam ranah association rule mining, salah satu algoritma yang paling banyak digunakan adalah Apriori. Algoritma ini bekerja dengan prinsip sederhana namun kuat, yaitu mencari frequent itemset atau kumpulan item yang sering muncul secara bersamaan dalam data transaksi. Prosesnya dilakukan secara iteratif dengan memanfaatkan ukuran support (tingkat kemunculan itemset dalam seluruh transaksi) dan confidence (tingkat kepercayaan bahwa suatu item muncul jika item lainnya sudah muncul). Dengan pendekatan ini, Apriori mampu menghasilkan aturan asosiasi yang menjelaskan pola pembelian konsumen, seperti jika konsumen membeli produk A, maka kemungkinan besar ia juga membeli produk B. Rasionalisasi pemilihan Apriori dalam penelitian ini didasarkan pada kemampuannya yang telah teruji dalam dunia ritel untuk menemukan pola keterkaitan antarproduk.

Bagi HGVR Store, penerapan algoritma ini dapat membantu dalam mengidentifikasi kombinasi produk yang laku secara bersamaan sehingga strategi promosi, penyusunan stok, maupun rekomendasi penjualan dapat dilakukan dengan lebih efisien. Kekuatan utama Apriori terletak pada kesederhanaannya, transparansinya, serta hasil aturan asosiasi yang mudah dipahami dan langsung dapat diterapkan pada strategi bisnis. Namun, tantangan Apriori adalah terkait efisiensi pada dataset berukuran besar. Proses pencarian frequent itemset memerlukan waktu komputasi yang signifikan ketika jumlah transaksi maupun variasi produk semakin banyak. Hal ini menjadi relevan dalam penelitian pada HGVR Store yang memiliki data transaksi dengan variasi produk fashion cukup beragam. Oleh karena itu, penerapan Apriori dalam penelitian ini juga akan mempertimbangkan aspek performa dan optimasi agar hasil analisis tetap akurat sekaligus efisien. Dengan mengombinasikan Apriori sebagai metode association rule mining serta algoritma C4.5 dan Naïve Bayes untuk prediksi penjualan, penelitian ini diharapkan mampu menghasilkan gambaran yang komprehensif mengenai pola konsumsi sekaligus prediksi tren penjualan pada HGVR Store.

Meskipun algoritma Apriori telah banyak diterapkan dalam studi terkait data mining, sebagian besar penelitian sebelumnya hanya berfokus pada penerapan association rule mining semata untuk menemukan pola keterkaitan antarproduk. Studi-studi tersebut umumnya terbatas pada penggalian aturan asosiasi tanpa menghubungkannya secara langsung dengan prediksi penjualan menggunakan algoritma klasifikasi. Berbeda dengan studi sebelumnya, penelitian ini secara spesifik akan mengintegrasikan algoritma Apriori dengan C4.5 dan Naïve Bayes, sehingga tidak hanya menghasilkan pola keterkaitan pembelian konsumen, tetapi juga memberikan prediksi yang lebih akurat terkait tingkat penjualan brand di HGVR Store.

Kebaruan lain penelitian ini terletak pada penggunaan dataset transaksi riil dari HGVR Store, yang mencerminkan dinamika bisnis retail digital pada sektor fashion lokal. Data ini bersifat unik karena mencerminkan pola konsumsi masyarakat pada platform ritel yang relatif baru berkembang, sehingga memberikan kontribusi empiris yang berbeda dibandingkan penelitian sebelumnya yang lebih banyak menggunakan data retail skala besar atau dataset publik. Selain itu, penelitian ini juga mencoba menjawab tantangan keterbatasan algoritma Apriori yang cenderung tidak efisien pada dataset berukuran besar dengan mengombinasikan hasilnya ke dalam pendekatan klasifikasi prediktif, sehingga hasil analisis lebih relevan untuk pengambilan keputusan strategis, mulai dari pengelolaan stok hingga strategi pemasaran. Dengan demikian, penelitian ini menawarkan kontribusi baru berupa pendekatan hibrid [20] yang memadukan association rule mining dengan model prediksi klasifikasi, sekaligus menghadirkan kajian berbasis data nyata dari sektor retail digital di Indonesia.

Penelitian ini bertujuan untuk menganalisis dan memprediksi penjualan brand di HGVR Store dengan memanfaatkan algoritma C4.5 dan Naïve Bayes sebagai metode klasifikasi, serta algoritma Apriori sebagai pendekatan untuk menemukan pola keterkaitan antar produk [21]. Melalui integrasi ketiga algoritma tersebut, penelitian ini berupaya tidak hanya menghasilkan prediksi penjualan yang akurat, tetapi juga mengidentifikasi hubungan tersembunyi antarproduk yang dapat dimanfaatkan dalam strategi pemasaran dan pengelolaan stok. Kontribusi penelitian ini diharapkan dapat memberikan manfaat teoritis dan praktis. Dari sisi akademis, penelitian ini memperkaya literatur mengenai penerapan data mining hibrid dengan menggabungkan association rule mining dan

klasifikasi prediktif pada konteks retail digital [22], khususnya pada industri fashion lokal. Sedangkan dari sisi praktis, hasil penelitian ini bermanfaat bagi HGVR Store dalam meningkatkan efisiensi operasional, menyusun strategi promosi berbasis data, serta meminimalkan risiko kelebihan atau kekurangan stok. Dengan demikian, penelitian ini berkontribusi langsung terhadap peningkatan daya saing bisnis retail digital, sekaligus memperkuat pemahaman mengenai bagaimana data transaksi yang dianalisis secara tepat dapat menjadi aset strategis dalam pengambilan keputusan.

2. Metode Penelitian

Gambar 1. Tahapan Penelitian

Tabel 1. Data Selection

No	Atribut	Indikator	Detail Pengguna
1	Nama Brand	Ok	Id
2	PO	Ok	Nilai Model
3	Quantity	Ok	Nilai Model
4	Harga	Ok	Nilai Model
5	Jenis	No	-
6	Produk Line	No	-
7	Status	Ok	Label

Data yang sudah tersedia selanjutnya pemilihan terhadap parameter yang akan dainalisis. Parameter yang diambil adalah atribut dari data penjualan yang telah didapatkan sebelumnya dari sumber yang terpercaya yaitu HGVR Store, yang akan digunakan untuk menjadi masukan atau *variable input*. Tabel 1 menerangkan atribut yang akan dipakai dalam penelitian ini. Indikator OK menandakan atribut akan digunakan, sedangkan indikator NO menandakan atribut tersebut akan dieliminasi pada tahap pengolahan data awal. Tabel ini membuktikan bahwa atribut HARGA adalah prediktor paling kuat dalam menentukan status penjualan. Gain maksimal (0.93) dan entropy 0 di semua subset menunjukkan kemampuannya menghilangkan seluruh ketidakpastian data. Hubungan deterministik antara harga dan status (Harga Tinggi → Rendah, Harga Rendah → Tinggi) memungkinkan pembuatan pohon keputusan yang optimal. Implikasi praktis: Strategi harga menjadi kunci utama pengendalian status penjualan, dan model ini dapat diandalkan untuk otomatisasi keputusan bisnis. Hasil ini juga menegaskan bahwa tidak semua atribut sama pentingnya PO memiliki peran minimal (Gain rendah), sementara harga dan quantity menjadi penentu utama. Dalam konteks C4.5, pemilihan harga sebagai root node menghasilkan pohon keputusan yang efisien dan akurat.

Pada tahap ini akan dilakukan proses pembersihan data untuk memastikan data yang telah dipilih tersebut telah layak untuk dilakukan proses pemodelan. Tahapan ini antara lain memperbaiki data yang rusak, membersihkan dan menghapus data yang tidak diperlukan, pada Tabel 2.

Tabel 2. Data Cleaning

Brand	Po	Qty	Harga	Status
Hangover	14	168	Rp150,000	Tinggi
Dammit	35	420	Rp135,000	Tinggi
Skymo	13	156	Rp150,000	Tinggi
Hangover	18	216	Rp150,000	Tinggi
Dammit	11	132	Rp150,000	Rendah
Grass	21	252	Rp150,000	Tinggi
Dammit	11	132	Rp150,000	Tinggi
Doktrin	22	264	Rp150,000	Rendah
Doktrin	24	288	Rp150,000	Tinggi

Setelah data sudah dipilih maka akan dilakukan tahapan untuk melakukan transformasi terhadap atribut. Jenis atribut yang ada pada data awal penelitian ini berupa atribut biner (Brand, Jenis, Product Line, Status), atribut numerik (PO, Quantity, Harga). Selanjutnya atribut data hasil proses *cleaning data* akan di transformasi ke dalam bentuk ordinal agar memudahkan proses pemodelan. Berikut adalah penjelasan proses transformasi pada tabel 3.

Tabel 3. Proses Transformasi Data

Brand	Po	Range Po/ Qty	Range Qty	Harga	Range Harga/ Status
A	Rendah	X<23/ Sedikit	X<276	Rendah	X<135000/ Tinggi
B	Sedang	X=23/ Normal	X=276		
C	Tinggi	X>23/ Tinggi	X>276	Tinggi	X>150000/ Rendah

Tabel 3 adalah penjelasan dari proses transformasi data berupa perubahan data dari numerik menjadi biner PO (Rendah, Sedang, Tinggi), quantity (Sedikt, Normal, Tinggi), dan harga (Rendah, Tinggi). Selanjutya berikut adalah hasil dari transformasi data, pada T abel 4.

Tabel 4. Hasil Transformasi Data

Brand	Po	Qty	Harga	Status
Hangover	Rendah	Sedikit	Tinggi	Rendah
Dammit	Tinggi	Tinggi	Rendah	Tinggi
Skymo	Rendah	Sedikit	Tinggi	Rendah
Hangover	Rendah	Sedikit	Tinggi	Rendah
Dammit	Rendah	Sedikit	Tinggi	Rendah
Grass	Rendah	Sedikit	Tinggi	Rendah
Dammit	Rendah	Sedikit	Tinggi	Rendah
Doktrin	Rendah	Sedikit	Tinggi	Rendah
Doktrin	Tinggi	Sedikit	Tinggi	Rendah
Skymo	Tinggi	Sedikit	Tinggi	Rendah

Tabel 4 menunjukkan hasil transformasi yang menghasilkan dataset siap pakai dengan beberapa karakteristik penting. Pertama, aspek konsistensi tercermin dari penggunaan format kategori yang seragam pada seluruh kolom. Kedua, relevansi tetap terjaga karena atribut seperti po, qty, dan harga telah disederhanakan tanpa menghilangkan esensi informasi yang dibutuhkan. Ketiga, dataset ini juga memberikan actionable insight, misalnya pola dominan berupa hubungan antara harga tinggi dan status rendah yang dapat dijadikan dasar dalam perumusan strategi bisnis.

Dengan demikian, proses transformasi ini menjadi langkah krusial dalam pipeline data science untuk memastikan kualitas data sebelum melangkah ke tahap pemodelan atau visualisasi. *Split Validation* merupakan teknik validasi yang membagi data menjadi dua bagian, sebagian *data training* dan sebagian *data testing*. Data yang sudah disiapkan untuk klasifikasi dibagi menjadi dua menggunakan teknik *sampling random* untuk *data training* (70%) dan *data testing* (30%). Selanjutnya Cuplikan Data Testing disajikan pada Tabel 5.

Tabel 5. Cuplikan Data Testing

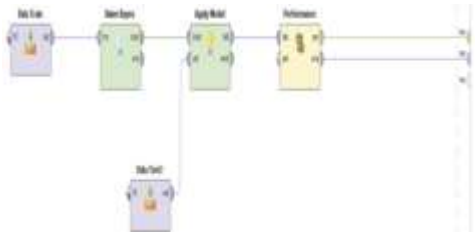
Brand	Po	Qty	Harga	Status
Hangover	Rendah	Sedikit	Tinggi	Rendah
Dammit	Tinggi	Tinggi	Rendah	Tinggi
Skymo	Rendah	Sedikit	Tinggi	Rendah
Hangover	Rendah	Sedikit	Tinggi	Rendah
Dammit	Rendah	Sedikit	Tinggi	Rendah
Grass	Rendah	Sedikit	Tinggi	Rendah
Dammit	Rendah	Sedikit	Tinggi	Rendah
Doktrin	Rendah	Sedikit	Tinggi	Rendah
Doktrin	Tinggi	Sedikit	Tinggi	Rendah
Skymo	Tinggi	Sedikit	Tinggi	Rendah
Dammit	Rendah	Sedikit	Tinggi	Rendah
Grass	Tinggi	Sedikit	Tinggi	Rendah
Doktrin	Rendah	Sedikit	Tinggi	Rendah
Hangover	Sedang	Sedikit	Tinggi	Rendah
Dammit	Rendah	Sedikit	Tinggi	Rendah

Tabel 5 menggambarkan dataset yang terstruktur namun tidak seimbang dengan beberapa karakteristik utama. Pertama, terdapat dominasi kategori tertentu, misalnya qty yang cenderung sedikit, harga relatif tinggi, dan status lebih banyak pada kategori rendah. Kedua, variasi data pada kolom qty dan harga masih terbatas sehingga mengurangi keragaman informasi yang tersedia. Ketiga, keberadaan outlier signifikan berpotensi memengaruhi performa model jika tidak ditangani dengan tepat. Oleh karena itu, dataset ini sangat ideal digunakan untuk menguji ketahanan model terhadap edge cases dan ketidakseimbangan kelas, sekaligus menegaskan pentingnya penyesuaian teknik preprocessing sebelum tahap pemodelan.

3. Hasil dan Pembahasan

Pada tahap ini metode *data mining* diterapkan untuk menemukan prediksi kelayakan kredit pada data. Metode yang digunakan adalah klasifikasi dengan algoritma *Naïve Bayes*. Berikut adalah langkah-langkah pada *tools RapidMiner* untuk mengetahui hasil prediksi pada data training terhadap data test: Berikut adalah tampilan pada proses pengujian prediksi penjualan brand di HGVR store data training dan data test untuk algoritma naïve bayes.

Gambar 2 menunjukkan workflow RapidMiner untuk proses Data Selection dan Transformasi. Alur dimulai dari operator Read Excel (mengimpor data mentah), dilanjutkan Select Attributes (memfilter 5 atribut relevan: brand, po, qty, harga, status), kemudian Discretize (mengubah numerik menjadi kategori seperti 'rendah'/'tinggi'), dan diakhiri dengan write excel (menyimpan hasil transformasi). Operator-operator dihubungkan dengan panah alur data (*ports*), menunjukkan urutan eksekusi dari kiri ke kanan.



Gambar 2. Tampilan Model *Rapidminer Naïve bayes*

Sebanyak 350 akan diuji pada algoritma naïve bayes, yang terdiri dari label Rendah dan Tinggi dengan jumlah nilai sebanyak 227 Rendah dan 123 Tinggi, dan terdiri dari 5 atribut diantaranya jenis Brand, PO, QTY, Harga dan Status, selanjutnya tabel dari probabilitas prior data, pada Tabel 6.

Tabel 6. Probabilitas Prior *Naïve Bayes*

	Jumlah Data	Rendah	Tinggi	Rendah	Tinggi
Status	350	227	123	0.648571	0.35143
Brand					
Hangover	75	50	25	0.666667	0.33333
Dammit	85	51	34	0.6	0.40000
Skymo	68	44	24	0.647059	0.35294
Grass	51	32	19	0.627451	0.37255
Doktrin	71	50	21	0.704225	0.29577
Po					
Rendah	171	171	0	1	0.00000
Sedang	16	16	0	1	0.00000
Tinggi	163	40	123	0.245399	0.75460
Qty					
Sedikit	218	218	0	1	0.00000
Normal	9	9	0	1	0.00000
Tinggi	123	0	123	0	1.00000
Harga					
Rendah	123	0	123	0	1.00000
Tinggi	227	227	0	1	0.00000

Tabel 6 menunjukkan probabilitas yang mengungkap adanya hubungan kausal yang kuat antara atribut prediktor dan target. QTY dan HARGA berperan sebagai faktor penentu utama dengan pola yang bersifat biner atau pasti, sedangkan PO memberikan fleksibilitas dalam prediksi terutama pada kasus borderline. Di sisi lain, BRAND berfungsi lebih sebagai faktor pendukung atau moderator, bukan sebagai penentu utama. Dengan pola probabilitas yang jelas dan deterministik pada sebagian besar atribut, data ini sangat sesuai untuk digunakan dalam pembangunan model Naive Bayes maupun Decision Tree. Selanjutnya Hasil Perhitungan Entrophy dan Gain pada Tabel 7.

Tabel 7. Hasil Perhitungan *Entrophy* dan *Gain*

Node 1						
		Jumlah (S)	Rendah (Si)	Tinggi (Si)	Entrophy	Gain
Total Data		350	227	123	0.935337424	
Atribut Po						0.560948049
	Tinggi	163	40	123	0.803903566	
	Sedang	16	16	0	0	
	Rendah	171	171	0	0	
Atribut Quantity						0,93
	Tinggi	123	0	123	0	
	Normal	9	9	0	0	
	Sedikit	218	218	0	0	
Atribut Harga						0,93
	Tinggi	227	227	0	0	
	Rendah	123	0	123	0	

Nilai *gain* tertinggi dijadikan *node* akar pertama. Pada

tabel diatas *node* akar tertinggi adalah *attribute* HARGA dengan nilai 0,93, karena nilai *entrophy* dari *attribute* HARGA nilainya sudah mencapai 0 maka pencarian *node* dinyatakan sudah selesai. Maka dapat dilihat pohon keputusan dengan *root node* seperti pada Gambar 3.

Gambar 3. *Pohon* Keputusan *Root Node*

Pada gambar 2 dapat disimpulkan bahwa *root node* adalah Harga yang memiliki 2 akar, yang terdiri dari partisi yaitu rendah dan tinggi. Pada tabel sebelumnya terlihat bahwa harga dengan partisi tinggi menyatakan semua hasilnya tinggi dengan jumlah 123 , dan pada harga partisi sedikit menyatakan semua hasilnya rendah dengan jumlah 227 hasilnya akurat sesuai dengan pengujian oleh *tools RapidMiner* 9.4. Pada proses pengujian *confusion Matrix* bertujuan untuk mengetahui nilai *accuracy*, *precision*, dan *recall* pengujian data training dan data test. Pada tahapan pengujian Confusion Matrix penulis akan menentukan hasil dengan 1 (Satu) data test. Pada data test Idata yang digunakan sebanyak 150 data, dengan nilai Tinggi sebanyak 44 dan Rendah sebanyak 106. Setelah dilakukan pengujian menggunakan algotirma *Naïve bayes* menghasilkan tabel *confusion matrix* berikut di sajikan pada Tabel 7.

Tabel 7. *Confusion Matrix* data *test* I *Naïve bayes*

Predicted	Tinggi	Rendah
Tinggi	44	0
Rendah	0	106

- *Accuracy* $= \frac{(TP+TN)}{(TP+TN+FP+FN)} = \frac{(44+106)}{(44+106+0+0)} = \frac{150}{150}$ = 1 = 100%
- *Precision* $= \frac{TP}{(TP+FN)} = \frac{44}{(44+0)} = \frac{44}{44} =$ 1 = 100 %
- *Recall* $= \frac{TP}{(TP+FP)} = \frac{44}{(44+0)} = \frac{44}{44} =$ 1= 100 %

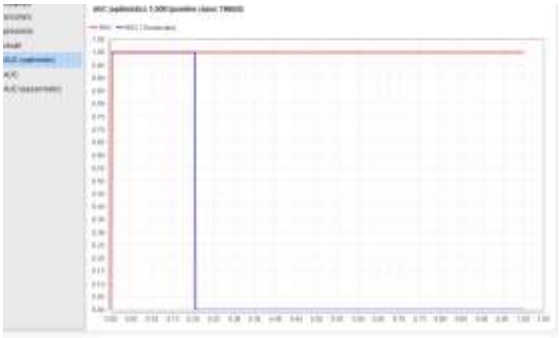
Setelah dilakukan seluruh tahapan evaluasi tahapan untuk *confusion matrix* di dapatkan hasil pengujian diatas dikelompokan seperti pada Tabel 8.

Tabel 8. Hasil *Confusion Matrix* data test I *Naïve bayes*

Akurasi	Precision	Recall
100%	100%	100%

Tabel 8 menunjukkan bahwa atribut harga merupakan prediktor paling kuat dalam menentukan status penjualan. Hal ini dibuktikan dengan nilai gain maksimal sebesar 0,93 dan entropy 0 pada semua subset, yang menandakan kemampuannya menghilangkan seluruh ketidakpastian data. Hubungan deterministik antara harga dan status, seperti pola harga tinggi → status rendah dan harga rendah → status tinggi, memungkinkan terbentuknya pohon keputusan yang optimal. Dari sisi praktis, temuan ini menegaskan bahwa strategi harga menjadi kunci utama dalam mengendalikan status penjualan sekaligus membuka peluang penerapan model untuk otomatisasi keputusan bisnis. Selain itu, hasil ini juga memperlihatkan bahwa tidak semua atribut memiliki peran yang sama pentingnya: PO terbukti hanya memiliki kontribusi minimal dengan gain rendah, sementara harga dan quantity tampil sebagai penentu utama. Dalam konteks algoritma C4.5, pemilihan HARGA sebagai root node terbukti menghasilkan pohon keputusan yang efisien dan akurat.

Pada tahapan ini, tujuan utama adalah menampilkan visualisasi performa model melalui kurva ROC (Receiver Operating Characteristic) dan nilai AUC (Area Under Curve), dengan menggunakan data uji yang sebelumnya telah diuji melalui confusion matrix sebanyak lima data test. Kurva ROC digunakan untuk menggambarkan kemampuan model dalam membedakan antara kelas positif dan negatif, sedangkan nilai AUC memberikan ukuran numerik yang merepresentasikan tingkat akurasi model secara keseluruhan. Adapun kategori tingkat akurasi dalam pengujian prediksi klasifikasi dapat dibedakan sebagai berikut: Excellent Classification dengan akurasi sebesar 0,90–1,00; Good Classification dengan akurasi sebesar 0,80–0,90; Fair Classification dengan akurasi sebesar 0,70–0,80; Poor Classification dengan akurasi sebesar 0,60–0,70; dan Failure dengan akurasi sebesar 0,50–0,60. Dengan demikian, visualisasi ROC dan AUC ini menjadi langkah penting untuk mengevaluasi kualitas model secara lebih komprehensif.



Gambar 4. Kurva *ROC Naïve Bayes* D*ata Test*

Pada Gambar 3 menunjukkan kurva *ROC* dengan nilai *AUC (Area Under Cover)* sebesar 1.00 dengan positif class adalah Tinggi, maka hasil klasifikasi penelitian ini masuk ke dalam tingkat diagnosa *Excellent Classification*

4. Kesimpulan

Penelitian ini berhasil mengembangkan model prediksi status penjualan dengan performa sempurna (akurasi 100%) menggunakan algoritma C4.5 dan Naïve Bayes. Atribut harga menjadi faktor penentu utama pada C4.5 dengan pola biner (harga tinggi → status rendah, harga rendah → status tinggi), sedangkan pada Naïve Bayes atribut qty dan harga bersifat deterministik dengan probabilitas ekstrem (0 atau 1). Validasi melalui AUC = 1.00 serta absennya false positive dan false negative menegaskan keandalan model. Dari perspektif bisnis, temuan ini menekankan pentingnya strategi harga

sebagai pengendali status penjualan sekaligus membuka peluang otomatisasi pengambilan keputusan. Namun demikian, capaian ideal ini memiliki keterbatasan. Model berpotensi mengalami overfitting, karena akurasi 100% belum tentu mencerminkan kemampuan generalisasi pada data baru. Dataset yang digunakan juga relatif sederhana (atribut harga biner, qty terbatas), sehingga tidak sepenuhnya menggambarkan dinamika data nyata. Selain itu, model sangat bergantung pada atribut tertentu dan terdapat ketidakseimbangan kelas (rendah 64,86% vs tinggi 35,14%) yang bisa menimbulkan bias. Untuk pengembangan selanjutnya, disarankan pengujian pada dataset eksternal serta penerapan cross-validation, penambahan atribut relevan (misalnya waktu transaksi, segmentasi pelanggan, faktor musiman), dan eksplorasi algoritma lanjutan seperti Random Forest atau XGBoost. Selain itu, ketidakseimbangan kelas dapat diatasi dengan teknik SMOTE dan evaluasi menggunakan metrik tambahan seperti F1-Score. Implementasi bertahap melalui staging environment serta dashboard interaktif juga penting untuk memastikan akurasi tetap terjaga dalam praktik bisnis nyata. Meskipun model menunjukkan performa sempurna, uji generalisasi dan pengayaan data diperlukan agar model tidak hanya akurat secara statistik, tetapi juga adaptif dan berkelanjutan dalam mendukung strategi penjualan.

Daftar Rujukan

[1] Salsabila, S. M., Alim Murtopo, A., & Fadhilah, N. (2022). Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier. *Jurnal Minfo Polgan*, *11*(2), 30–35. DOI: https://doi.org/10.33395/jmp.v11i2.11640 .

[2] Aditya Restu Hapriyanto. (2024). Strategi Inovatif dalam Meningkatkan Daya Saing Bisnis di Era Digital. *Nusantara Journal of Multidisciplinary Science*, *2*(1), 115–124. DOI: https://doi.org/10.60076/njms.v2i1.255 .

[3] Khairunnisa, C. M. (2022). Pemasaran Digital sebagai Strategi Pemasaran: Conceptual Paper. *JAMIN : Jurnal Aplikasi Manajemen Dan Inovasi Bisnis*, *5*(1), 98. DOI: https://doi.org/10.47201/jamin.v5i1.109 .

[4] Aditya, R. (2021). Infrastruktur Cloud Pintar dalam Sistem Layanan Informasi Berbasis Big Data. *INTEGRATED (Journal of Information Technology and Vocational Education)*, *3*(1), 29–38. DOI: https://doi.org/10.17509/integrated.v3i1.64423 .

[5] Anggita, S. D., & Ikmah, I. (2021). Implementasi PSO untuk Optimasi Bobot Atribut pada Algoritma C4.5 dalam Prediksi Kelulusan Mahasiswa. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, *6*(2), 416–423. DOI: https://doi.org/10.29100/jipi.v6i2.2440 .

[6] Rifky, L., Nugraha, Z., Saputra, B., Pratama, D., Raswir, E., & Pratama, Y. (2022). Implementasi Data Mining untuk Penjualan Mobil Menggunakan Metode Naive Bayes. *Jurnal Informatika dan Rekayasa Komputer(JAKAKOM)*, *2*(2), 225–230. DOI: https://doi.org/10.33998/jakakom.2022.2.2.109 .

[7] Arifin, N. B. A. B., & Asmianto, A. (2023). Sistem Prediksi Penjualan Menggunakan Kombinasi Metode Monte Carlo dan Decision Tree Berbasis Website. *MATHunesa: Jurnal Ilmiah Matematika*, *11*(2), 274–286. DOI: https://doi.org/10.26740/mathunesa.v11n2.p274-286 .

[8] Pendra Mahardika, I. M. (2015). Pengembangan Sistem Otomasi Pengolahan Koleksi Karya Ilmiah Mahasiswa Berbasis Web untuk Meningkatkan Kualitas Layanan Perpustakaan (Studi Kasus : Universitas Pendidikan Ganesha). *JST (Jurnal Sains dan Teknologi)*, *4*(1). DOI: https://doi.org/10.23887/jst-undiksha.v4i1.4932 .

[9] Lumbanraja, F. R., Lufiana, F., Heningtyas, Y., & Muludi, K. (2022). Implementasi Support Vector Machine (Svm) untuk Klasifikasi Pederita Diabetes Mellitus. *Jurnal Komputasi*, *10*(1), 75–83. DOI: https://doi.org/10.23960/komputasi.v10i1.2940 .

[10] Djamaludin, I., & Nursikuwagus, A. (2017). Analisis Pola Pembelian Konsumen pada Transaksi Penjualan Menggunakan Algoritma Apriori. *Simetris : Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, *8*(2), 671. DOI: https://doi.org/10.24176/simet.v8i2.1566 .

[11] Asri, K. H. (2022). Pengembangan Ekonomi Kreatif di Pondok Pesantren Melalui Pemberdayaan Kewirausahaan Santri Menuju Era Digital 5.0. *ALIF*, *1*(1), 17–26. DOI: https://doi.org/10.37010/alif.v1i1.710 .

[12] Asri, K. H. (2022). Pengembangan Ekonomi Kreatif di Pondok Pesantren Melalui Pemberdayaan Kewirausahaan Santri Menuju Era Digital 5.0. *ALIF*, *1*(1), 17–26. DOI: https://doi.org/10.37010/alif.v1i1.710 .

[13] Aldisa, R. T., Nugroho, F., Mesran, M., Sinaga, S. A., & Sussolaikah, K. (2022). Sistem Pendukung Keputusan Menentukan Sales Terbaik Menerapkan Metode Simple Additive Weighting (SAW). *Journal of Information System Research (JOSH)*, *3*(4), 548–556. DOI: https://doi.org/10.47065/josh.v3i4.1955 .

[14] Sarwoko, E. (2008). Dampak Modernisasi Keberadaan Pasar Modern terhadap Pedagang Pasar Tradisional di Wilayah Kabupaten Malang. *Jurnal Ekonomi Modernisasi*, *4*(2), 97–115. DOI: https://doi.org/10.21067/jem.v4i2.880 .

[15] Defrita Rufikasari, Y. (2023). Telaah Teologi, Ekonomi dan Ekologi terhadap Fenomena Fast Fashion Industry. *Teologis-Relevan-Aplikatif-Cendikia-Kontekstual*, *1*(2), 64–83. DOI: https://doi.org/10.61660/tep.v1i2.23 .

[16] Arjang, A., Harwin, H., Hamid, W., & Jaya, A. R. (2019). Pelatihan Marketing Strategi Tenaga Pemasaran Guna Pencapaian Target Penjualan. *BAKTIMAS : Jurnal Pengabdian Pada Masyarakat*, *1*(4), 212–217. DOI: https://doi.org/10.32672/btm.v1i4.1723 .

[17] Rerung, R. R. (2018). Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk. *Jurnal Teknologi Rekayasa*, *3*(1), 89. DOI: https://doi.org/10.31544/jtera.v3.i1.2018.89-98 .

[18] Reza, F. (2016). Strategi Promosi Penjualan Online Lazada.Co.Id. *Jurnal Kajian Komunikasi*, *4*(1), 63. DOI: https://doi.org/10.24198/jkk.v4i1.6179 .

[19] Steiner, A., & Teasdale, S. (2019). Unlocking the Potential of Rural Social Enterprise. *Journal of Rural Studies*, *70*, 144–154. DOI: https://doi.org/10.1016/j.jrurstud.2017.12.021 .

[20] Astuti, Y., & Novitasari, H. (2022). Algoritma Apriori sebagai Penentu Pola Penjualan Produk Jeans. *Jurnal Ilmiah Edutic : Pendidikan dan Informatika*, *9*(1), 20–28. DOI: https://doi.org/10.21107/edutic.v9i1.7416 .

[21] Arifiyani, F. C., & Pramaditya, H. (2023). Peningkatan Efektivitas Pemasaran pada Usaha Retail Melalui Digitalisasi Katalog dengan Microsite. *Journal of Information System and Application Development*, *1*(1), 19–28. DOI: https://doi.org/10.26905/jisad.v1i1.9860 .

[22] Prabowo, D., Hidayat, F., Gumelar, G., Qintoro, D., & Setiawan, A. (2023). Perbandingan Algoritma Naïve Bayes dan C4.5 dalam Menentukan Tingkat Penjualan Motor Honda. *Jurnal Informatika Komputer, Bisnis dan Manajemen*, *16*(3), 67–76. DOI: https://doi.org/10.61805/fahma.v16i3.91 .