

# Penerapan Metode *Naïve Bayes* untuk Klasifikasi Produk Berdasarkan Kategori Penjualan di Toko Artemist

Yuma Akbar <sup>1</sup>, Mariyatul Qibthiyah <sup>2\*</sup>

<sup>1,2\*</sup> Program Studi Sistem Informasi, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia.

*Email:* yuma.pjj@gmail.com <sup>1</sup>, mariyatulqibthiyah28@gmail.com <sup>2\*</sup>

## Histori Artikel:

*Dikirim* 30 Juli 2025; *Diterima dalam bentuk revisi* 15 Agustus 2025; *Diterima* 30 Agustus 2025; *Diterbitkan* 10 September 2025. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

## Abstrak

Penerapan analisis data dalam sektor ritel menjadi kebutuhan penting untuk mendukung pengambilan keputusan yang akurat dan efisien. Penelitian ini bertujuan untuk mengklasifikasikan produk di Toko Artemist ke dalam dua kategori, yaitu diminati dan kurang diminati, menggunakan metode *Naïve Bayes*. Data yang digunakan berupa catatan penjualan selama satu tahun dengan total 8.106 transaksi yang setelah tahap preprocessing menghasilkan 148 produk. Label kelas ditentukan berdasarkan ambang rata-rata jumlah penjualan. Dataset dibagi menggunakan skema stratifikasi 70% data latih (103 produk) dan 30% data uji (45 produk). Algoritma *Naïve Bayes* diimplementasikan pada perangkat lunak RapidMiner Studio. Hasil evaluasi terhadap data uji menunjukkan akurasi 93,33%, dengan precision 89,29% dan recall 100% pada kelas diminati, serta precision 100% dan recall 85% pada kelas kurang diminati. Temuan ini membuktikan bahwa *Naïve Bayes* efektif dalam mengidentifikasi produk dengan tingkat minat konsumen berbeda, sekaligus memberikan manfaat praktis berupa rekomendasi pengelolaan stok, perencanaan promosi, dan strategi pemasaran berbasis data bagi ritel skala menengah.

**Kata Kunci:** *Naïve Bayes*; Klasifikasi Produk; Data Penjualan; Data Mining; Toko Ritel.

## Abstract

The implementation of this model offers practical benefits for stock management, promotional planning, and data-driven product strategy decisions, thereby improving operational efficiency for medium-scale retail businesses. The application of data analysis in the retail sector is essential to support accurate and efficient decision-making. This study aims to classify products at Artemist Store into two categories: high demand and low demand, using the *Naïve Bayes* method. The data used are sales records for one year with a total of 8,106 transactions, which after preprocessing resulted in 148 products. Class labels are determined based on the average sales threshold. The dataset is divided using a stratification scheme of 70% training data (103 products) and 30% test data (45 products). The *Naïve Bayes* algorithm is implemented in RapidMiner Studio software. The evaluation results on the test data show an accuracy of 93.33%, with 89.29% precision and 100% recall in the high demand class, and 100% precision and 85% recall in the low demand class. These findings prove that *Naïve Bayes* is effective in identifying products with different levels of consumer interest, while also providing practical benefits in the form of stock management recommendations, promotional planning, and data-driven marketing strategies for medium-scale retailers.

**Keyword:** *Naïve Bayes*; Product Classification; Sales Data; Data Mining; Retail Store.

## 1. Pendahuluan

Industri ritel mencakup berbagai aktivitas yang melibatkan penjualan barang dan jasa langsung kepada konsumen akhir. Aktivitas ini meliputi pemilihan lokasi, strategi produk, penetapan harga, promosi, serta layanan pelanggan yang dirancang untuk memberikan pengalaman terbaik bagi konsumen (Berman & Evans, 2004). Keberadaan sektor ini sangat penting dalam perekonomian karena mempengaruhi banyak aspek, mulai dari penciptaan lapangan kerja hingga distribusi produk yang mendukung sektor lainnya. Perkembangan teknologi informasi dalam beberapa tahun terakhir membuka banyak peluang bagi bisnis ritel untuk mengelola dan menganalisis data secara lebih efisien. Data penjualan yang sebelumnya hanya digunakan untuk laporan rutin kini dapat dianalisis lebih lanjut untuk memberikan pemahaman yang lebih baik dan mendukung keputusan yang lebih tepat. Dalam dunia data mining, teknik klasifikasi adalah salah satu cara yang digunakan untuk memisahkan data ke dalam kategori tertentu. Proses ini memudahkan untuk mengidentifikasi pola dan hubungan yang ada dalam data, sehingga informasi yang semula sederhana dapat memberikan nilai lebih. Misalnya, dalam konteks bisnis ritel, klasifikasi data penjualan dapat membantu toko menentukan produk mana yang banyak diminati konsumen dan mana yang tidak, serta membantu merancang strategi yang lebih efektif untuk pengelolaan stok. Toko Artemist adalah salah satu usaha ritel pakaian yang memiliki lebih dari 140 *stock keeping* unit (SKU). Meski penjualannya cukup signifikan, data yang ada selama ini hanya digunakan untuk keperluan pelaporan rutin tanpa adanya analisis yang lebih lanjut. Salah satu masalah yang muncul adalah penumpukan produk yang tidak diminati oleh konsumen, yang pada gilirannya dapat mengurangi efisiensi operasional dan menambah biaya. Hal ini disebabkan oleh tidak adanya sistem yang mampu mengidentifikasi produk mana yang seharusnya dipromosikan lebih intensif dan mana yang perlu dikurangi dari stok. Penumpukan produk slow moving ( $\pm 35\%$  dari total SKU) menambah tantangan dalam mengelola persediaan. Untuk mengatasi masalah ini, penelitian ini memilih metode Naïve Bayes, sebuah teknik klasifikasi probabilistik yang efisien dan efektif dalam menangani data besar dengan sumber daya komputasi terbatas. Keunggulan utama Naïve Bayes terletak pada kemampuannya untuk memberikan hasil yang baik meskipun dengan asumsi independensi antaratribut, yang menjadikannya cocok untuk aplikasi di sektor ritel seperti Toko Artemist. Dibandingkan dengan metode lain seperti Decision Tree atau Random Forest, Naïve Bayes lebih sederhana dan mudah diterapkan, meskipun terdapat beberapa keterbatasan yang perlu diperhatikan dalam penggunaannya.

Penelitian ini bertujuan untuk mengklasifikasikan produk Toko Artemist menjadi dua kategori: produk yang diminati dan yang kurang diminati berdasarkan data penjualan yang ada. Dengan cara ini, manajemen toko dapat mengetahui produk mana yang sebaiknya mendapatkan perhatian lebih dalam hal stok dan promosi. Selain itu, penelitian ini juga akan mengevaluasi efektivitas Naïve Bayes dalam membangun model klasifikasi yang akurat dan mampu mengidentifikasi produk dengan berbagai pola musiman yang memengaruhi penjualan. Penting untuk dicatat bahwa penelitian ini berbeda dari studi-studi sebelumnya yang lebih banyak berfokus pada produk fast moving atau produk yang sering terjual. Fokus penelitian ini akan lebih mendalam pada produk yang kurang diminati, kategori yang seringkali tidak mendapat perhatian cukup dalam penelitian sebelumnya. Dengan menggunakan data riil dari sektor ritel fashion skala menengah, penelitian ini mencoba memberikan gambaran yang lebih jelas mengenai tantangan yang dihadapi oleh bisnis ritel seperti Toko Artemist dalam mengelola produk-produk yang memiliki tingkat permintaan rendah. Hasil dari penelitian ini diharapkan tidak hanya dapat memperbaiki pengelolaan stok dan perencanaan promosi di Toko Artemist, tetapi juga dapat digunakan sebagai dasar bagi pengembangan sistem yang lebih canggih di masa mendatang, serta memberikan pendekatan berbasis data yang lebih kuat untuk pengambilan keputusan yang lebih tepat.

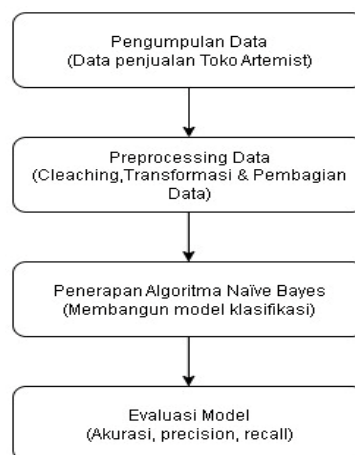
## 2. Metode Penelitian

Data penelitian ini diperoleh dari catatan penjualan historis Toko Artemist selama satu tahun dengan total 8.106 transaksi. Data tersebut kemudian diproses dan diolah untuk menghasilkan 148 produk unik setelah melalui tahap pembersihan yang meliputi penghapusan duplikasi, penghilangan entri yang tidak relevan, penghapusan produk yang tidak terjual, serta penyeragaman penamaan produk. Langkah-langkah pembersihan ini bertujuan untuk memastikan bahwa data yang digunakan konsisten dan siap untuk analisis lebih lanjut. Beberapa atribut yang digunakan dalam penelitian ini meliputi kategori produk, model pakaian, metode pembayaran, dan harga. Agar tidak terjadi target leakage, jumlah penjualan yang menjadi dasar dalam pelabelan produk tidak digunakan sebagai fitur input dalam model. Variabel numerik berupa harga kemudian didiskretisasi dengan metode *equal frequency*, membagi harga ke dalam lima interval untuk menyesuaikan kebutuhan algoritma Naïve Bayes. Setelah tahap pra-pemrosesan, dataset dibagi menjadi dua bagian: 70% untuk data latih dan 30% untuk data uji. Pembagian data dilakukan secara stratifikasi untuk memastikan distribusi kelas tetap seimbang. Dengan pembagian ini, data latih yang digunakan berjumlah 103 produk, sementara data uji berjumlah 45 produk. Jumlah data latih dan data uji yang digunakan pada penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Jumlah Data Latih & Data Uji

Jenis Data	Persentase	Jumlah
Data Latih	70%	103
Data Uji	30%	45

Metode yang digunakan dalam penelitian ini adalah Naïve Bayes. Algoritma ini dipilih karena kesesuaiannya dengan data yang telah dikategorikan dan kemampuannya untuk mengolah data besar dengan sumber daya komputasi yang terbatas. Implementasi Naïve Bayes dilakukan menggunakan perangkat lunak *RapidMiner Studio* versi 9.0. Proses yang dilakukan mencakup beberapa tahapan, yaitu pemuatan data, pengaturan peran atribut, pembagian data secara stratifikasi, pelatihan model, penerapan model pada data uji, dan evaluasi hasil. Evaluasi kinerja model dilakukan dengan menggunakan *confusion matrix* untuk memperoleh metrik evaluasi seperti akurasi, precision, recall, dan F1-score untuk setiap kelas. Selain itu, sebagai pembanding, digunakan akurasi baseline yang dihitung berdasarkan kelas mayoritas. Model dikatakan berhasil apabila mampu mencapai akurasi minimal 80% dan menunjukkan performa yang seimbang pada kedua kelas, dengan perhatian khusus pada kelas "kurang diminati", yang sangat penting dalam mengoptimalkan pengelolaan stok dan strategi promosi. Proses penerapan model secara keseluruhan dapat dilihat pada Gambar 1.



Gambar 1. Model Proses Penelitian

### 3. Hasil dan Pembahasan

#### 3.1 Hasil

##### 3.1.1 Tahap Pengumpulan Data

Tahap pengumpulan data dalam penelitian ini dilakukan dengan menggunakan catatan penjualan historis Toko Artemist yang mencatat transaksi penjualan dan interaksi pelanggan selama satu tahun penuh. Dari total data yang tersedia, dipilih lima atribut utama yang dianggap relevan untuk proses klasifikasi, yaitu model pakaian, kategori produk, harga, metode pembayaran, dan jumlah penjualan. Pemilihan atribut ini didasarkan pada kaitannya yang kuat dengan analisis minat pembeli, yang menjadi fokus utama penelitian. Data yang terkumpul kemudian disusun dalam format terstruktur untuk memudahkan proses preprocessing dan penerapan algoritma klasifikasi pada tahap berikutnya. Gambar 2 menampilkan contoh data penjualan tahun 2024 yang digunakan dalam penelitian ini.

No	Model Pakaian	Kategori	Harga	Pembayaran	Jumlah Penjualan	Label
1	Lingerie Satin Motif	Lingerie	Sedang	Transfer	395	Diminati
2	Dress Katun Polos	Dress	Sedang	Transfer	306	Diminati
3	Kaos Bunga	Top	Murah	Transfer	183	Diminati
4	Baju Tidur Luxury Silk	Baju Tidur	Sedang	Transfer	279	Diminati
5	Dress Katun Polos	Dress	Sedang	Shopee	258	Diminati
6	Dress Formal	Dress	Sedang	Transfer	135	Diminati
8	Baju Tidur Motif	Baju Tidur	Sedang	Shopee	257	Diminati
9	Baju Tidur Polos	Baju Tidur	Sedang	Shopee	296	Diminati
10	Kaos Oversize	Top	Murah	Transfer	114	Diminati
11	Dress Bunga	Dress	Sedang	Transfer	223	Diminati
12	Lingerie Transparan	Lingerie	Murah	Transfer	234	Diminati
13	Dress Satin	Dress	Sedang	Transfer	65	Diminati
15	Rok Panjang Vintage	Rok	Murah	Transfer	261	Diminati
17	Gaun Midi Motif	Gaun	Sedang	Transfer	12	Kurang Diminati
18	Kemeja Kotak	Top	Murah	Shopee	184	Diminati
19	Kaos Polos	Top	Murah	Transfer	29	Kurang Diminati
20	Rok Panjang Vintage	Rok	Murah	Shopee	212	Diminati
22	Kemeja Kotak	Top	Murah	Transfer	206	Diminati
23	Dress Bunga	Dress	Sedang	Shopee	161	Diminati
24	Lingerie Satin Motif	Lingerie	Sedang	Shopee	303	Diminati
26	Baju Tidur Polos	Baju Tidur	Sedang	Transfer	368	Diminati
27	Dress Silk Premium	Dress	Sedang	Transfer	40	Kurang Diminati
33	Gaun Bermotif	Gaun	Sedang	Transfer	34	Kurang Diminati
38	Lingerie Transparan	Lingerie	Murah	Shopee	157	Diminati
39	Gaun Panjang Satin	Gaun	Mahal	Shopee	6	Kurang Diminati
42	Kaos Vintage	Top	Murah	Shopee	6	Kurang Diminati
43	Baju Tidur Premium	Baju Tidur	Sedang	Transfer	143	Diminati
46	Rok Formal Premium	Rok	Murah	Transfer	444	Diminati
58	Gaun Pesta Standar	Gaun	Sedang	Transfer	20	Kurang Diminati

Gambar 2. Contoh Data Penjualan 2024

##### 3.1.2 Tahap Preprocessing Data

Tahap *preprocessing* data merupakan langkah awal yang sangat penting sebelum melakukan analisis menggunakan algoritma klasifikasi *Naïve Bayes*. Tujuan utama dari tahapan ini adalah untuk memastikan bahwa data siap digunakan secara optimal dalam proses klasifikasi. Proses dimulai dengan *data cleaning*, yaitu menghapus produk yang tidak memiliki riwayat penjualan karena dianggap tidak relevan, mempertahankan hanya kategori produk yang sesuai dengan ruang lingkup penelitian, menghapus entri duplikat untuk menghindari bias, serta menyetarakan penamaan produk agar konsisten dan sesuai dengan standar sistem. Selanjutnya, dilakukan transformasi data dengan mengubah nilai numerik pada beberapa atribut menjadi bentuk narasi. Selain itu, dilakukan penambahan label klasifikasi "Diminati" atau "Kurang Diminati" berdasarkan nilai ambang rata-rata penjualan. Produk yang memiliki jumlah penjualan sama dengan atau lebih tinggi dari rata-rata dikategorikan sebagai "Diminati", sementara yang berada di bawah rata-rata digolongkan ke dalam kategori "Kurang Diminati". Atribut yang tidak memberikan kontribusi langsung terhadap klasifikasi, seperti kolom Nama Produk dan No, dikeluarkan (*excluded*) untuk menghindari gangguan dalam pembentukan pola. Tahap akhir dalam preprocessing adalah pembagian data menggunakan metode *split validation* dengan rasio 70:30. Artinya, 70% dari data digunakan untuk pelatihan model (*training set*), sementara 30% sisanya digunakan untuk pengujian model (*testing set*). Pembagian ini bertujuan untuk mengevaluasi kinerja model secara objektif terhadap data yang belum pernah dipelajari sebelumnya dan untuk mengukur kemampuan generalisasi dari algoritma. Data uji ditampilkan pada Gambar 3, sementara data latih dapat dilihat pada Gambar 4.

Kategori	Harga	Pembayaran	Jumlah Penjualan	Label
Lingerie	Sedang	Transfer	395	Diminati
Dress	Sedang	Transfer	306	Diminati
Top	Murah	Transfer	183	Diminati
Baju Tidur	Sedang	Transfer	279	Diminati
Dress	Sedang	Shopee	258	Diminati
Dress	Sedang	Transfer	135	Diminati
Baju Tidur	Sedang	Shopee	257	Diminati
Lingerie	Sedang	Transfer	26	Kurang Diminati
Top	Sedang	Transfer	3	Kurang Diminati
Gaun	Sedang	Transfer	12	Kurang Diminati
Gaun	Sedang	Transfer	7	Kurang Diminati
Gaun	Sedang	Shopee	5	Kurang Diminati
Lingerie	Sedang	Shopee	14	Kurang Diminati
Gaun	Sedang	Shopee	22	Kurang Diminati
Lingerie	Murah	Transfer	149	Diminati
Gaun	Mahal	Transfer	7	Kurang Diminati
Dress	Sedang	Transfer	14	Kurang Diminati
Gaun	Sedang	Shopee	14	Kurang Diminati
Top	Sedang	Shopee	14	Kurang Diminati
Gaun	Sedang	Shopee	1	Kurang Diminati
Top	Sedang	Transfer	2	Kurang Diminati
Top	Murah	Shopee	73	Diminati
Dress	Sedang	Transfer	30	Kurang Diminati
Dress	Sedang	Shopee	17	Kurang Diminati
Dress	Sedang	Shopee	26	Kurang Diminati
Top	Sedang	Transfer	16	Kurang Diminati
Rok	Murah	Shopee	29	Kurang Diminati
Gaun	Sedang	Shopee	11	Kurang Diminati
Top	Sedang	Shopee	17	Kurang Diminati
Top	Murah	Transfer	12	Kurang Diminati
Gaun	Mahal	Shopee	6	Kurang Diminati
Gaun	Mahal	Transfer	5	Kurang Diminati
Dress	Sedang	Shopee	53	Kurang Diminati
Rok	Murah	Transfer	45	Kurang Diminati

Gambar 3. Data Uji

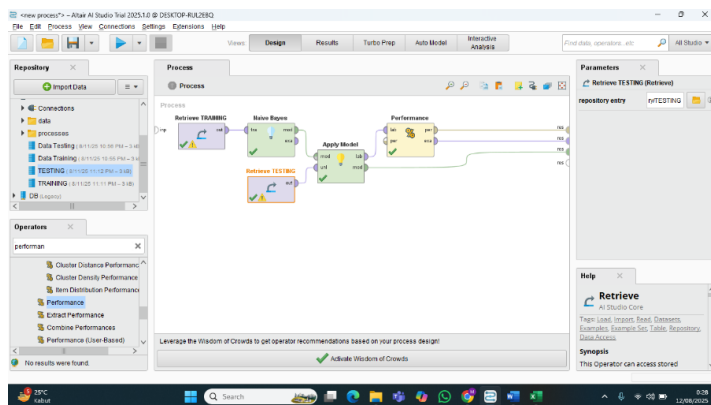
Kategori	Harga	Pembayaran	Jumlah Penjualan	Label
Baju Tidur	Sedang	Shopee	236	Diminati
Top	Murah	Transfer	114	Diminati
Dress	Sedang	Transfer	223	Diminati
Lingerie	Murah	Transfer	234	Diminati
Dress	Sedang	Transfer	65	Diminati
Rok	Murah	Transfer	261	Diminati
Gaun	Sedang	Transfer	12	Kurang Diminati
Top	Murah	Shopee	184	Diminati
Top	Murah	Transfer	29	Kurang Diminati
Rok	Murah	Shopee	212	Diminati
Top	Murah	Transfer	206	Diminati
Dress	Sedang	Shopee	161	Diminati
Lingerie	Sedang	Shopee	303	Diminati
Baju Tidur	Sedang	Transfer	368	Diminati
Dress	Sedang	Transfer	40	Kurang Diminati
Gaun	Sedang	Transfer	34	Kurang Diminati
Lingerie	Murah	Shopee	157	Diminati
Gaun	Mahal	Shopee	6	Kurang Diminati
Top	Murah	Shopee	6	Kurang Diminati
Baju Tidur	Sedang	Transfer	143	Diminati
Rok	Murah	Transfer	444	Diminati
Gaun	Sedang	Transfer	20	Kurang Diminati
Rok	Murah	Shopee	305	Diminati
Dress	Sedang	Shopee	34	Diminati
Gaun	Sedang	Shopee	8	Kurang Diminati
Lingerie	Murah	Shopee	130	Diminati
Dress	Sedang	Transfer	37	Kurang Diminati
Top	Murah	Shopee	92	Diminati
Top	Murah	Transfer	10	Kurang Diminati
Baju Tidur	Sedang	Shopee	211	Diminati
Top	Murah	Transfer	147	Diminati
Gaun	Sedang	Transfer	18	Kurang Diminati
Baju Tidur	Sedang	Shopee	126	Diminati
Dress	Sedang	Shopee	87	Diminati

Gambar 4. Data Latih

### 3.1.3 Penerapan Algoritma *Naïve Bayes*

Tahap penerapan algoritma *Naïve Bayes* dalam penelitian ini dilakukan dengan mengimpor data penjualan yang telah melalui proses preprocessing dan diberi label klasifikasi “Diminati” serta “Kurang Diminati” ke dalam aplikasi *RapidMiner*. Proses dimulai dengan memuat data latih (*training data*) dan data uji (*testing data*) menggunakan operator *Read CSV* yang terhubung dengan file dataset berformat CSV. Data latih kemudian diproses menggunakan operator *Naïve Bayes* untuk membangun model klasifikasi berdasarkan distribusi probabilistik atribut-atribut dalam dataset. Model yang terbentuk diaplikasikan pada data uji melalui operator *Apply Model* untuk menguji kemampuan prediksi terhadap data yang belum pernah dikenali sebelumnya. Selanjutnya, operator *Performance* digunakan untuk mengevaluasi hasil klasifikasi dengan metrik akurasi, precision, dan recall, sehingga dapat diukur tingkat ketepatan dan kelengkapan model dalam memprediksi kategori produk.

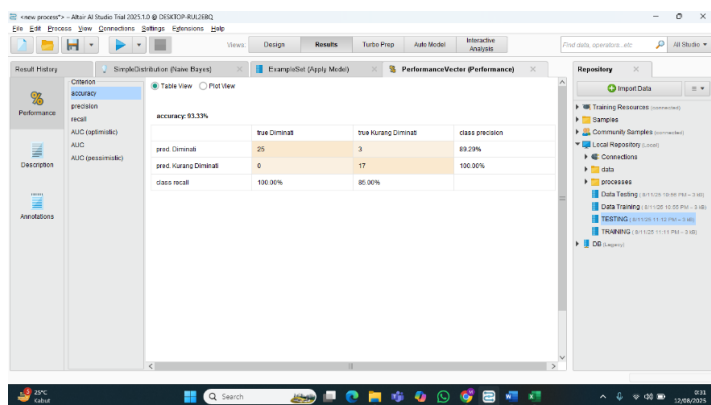
Alur ini menggambarkan proses penerapan *Naïve Bayes* secara terstruktur, mulai dari penginputan data, pembentukan model, pengujian model, hingga penilaian performa. Tujuan dari seluruh tahapan ini adalah untuk memastikan model dapat memberikan hasil klasifikasi yang akurat dan andal. Tampilan *Model Analysis* di *RapidMiner* disajikan pada Gambar 5.



Gambar 5. Tampilan Model Analysis

### 3.1.4 Evaluasi Model

Tahap evaluasi model pada penelitian ini bertujuan mengukur kinerja algoritma *Naïve Bayes* dalam mengklasifikasikan produk ke dalam kategori “*Diminati*” dan “*Kurang Diminati*” menggunakan *confusion matrix* serta metrik akurasi, *precision*, dan *recall*. Berdasarkan hasil pengujian terhadap 45 data uji, diperoleh *confusion matrix* sebagai berikut: prediksi *Diminati* terhadap data sebenarnya *Diminati* (True Positive/TP) sebanyak 25, prediksi *Kurang Diminati* terhadap data sebenarnya *Kurang Diminati* (True Negative/TN) sebanyak 17, prediksi *Diminati* terhadap data sebenarnya *Kurang Diminati* (False Positive/FP) sebanyak 3, dan tidak terdapat kasus prediksi *Kurang Diminati* terhadap data sebenarnya *Diminati* (False Negative/FN = 0). Nilai akurasi yang dihasilkan mencapai 93,33%, menunjukkan bahwa sebagian besar prediksi model sesuai dengan label sebenarnya. Untuk kelas *Kurang Diminati*, nilai *precision* mencapai 100% dan *recall* sebesar 85%, yang berarti seluruh prediksi *Kurang Diminati* benar, meskipun masih ada sebagian data kelas ini yang terlewat teridentifikasi. Sementara itu, untuk kelas *Diminati* diperoleh *precision* sebesar 89,29% dan *recall* 100%, yang mengindikasikan bahwa seluruh data *Diminati* berhasil dikenali, meskipun terdapat sedikit kesalahan prediksi pada kelas ini. Secara keseluruhan, evaluasi menunjukkan bahwa algoritma *Naïve Bayes* mampu memberikan hasil klasifikasi yang sangat baik, terutama dalam mendeteksi produk *Diminati* dan *Kurang Diminati*, sehingga dapat diandalkan sebagai alat bantu pengambilan keputusan dalam manajemen stok dan strategi pemasaran di Toko Artemist.



Gambar 6. Performance Implementasi Naïve Bayes

Penelitian ini menunjukkan bahwa identifikasi produk kurang diminati di Toko Laris Eksis berhasil dilakukan menggunakan metode klasifikasi Naïve Bayes yang diimplementasikan melalui aplikasi RapidMiner. Data penjualan produk dikumpulkan, diberi label “Diminati” atau “Kurang Diminati” berdasarkan rata-rata penjualan, kemudian dibagi menggunakan metode *split validation* dengan rasio 70% data latih dan 30% data uji. Dari 73 data uji, model berhasil memprediksi dengan benar 60 data kategori Kurang Diminati dan 11 data kategori Diminati, sedangkan 2 data kategori Diminati diprediksi keliru sebagai Kurang Diminati. Berdasarkan *confusion matrix*, diperoleh akurasi sebesar 97,26%, *precision* untuk kelas Kurang Diminati mencapai 100% dengan *recall* sebesar 95,56%, sedangkan kelas Diminati memiliki *precision* sebesar 84,62% dan *recall* sebesar 100%. Hasil ini membuktikan bahwa Naïve Bayes mampu memberikan klasifikasi yang sangat baik, khususnya dalam mendeteksi produk Kurang Diminati, sehingga dapat digunakan sebagai alat bantu pengambilan keputusan manajemen stok. Dengan proses klasifikasi yang cepat dan akurat, metode ini mampu memproses data penjualan secara efisien, memberikan prediksi tingkat permintaan produk secara tepat, serta membantu menentukan strategi penambahan atau pengurangan stok untuk meningkatkan efektivitas pengelolaan persediaan. Dataset awal berjumlah 8.106 transaksi penjualan, yang setelah melalui tahap pembersihan dan agregasi menghasilkan 148 produk unik. Label kelas ditentukan menggunakan threshold rata-rata penjualan tahunan, sehingga 92 produk masuk kategori *Diminati* dan 56 produk masuk kategori *Kurang Diminati*. Setelah pembagian data secara stratifikasi dengan perbandingan 70%:30%, distribusi data disajikan pada Tabel 2.

Tabel 2. Distribusi Data Latih dan Uji

Kelas	Data Latih	Data Uji	Total
Diminati	67	25	92
Kurang Diminati	36	20	56
Total	103	45	148

Analisis distribusi menunjukkan bahwa produk dalam kategori *Diminati* cenderung memiliki harga yang relatif bervariasi namun tetap berada pada kisaran menengah, dengan metode pembayaran dominan melalui transfer dan e-wallet. Sebaliknya, produk *Kurang Diminati* lebih banyak terdapat pada kategori model tertentu seperti rok atau gaun yang bersifat musiman, dengan harga cenderung lebih tinggi namun tingkat rotasi penjualan rendah. Rata-rata penjualan produk diminati berada di atas ambang mean 53 unit per tahun, sedangkan produk kurang diminati umumnya hanya terjual di bawah 30 unit. Hal ini menegaskan bahwa pemilihan threshold rata-rata cukup efektif untuk memisahkan kedua kategori. Model Naïve Bayes diuji menggunakan 45 data uji, menghasilkan *confusion matrix* pada Tabel 3.

Tabel 3. Confusion Matrix

	True Kurang Diminati	True Diminati
Pred Kurang Diminati	17	0
Pred Diminati	3	25

Dari tabel tersebut diperoleh metrik kinerja sebagai berikut:

Akurasi : 93,33%

Precision (Diminati) : 89,29%

Recall (Diminati) : 100%

Precision (Kurang Diminati) : 100%

Recall (Kurang Diminati) : 85%

Jika seluruh produk diprediksi masuk kelas mayoritas (*Diminati* dengan 92 produk), maka baseline akurasi hanya sebesar  $(92/148) \times 100\% = 62,16\%$ . Dengan demikian, penerapan Naïve Bayes

meningkatkan akurasi relatif sebesar +31,17% dibanding baseline, sekaligus memberikan kemampuan nyata dalam mengenali kelas minoritas (*Kurang Diminati*).

### 3.2 Pembahasan

Penerapan algoritma *Naïve Bayes* pada data penjualan di Toko Artemist menghasilkan akurasi sebesar 93,33%, jauh melampaui ambang minimal yang ditetapkan. Model mampu mengenali produk yang diminati dengan tingkat recall 100%, serta menunjukkan ketelitian maksimal dalam mengidentifikasi barang yang kurang laku melalui precision 100%. Hampir seluruh produk yang sering dibeli terdeteksi secara tepat, sementara sebagian kecil barang dengan penjualan rendah masih terlewat. Perbandingan dengan baseline accuracy sebesar 62,16% memperlihatkan lonjakan performa sekitar 31%. Hasil ini menegaskan keampuhan *Naïve Bayes* dalam menemukan produk yang jarang terjual, yang sering kali menjadi sumber masalah persediaan dan potensi kerugian jika tidak segera diantisipasi (Wardani *et al.*, 2023; Abdullah *et al.*, 2022). Studi serupa yang dilakukan oleh No *et al.* (2024), Kuota *et al.* (2023), dan Persediaan *et al.* (2022) juga mencatat tingkat akurasi tinggi pada klasifikasi produk terlaris maupun barang yang kurang diminati. Dari sisi efisiensi, *Naïve Bayes* menawarkan proses komputasi yang ringan tanpa mengorbankan ketepatan klasifikasi. Penelitian oleh Akbar & Sugiharto (2023) serta Vibrianti *et al.* (2024) menunjukkan bahwa algoritma ini tetap kompetitif dibandingkan metode lain seperti *Decision Tree* atau *Random Forest*, terutama untuk data penjualan dengan variabel yang beragam. Atribut seperti kategori, model pakaian, harga yang telah dikelompokkan, dan metode pembayaran turut membantu membedakan pola pembelian pelanggan. Produk dengan model tertentu, misalnya rok atau gaun berharga lebih tinggi, cenderung kurang diminati, sedangkan pakaian kasual dengan harga menengah lebih laris. Hal tersebut selaras dengan hasil penelitian Rahayu *et al.* (2018) dan Muttaqin *et al.* (2023), yang menekankan pentingnya pemilihan fitur yang tepat untuk meningkatkan hasil klasifikasi. Berbeda dari penelitian yang hanya berfokus pada produk terlaris (No *et al.*, 2024; Kuota *et al.*, 2023), penelitian ini juga memperhatikan barang yang kurang diminati. Identifikasi kelompok ini sangat berguna bagi pengelola toko, terutama dalam mengatur stok dan merancang strategi pemasaran agar tidak terjadi penumpukan barang yang sulit terjual (Musfita *et al.*, n.d.; Syahputra, 2022). Penggunaan *Naïve Bayes* terbukti memberikan hasil yang solid dalam membantu pengambilan keputusan di bisnis ritel, khususnya dalam pengelolaan inventori dan perencanaan promosi. Pendekatan berbasis data seperti ini semakin relevan bagi perusahaan yang ingin meningkatkan efisiensi operasional dan daya saing di era digital (Muhammad Arhami *et al.*, 2020).

## 4. Kesimpulan

Penelitian yang dilakukan berhasil mengklasifikasikan produk-produk di Toko Artemist berdasarkan data penjualan selama satu tahun, yang setelah proses pembersihan dan agregasi mencakup 148 produk. Setiap produk dikelompokkan ke dalam dua kategori, yaitu diminati dan kurang diminati, menggunakan rata-rata penjualan tahunan sebagai batas pemisah. Proses pengolahan data dilakukan secara bertahap, mulai dari pengumpulan, pembersihan, pelabelan, pemisahan data, pelatihan model, hingga evaluasi. Hasilnya, model *Naïve Bayes* yang diterapkan menunjukkan performa yang sangat baik. Pengujian pada 45 data uji menghasilkan akurasi sebesar 93,33%. Untuk kategori diminati, precision mencapai 89,29% dan recall 100%, sedangkan pada kategori kurang diminati, precision 100% dan recall 85%. Nilai F1-score pada kedua kelas juga seimbang. Temuan ini membuka peluang penerapan sistem klasifikasi otomatis yang dapat membantu pengelolaan stok serta mendukung strategi pemasaran. Model yang dibangun mampu mengenali produk-produk yang kurang diminati, sehingga manajemen dapat segera mengambil langkah seperti promosi, mengurangi pembelian ulang produk dengan perputaran lambat, atau merencanakan penyesuaian harga. Seluruh tujuan penelitian telah tercapai: mulai dari identifikasi produk, penerapan algoritma klasifikasi, hingga evaluasi kinerja metode *Naïve Bayes*.

Bagi pemilik toko, penggunaan metode *Naive Bayes* secara rutin sangat dianjurkan agar tren penjualan dapat dipantau secara berkala. Peneliti selanjutnya dapat memperluas cakupan dengan menerapkan *k-fold cross validation stratifikasi* untuk memperoleh gambaran performa model yang lebih stabil. Selain itu, evaluasi berbasis biaya juga penting dipertimbangkan, mengingat dampak kesalahan klasifikasi tidak selalu seimbang antara satu kategori dengan kategori lain. Pengembangan ambang klasifikasi yang lebih fleksibel sesuai jenis produk juga dapat menjadi langkah berikutnya. Perbandingan dengan metode lain seperti *Decision Tree*, *Random Forest*, atau *SVM* juga layak dicoba untuk mendapatkan hasil yang lebih optimal. Terakhir, integrasi hasil prediksi ke dalam dasbor pemantauan akan sangat membantu pengambilan keputusan secara langsung di tingkat manajerial.

## 5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada kedua orang tua tercinta, Ibu Ciknah dan Bapak Komarudin, atas doa, dukungan, serta Adik tercinta Ahmad Muzaki dan Lailatul Komariyah dan kasih sayang yang tiada henti. Ucapan terima kasih juga disampaikan kepada dosen pembimbing, Bapak Yuma, yang telah memberikan arahan, bimbingan, serta motivasi selama proses penelitian ini. Penulis juga berterima kasih kepada teman terbaik, Samsir Alamsah serta Teman-teman atas dukungan moral dan semangat yang selalu diberikan. Tidak lupa, penulis menyampaikan apresiasi yang tulus kepada teman-teman tercinta yang selalu menemani dan memberikan dukungan hingga penelitian ini dapat terselesaikan dengan baik.

## 6. Daftar Pustaka

- Abdullah, R. W., Hartanti, D., Permatasari, H., Septyanto, A. W., & Abi Bagaskara, Y. (2022). Penerapan Data Mining untuk Memprediksi Jumlah Produk Terlaris Menggunakan Algoritma Naive Bayes Studi Kasus (Toko Prapti). *Jurnal Ilmiah Informatika Global*, 13(1). <https://doi.org/10.36982/jiig.v13i1.2060>.
- Akbar, Y., & Sugiharto, T. (2023). Analisis Sentimen Pengguna Twitter di Indonesia Terhadap ChatGPT Menggunakan Algoritma C4. 5 dan Naive Bayes. *Jurnal Sains dan Teknologi*, 5(1), 115-122.
- Alam, S., & Sunardi, S. (2023, September). Analisis Prediksi Penjualan Kue menggunakan Metode Naive Bayes. In *SISITI: Seminar Ilmiah Sistem Informasi dan Teknologi Informasi* (Vol. 12, No. 2, pp. 54-67). <https://doi.org/10.36774/sisiti.v12i2.1327>.
- Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan dan Minuman Menggunakan Metode Algoritma Naive Bayes: Studi Kasus: Makan Barbeque Sepuasnya. *Jurnal Ilmiah Informatika*, 9(02), 75-81.
- Arhami, M., Kom, M., & Muhammad Nasir, S. T. (2020). *Data Mining-Algoritma dan Implementasi*. Penerbit Andi.
- Harahap, F., Fahrozi, W., Adawiyah, R., Siregar, E. T., & Harahap, A. Y. N. (2023). Implementasi Data Mining dalam Memprediksi Produk AC Terlaris untuk Meningkatkan Penjualan Menggunakan Metode Naive Bayes. *Jurnal Unitek*, 16(1), 41-51. <https://doi.org/10.52072/unitek.v16i1.541>.
- Hidayat, B. (2021). Implementasi Data Mining Untuk Memprediksi Penjualan Pakaian Menggunakan Algoritma Naive Bayes. In *ITB Graduate School Conference* (Vol. 1, No. 1, pp. 663-673).

- Husaini, A. P., & Lisdiyanto, A. (2024). Sistem Prediksi Penjualan Produk APD Terlaris di PT A3 Karunia Sidoarjo menggunakan Metode Naive Bayes. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 6(2), 431-437. <https://doi.org/10.47233/jteksis.v6i2.1266>.
- Mahmudati, R., Rohman, S., & Sa'adah, I. (2025). SISTEM PREDIKSI HASIL LABA PENJUALAN DI UNSIQ MART MENGGUNAKAN METODE NAIVE BAYES. *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, 4(1), 29-33. <https://doi.org/10.55123/storage.v4i1.4849>.
- Musfita, N., Fitriyani, N., & Baskara, Z. W. (2023). Klasifikasi Penjualan Provider Pulsa di Kecamatan Masbagik Lombok Timur Menggunakan Metode Naive Bayes. *ESTIMASI: Journal of Statistics and Its Application*, 261-272.
- Purnama, P. A. W., & Putra, T. A. (2024). Klasifikasi Penjualan Produk Menggunakan Algoritma Naive Bayes pada Konter HP Bayu Cell. *REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer*, 8(1), 286-292. <https://doi.org/10.33395/remik.v8i1.13207>.
- Rahman, M. A. F., Mair, Z. R., & Sartika, D. (2024). Klasifikasi Ulasan Pelanggan Shopee Mall Terhadap E-Commerce Penjualan Baju Batik Metode Naive Bayes. *IDEALIS: InDonEsiA journal Information System*, 7(2), 164-177. <https://doi.org/10.36080/idealisis.v7i2.3178>.
- Rifky, L., Nugraha, Z., Pratama, D., & Raswir, E. (2022). Implementasi Data Mining Untuk Penjualan Mobil Menggunakan Metode Naive Bayes. *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, 2(2), 225-230. <https://doi.org/10.33998/jakakom.2022.2.2.109>.
- Saleh, A. (2015). Implementasi metode klasifikasi naive bayes dalam memprediksi besarnya penggunaan listrik rumah tangga. *Creative Information Technology Journal*, 2(3), 207-217.
- Saputra, M. J., & Herdiansyah, M. I. (2022). Penerapan Naive Bayes Dalam Memprediksi Penjualan Dan Persediaan Kain Jumputan Pada Toko Batiq Colet Tuan Kentang Palembang. *Jurnal Mantik*, 6(2), 2502-2507.
- Saputra, M. J., & Herdiansyah, M. I. (2022). Penerapan Naive Bayes Dalam Memprediksi Penjualan Dan Persediaan Kain Jumputan Pada Toko Batiq Colet Tuan Kentang Palembang. *Jurnal Mantik*, 6(2), 2502-2507.
- Senika, A., Rasiban, R., & Iskandar, D. (2022). Implementasi metode Naive Bayes dalam penilaian kinerja sales marketing pada PT. Pachira Distrinusa. *Jurnal Media Informatika Budidarma*, 6(1), 701.
- Torrijos, P., Alfaro, J. C., Gámez, J. A., & Puerta, J. M. (2024, November). Federated Learning with Discriminative Naive Bayes Classifier. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 328-339). Cham: Springer Nature Switzerland.
- Vibrianti, V., Wahyudin, E., Kaslani, K., Pratama, D., & Dwilestari, G. (2024). Klasifikasi Barang Produksi Pada Tnt. Guitar Workshop Dengan Metode Naive Bayes Menggunakan Rapid Miner. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1432-1438.
- Wardani, N. W., Nugraha, P. G. S. C., & Mahendra, G. S. (2023). Implementasi Naive Bayes Pada Data Mining Untuk Mengklasifikasikan Penjualan Barang Terlaris Pada Perusahaan Ritel. *JST (Jurnal Sains dan Teknologi)*, 12(3), 656-668. <https://doi.org/10.23887/jstundiksha.v12i3.38605>.