

Klasifikasi Tingkat Penjualan Produk pada Toko Jati Karebet Menggunakan Algoritma Naïve Bayes

Revi Setiawan¹, Bayu Priyatna², Elfina Novalia³, Baenil Huda⁴

^{1,2,3,4}Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang

¹si22.revisetiawan@mhs.ubpkarawang.ac.id, ²bayu.priyatna@ubpkarawang.ac.id, ³baenil88@ubpkarawang.ac.id,

⁴elfina.novalia@ubpkarawang.ac.id

Abstract

This study aims to apply the Naive Bayes algorithm to classify product sales levels at Toko Jati Karebet throughout 2024. The background of this research is the suboptimal use of data-driven sales strategies in determining stock and promotion priorities, which often leads to inventory inefficiencies in Micro, Small, and Medium Enterprises (MSMEs). Historical sales data were analyzed using a data mining approach to identify sales patterns and build a predictive model. The initial stage involved data preprocessing, selecting only completed orders, aggregating sales per product, and labeling sales categories into three classes: high-sales (>100 units), medium-sales (20–100 units), and low-sales (<20 units). The Gaussian Naive Bayes model was trained and tested using a supervised learning method with a 70:30 train-test split. Model evaluation was conducted using a confusion matrix and classification metrics. The testing results showed an accuracy of 76%, with a precision of 0.79, recall of 0.98, and F1-score of 0.87 for the high-sales category. These findings indicate that Naive Bayes can provide reliable predictions for the majority class; however, its performance decreases for minority categories due to data imbalance. This research concludes that the Naive Bayes algorithm can serve as a decision-support tool for inventory management and sales strategy in MSMEs, and recommends the application of data balancing techniques or the exploration of other algorithms in future studies to improve performance across all categories

Keywords: product classification, data mining, naive bayes, sales.

Abstrak

Penelitian ini bertujuan menerapkan algoritma Naive Bayes untuk mengklasifikasikan tingkat penjualan produk di Toko Jati Karebet selama tahun 2024. Latar belakang penelitian ini adalah belum optimalnya pemanfaatan strategi penjualan berbasis data dalam menentukan prioritas stok dan promosi, yang sering menyebabkan inefisiensi persediaan pada Usaha Mikro, Kecil, dan Menengah (UMKM). Data penjualan historis dianalisis menggunakan pendekatan data mining untuk mengenali pola penjualan dan membangun model prediksi. Tahap awal meliputi *preprocessing* data, seleksi pesanan yang berstatus selesai, agregasi penjualan per produk, dan pelabelan kategori kelerisan menjadi tiga kelas: laris (>100 unit), kurang laris (20–100 unit), dan tidak laris (<20 unit). Model Gaussian Naive Bayes dilatih dan diuji dengan metode *supervised learning* menggunakan pembagian data 70% untuk pelatihan dan 30% untuk pengujian. Evaluasi model dilakukan dengan *confusion matrix* dan metrik klasifikasi. Hasil pengujian menunjukkan akurasi sebesar 76%, dengan *precision* 0,79, *recall* 0,98, dan *F1-score* 0,87 pada kategori laris. Temuan ini membuktikan bahwa Naive Bayes mampu memberikan hasil prediksi yang cukup andal untuk kategori mayoritas, namun kinerjanya menurun pada kategori minoritas akibat ketidakseimbangan distribusi data. Penelitian ini menyimpulkan bahwa algoritma Naive Bayes dapat digunakan sebagai alat bantu pengambilan keputusan dalam manajemen stok dan strategi penjualan UMKM, serta merekomendasikan penerapan teknik penyeimbangan data atau eksplorasi algoritma lain pada penelitian berikutnya untuk meningkatkan performa di semua kategori.

Kata kunci: klasifikasi produk, data mining, naive bayes, penjualan.

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

1. Pendahuluan

Dalam era digital saat ini, bisnis ritel tidak lagi hanya bergantung pada intuisi atau pengalaman semata dalam menentukan strategi penjualan, tetapi juga pada analisis data penjualan historis yang tersedia. Salah satu metode yang berkembang pesat dalam mengolah data penjualan adalah data mining, yang mampu menggali

informasi tersembunyi dan pola yang berguna untuk pengambilan keputusan bisnis berbasis data [1]. Pengambilan keputusan yang tepat dan cepat merupakan elemen kritis dalam keberlangsungan Usaha Mikro, Kecil, dan Menengah (UMKM). Hal ini menjadi semakin penting mengingat UMKM sering dihadapkan pada berbagai tantangan yang berkaitan

dengan kondisi pasar yang dinamis. Salah satu faktor penentu kesuksesan adalah kemampuan pengelola UMKM dalam merespons pergeseran kebutuhan dan harapan konsumen secara efisien [2].

Toko Jati Karebet, sebagai salah satu pelaku usaha mikro, kecil, dan menengah (UMKM) yang bergerak di bidang penjualan berbagai kelengkapan gamelan, menghadapi tantangan dalam menentukan produk mana yang sebaiknya diprioritaskan dalam stok dan promosi. Sebagai entitas usaha yang menghadapi keterbatasan dalam hal sumber daya dan sistem manajemen berbasis teknologi, toko ini dihadapkan pada tantangan dalam mengelola persediaan barang dan menentukan prioritas produk yang layak untuk dioptimalkan dalam hal penyediaan stok maupun strategi promosi. Ketidadaan sistem yang mampu menganalisis tren penjualan historis menyebabkan terjadinya inefisiensi. Penelitian lain oleh [3] juga menyoroti bahwa keterbatasan akses terhadap teknologi digital dan rendahnya literasi digital menjadi hambatan utama bagi UMKM dalam mengadopsi sistem manajemen persediaan yang lebih efisien.

Dalam rangka mengatasi permasalahan tersebut, penelitian ini diarahkan untuk merancang dan membangun model klasifikasi penjualan produk dengan memanfaatkan algoritma *Naive Bayes* [4], salah satu metode klasifikasi yang populer dalam domain *data mining* [5]. *Naive Bayes* adalah salah satu algoritma pengklasifikasian yang populer di bidang pembelajaran mesin (*machine learning*), terutama karena kesederhanaannya dan efektivitasnya dalam menangani berbagai masalah klasifikasi [6]. Metode ini berdasarkan pada Teorema Bayes dan menggunakan asumsi independensi antar fitur untuk menentukan probabilitas keanggotaan kelas dari data yang diberikan [7].

Tujuan utama dari penelitian ini adalah untuk menghasilkan model prediktif yang mampu mengklasifikasikan produk-produk yang memiliki potensi penjualan tinggi maupun rendah berdasarkan data penjualan historis [8]. Dengan adanya model ini, diharapkan manajemen Toko Jati Karebet dapat memperoleh dasar pertimbangan yang lebih objektif dan sistematis dalam pengambilan keputusan terkait manajemen persediaan dan strategi pemasaran. Selain itu, hasil penelitian ini juga diharapkan dapat berkontribusi pada adopsi teknologi analitik dalam meningkatkan daya saing usaha kecil di era digital.

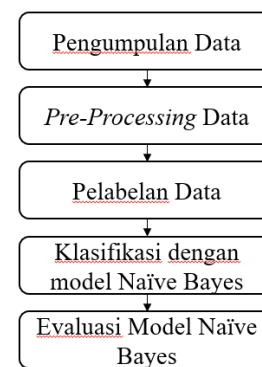
Penelitian ini juga berupaya memberikan gambaran mengenai bagaimana UMKM dapat memanfaatkan pendekatan analitik untuk mengatasi keterbatasan yang mereka miliki, terutama dalam konteks digitalisasi. Dengan adanya dukungan model klasifikasi berbasis *Naive Bayes*, UMKM dapat memprediksi pola penjualan secara lebih akurat, sehingga keputusan bisnis yang diambil tidak hanya mengandalkan perkiraan, tetapi didukung oleh hasil analisis data yang terukur [9]. Pendekatan ini sekaligus menunjukkan

bahwa penerapan algoritma sederhana namun efektif dapat menjadi solusi nyata bagi pelaku usaha kecil yang ingin mengintegrasikan data-driven decision making ke dalam operasional bisnis mereka.

Penerapan metode ini diharapkan dapat membuka peluang penelitian lanjutan yang lebih luas, seperti integrasi dengan algoritma lainnya (misalnya *Decision Tree*, *Random Forest*, atau *Support Vector Machine*) untuk membandingkan tingkat akurasi prediksi. Dengan demikian, penelitian ini tidak hanya memberikan manfaat praktis bagi Toko Jati Karebet, tetapi juga kontribusi ilmiah yang relevan dalam bidang analisis data penjualan UMKM di Indonesia.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode deskriptif analitis [10]. Pendekatan ini dipilih untuk menggambarkan secara menyeluruh pola kelarisan produk berdasarkan data penjualan aktual, kemudian menganalisisnya menggunakan algoritma klasifikasi *Naive Bayes* [11]. Tujuan dari pendekatan ini adalah menghasilkan model klasifikasi yang tidak hanya akurat, tetapi juga dapat digunakan sebagai dasar pengambilan keputusan dalam manajemen stok dan strategi penjualan.



Gambar 1. Metode Penelitian

Berdasarkan Gambar 1, metode penelitian merupakan suatu prosedur sistematis yang digunakan untuk merencanakan, melaksanakan, dan mengevaluasi penelitian. Pada penelitian ini, metode yang digunakan meliputi beberapa tahapan utama, yaitu: (1) *Pengumpulan Data*, dilakukan dengan mengambil arsip penjualan Toko Jati Karebet sepanjang tahun 2024; (2) *Preprocessing Data*, meliputi seleksi pesanan selesai, ekstraksi fitur waktu, agregasi penjualan, dan pembersihan format data; (3) *Pelabelan Data*, mengklasifikasikan produk ke dalam kategori laris, kurang laris, dan tidak laris berdasarkan jumlah unit terjual; (4) *Pembangunan Model Naive Bayes*, menggunakan algoritma Gaussian *Naive Bayes* dengan pembagian data 70% pelatihan dan 30% pengujian; dan (5) *Evaluasi Model*, menggunakan *confusion matrix* serta metrik akurasi, presisi, recall, dan F1-score untuk menilai kinerja model. Dalam metode penelitian terdapat beberapa tahapan yang dapat dijelaskan sebagai berikut:

1. Tahap Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari arsip transaksi penjualan Toko Jati Karebet sepanjang tahun 2024. Seluruh catatan transaksi harian yang tersedia kemudian dikompilasi ke dalam format *Microsoft Excel* (.xlsx) agar lebih mudah dianalisis. Data tersebut memuat sejumlah atribut penting yang menjadi dasar analisis, antara lain nama produk, jumlah terjual, harga awal, harga setelah diskon, total harga produk, total pembayaran, serta waktu pemesanan. Setiap atribut memiliki fungsi yang berbeda, misalnya nama produk digunakan untuk mengidentifikasi jenis barang yang dijual, jumlah terjual untuk menunjukkan kuantitas penjualan, harga awal dan harga setelah diskon untuk menggambarkan strategi promosi, sedangkan total harga produk dan total pembayaran mencerminkan nilai transaksi yang sebenarnya. Adapun atribut waktu pemesanan memberikan informasi mengenai dimensi waktu yang dapat dimanfaatkan untuk melihat pola musiman atau tren penjualan tertentu. Dengan cakupan data transaksi selama satu tahun penuh, penelitian ini memperoleh basis data yang representatif sehingga hasil analisis yang diperoleh dapat mencerminkan kondisi riil penjualan di Toko Jati Karebet.

2. Tahap Pre-Processing Data

Sebelum data dianalisis menggunakan algoritma Naive Bayes, data penjualan yang telah dihimpun perlu melalui proses pra-pemrosesan agar siap diolah. Tahap pertama adalah seleksi data, yaitu memilih hanya transaksi dengan status pesanan “Selesai”, sehingga data yang dianalisis benar-benar valid dan mencerminkan penjualan aktual. Selanjutnya dilakukan ekstraksi fitur waktu, di mana dari atribut waktu pemesanan diturunkan variabel baru berupa bulan dan kuartal. Hal ini bertujuan untuk mempermudah analisis pola penjualan berdasarkan dimensi waktu tertentu. Setelah itu, dilakukan agregasi penjualan dengan cara mengelompokkan data berdasarkan nama produk dan menghitung total kuantitas penjualan untuk setiap produk selama periode penelitian. Tahap berikutnya adalah pembersihan dan konversi data, yang dilakukan dengan memperbaiki format penulisan pada kolom numerik, misalnya menghapus tanda ribuan atau menyesuaikan titik desimal, serta memastikan tipe data sesuai dengan kebutuhan analisis. Proses ini juga melibatkan konversi tipe data agar nilai numerik dapat diperlakukan sebagai angka murni dan bukan teks. Dengan melalui serangkaian tahapan preprocessing ini, data mentah yang semula berpotensi mengandung kesalahan atau inkonsistensi menjadi lebih bersih, terstruktur, dan siap digunakan dalam proses klasifikasi, sehingga dapat meningkatkan akurasi hasil yang diperoleh dari model yang dibangun.

3. Pelabelan Data

Tahap pelabelan data bertujuan untuk menentukan kategori atau kelas dari setiap produk berdasarkan jumlah penjualan yang telah diolah pada tahap

sebelumnya [12]. Proses pelabelan ini sangat penting karena label tersebut akan menjadi target kelas dalam proses klasifikasi menggunakan algoritma Naive Bayes [13]. Langkah-langkah pelabelan dilakukan sebagai berikut:

a. Penentuan Kriteria Klarisan

Kriteria klarisan ditentukan berdasarkan nilai kuantitas penjualan masing-masing produk, penelitian ini, digunakan ambang batas.

- Laris (>100 unit terjual)
- Kurang Laris (20–100 unit)
- Tidak Laris (<20 unit)

b. Implementasi Logika Pelabelan dengan Python.

Setelah data hasil agregasi penjualan diperoleh menggunakan fungsi `groupby()` dari Pandas, pelabelan dilakukan dengan menambahkan kolom baru bernama *Kategori* menggunakan fungsi `apply()` pada Pandas. Fungsi ini menerapkan logika klasifikasi berbasis jumlah penjualan untuk mengelompokkan produk ke dalam tiga kategori: "Tidak Laris", "Kurang Laris", dan "Laris".

c. Pemeriksaan Hasil Pelabelan

Setelah label ditambahkan, dilakukan pengecekan jumlah masing-masing kategori untuk memastikan distribusi data seimbang dan tidak ada kesalahan teknis saat proses pelabelan otomatis.

4. Klasifikasi Model Naive Bayes

Tahap klasifikasi yang dilakukan dalam penelitian ini yaitu dengan menggunakan algoritma *Naive Bayes Classifier* (NBC). Teorema Bayes yang mengasumsikan independensi antar fitur [14]. Rumus umum dari Teorema Bayes adalah:

$$P(C_i|X) = P(X|C_i) \cdot P(C_i)/P(X) \quad (1)$$

Dimana:

$P(C_i|X)$: probabilitas suatu data X termasuk ke dalam kelas C_i

$P(X|C_i)$: probabilitas data X muncul pada kelas C_i

$P(C_i)$: probabilitas a priori dari kelas C_i

$P(X)$: probabilitas a priori dari fitur X

Karena $P(X)$ bernilai sama untuk seluruh kelas, maka dapat disederhanakan menjadi:

$$P(C_i|X) \propto P(X|C_i) \cdot P(C_i) \quad (2)$$

Dalam penelitian ini digunakan varian Gaussian Naive Bayes, yang mengasumsikan bahwa fitur numerik memiliki distribusi normal. Probabilitas dari fitur x_k pada kelas C_i dihitung menggunakan rumus distribusi normal sebagai berikut:

$$P(x_k|C_i) = (1/\sqrt{2\pi\sigma_i^2}) \cdot \exp(-(x_k - \mu_i)^2/(2\sigma_i^2)) \quad (3)$$

Dengan:

μ_i : rata-rata fitur dalam kelas C_i

σ_i^2 : varians fitur dalam kelas C_i

Proses klasifikasi data penjualan Toko Jati Karebet dengan algoritma *Naïve Bayes Classifier* (NBC) dilakukan menggunakan platform *Google Colab*, yang dipilih karena mendukung integrasi dengan bahasa pemrograman Python dan berbagai pustaka machine learning. Langkah pertama yang dilakukan adalah menyiapkan dataset yang telah melalui tahap *pre-processing*. Dataset ini kemudian dibagi menjadi dua bagian, yaitu data training dan data testing, dengan menggunakan teknik *holdout* pada rasio 70:30. Artinya, 70% data digunakan untuk melatih model, sedangkan 30% sisanya dipakai untuk menguji performa model yang telah dibangun.

Selanjutnya, model *Naïve Bayes Classifier* dibangun dengan memanfaatkan data training. Pada tahap ini, algoritma menghitung probabilitas bersyarat dari setiap atribut terhadap kelas target berdasarkan data historis. Setelah model terbentuk, langkah berikutnya adalah menerapkan model klasifikasi pada data training itu sendiri untuk mengevaluasi sejauh mana model mampu mengenali pola data yang sudah dikenalnya. Akurasi hasil klasifikasi pada data training kemudian dihitung dengan menggunakan *confusion matrix*, yang memberikan gambaran mengenai jumlah prediksi benar (*true positive* dan *true negative*) serta prediksi salah (*false positive* dan *false negative*).

Tahapan selanjutnya adalah menguji kemampuan generalisasi model dengan menggunakan data testing. Sebuah model *Naïve Bayes* baru tidak perlu dibangun ulang, melainkan model hasil training diterapkan langsung pada data testing. Melalui proses ini dapat diketahui seberapa baik model dapat mengklasifikasikan data baru yang belum pernah dilihat sebelumnya. Sama seperti pada data training, hasil prediksi pada data testing juga diukur dengan *confusion matrix*, sehingga diperoleh nilai akurasi, *presisi*, *recall*, dan *f1-score*.

Dengan melalui tahapan ini secara berurutan, proses implementasi *Naïve Bayes* tidak hanya memberikan ukuran akurasi pada data training, tetapi juga menilai kemampuan model dalam memprediksi data baru. Evaluasi tersebut menjadi dasar untuk menyimpulkan apakah model yang dibangun sudah cukup andal untuk digunakan sebagai alat bantu pengambilan keputusan dalam klasifikasi kelarisan produk di Toko Jati Karebet.

5. Evaluasi

Evaluasi merupakan hasil pengukuran dengan *confusion matrix*. *Confusion matrix* merupakan suatu tabel yang digunakan untuk mengukur kinerja terhadap suatu algoritma klasifikasi[15]. Diketahui ada beberapa istilah didalam *confusion matrix* diantaranya adalah *True Positif* (TP), *True Negatif* (TN), *False Positif* (FP), serta *False Negatif* (FN). Berdasarkan nilai

confusion matrix tersebut, maka dapat diketahui akurasi klasifikasi secara keseluruhan dengan menggunakan akurasi.

3. Hasil dan Pembahasan

1. Data Awal

Penelitian ini diawali dengan penggunaan data mentah yang diperoleh dari hasil ekspor sistem penjualan Toko Jati Karebet selama periode tahun 2024. Data mentah tersebut masih berupa catatan transaksi harian yang belum melalui proses pembersihan, dengan total sebanyak 2101 baris transaksi yang masing-masing merepresentasikan satu aktivitas pembelian konsumen. Setiap transaksi memuat berbagai atribut penting, antara lain nama produk, yang digunakan untuk mengidentifikasi jenis barang yang dijual; harga produk, baik harga normal maupun harga setelah diskon; jumlah pembelian, yang menunjukkan kuantitas unit produk yang dipesan; status pesanan, yang memberikan informasi apakah transaksi telah selesai, dibatalkan, atau masih dalam proses; serta waktu pemesanan, yang memuat tanggal dan jam terjadinya transaksi. Kondisi awal data ini masih berpotensi mengandung entri yang tidak konsisten, duplikasi, ataupun informasi yang kurang relevan sehingga perlu dilakukan tahapan *pre-processing* untuk memastikan kualitas dan kelayakan data sebelum dianalisis lebih lanjut menggunakan algoritma klasifikasi *Naïve Bayes*.

Tabel 1. Data Penjualan Tahun 2024 (Sebelum *Preprocessing*)

No	Nama Produk	Jumlah	Harga Awal	Harga Setelah Diskon	Total Harga Produk	Total Pembayaran	Waktu Pesanan Dibuat	Bulan	Status Pesanan
1	tabuh/pemukul gamelan (Saron)	10	Rp11.000,00	Rp110.000,00	Rp340.156,00	Rp11.000,00	2024-01-01 11:46:00	1	Selesai
2	tabuh/pemukul saron penerus/peking	10	Rp21.000,00	Rp210.000,00	Rp340.156,00	Rp21.000,00	2024-01-01 11:46:00	1	Selesai
3	tabuh/pemukul gamelan (Saron)	10	Rp11.000,00	Rp110.000,00	Rp115.400,00	Rp11.000,00	2024-01-01 12:03:00	1	Selesai
4	Tabuh Bonang penerus	1	Rp22.795,00	Rp22.795,00	Rp93.409,00	Rp22.795,00	2024-01-01 20:07:00	1	Selesai
5	tabuh/pemukul gamelan (Saron)	1	Rp13.095,00	Rp13.095,00	Rp93.409,00	Rp13.095,00	2024-01-01 20:07:00	1	Selesai
...	...	...	...	...	...	...	...	...	...
2101	senar/kawat Rebab kuning	2	Rp10.000,00	Rp20.000,00	Rp29.600,00	Rp10.000,00	2024-12-31 23:46:00	12	Selesai

2. Hasil Preprocessing

Setelah dilakukan tahap *pre-processing*, data penjualan yang telah bersih kemudian diklasifikasikan berdasarkan total jumlah unit penjualan per produk sepanjang tahun 2024, di mana setiap produk diberi label kategori kelarisan untuk memudahkan proses klasifikasi dengan algoritma *Naïve Bayes*. Penentuan kategori dilakukan melalui analisis distribusi frekuensi penjualan historis yang mencerminkan sebaran jumlah unit terjual, kemudian ditetapkan batasan nilai untuk membagi produk ke dalam tiga kelompok, yaitu laris, kurang laris, serta tidak laris. Batasan ini tidak hanya ditentukan secara statistik, tetapi juga mempertimbangkan karakteristik operasional Toko Jati Karebet sebagai pelaku UMKM yang memiliki keterbatasan sumber daya, fluktuasi permintaan musiman, serta dinamika ketersediaan stok, sehingga klasifikasi yang dilakukan lebih relevan dengan kondisi

riil usaha dan dapat menjadi dasar yang kuat dalam membangun model prediksi kelarisan produk.pertimbangan karakteristik operasional Toko Jati Karebet sebagai pelaku UMKM. sebagai berikut:

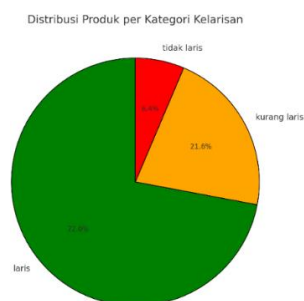
- Laris (>100 unit terjual)
- Kurang Laris (20–100 unit)
- Tidak Laris (<20 unit)

Tabel 2. Data Penjualan Tahun 2024 (Setelah *Preprocessing*)

No	Nama Produk	Jumlah	Harga Awal	Harga Setelah Diskon	Total Harga Produk	Total Pembayaran	Waktu Pesanan Dibuat	Bulan	Kuartal	Label Klarisan	Encoding
1	tabuh/pemukul gamelan (Saron)	10	13.50	0	110.000	340.156	2024-01-01 11:46:00	1	1	laris	0
2	tabuh/pemukul saron penerus/peling	10	22.50	0	210.000	340.156	2024-01-01 11:46:00	1	1	kurang laris	1
3	tabuh/pemukul gamelan (Saron)	10	13.50	0	110.000	115.400	2024-01-01 12:03:00	1	1	laris	0
4	Tabuh Bonang penerus	1	23.50	0	22.795	93.409	2024-01-01 20:07:00	1	1	laris	0
5	tabuh/pemukul gamelan (Saron)	1	13.50	0	13.095	93.409	2024-01-01 20:07:00	1	1	laris	0
...	...	...	...	...	...	...	...	...	...	...	...
21	senar/kawat	2	10.00	0	10.000	29.600	2024-12-31 23:46:00	12	4	laris	0
01	Rebab kuningan										

Pada Tabel 2 memperlihatkan ringkasan data penjualan produk di Toko Jati Karebet pada tahun 2024 setelah melalui tahap *preprocessing*. Data yang disajikan hanya mencakup pesanan dengan status *selesai* sehingga dapat merepresentasikan penjualan yang valid. Setiap baris pada tabel menunjukkan informasi produk berupa nama, jumlah terjual, total harga, harga setelah diskon, total pembayaran, serta kategori kelarisan yang telah ditetapkan berdasarkan jumlah unit terjual. Tabel ini menjadi dasar dalam proses pelatihan model karena seluruh fitur numerik dan label kelas diambil dari data ini setelah pembersihan dan seleksi.Poin penting yang harus ditegaskan:

- File *pre-processing* menunjukkan data penjualan per transaksi (bukan agregat tahunan).
- Model klasifikasi menggunakan data penjualan terakumulasi per produk untuk menentukan kategori kelarisan.



Gambar 2. Diagram Lingkaran – Distribusi Produk

3. Pelatihan Model Naive Bayes

Model Gaussian Naive Bayes digunakan dalam penelitian ini. Data dibagi menjadi 70% untuk pelatihan dan 30% untuk pengujian. Fitur yang digunakan meliputi harga setelah diskon, total harga produk, total pembayaran. Sebelum pelatihan, seluruh fitur numerik distandarisasi menggunakan *StandardScaler* agar distribusi datanya sesuai dengan

asumsi distribusi normal (Gaussian) yang digunakan dalam algoritma ini.

Tabel 3. Distribusi Data *Training* dan Data *Testing*

Kategori	Jumlah Data	Jumlah Data
	Training	Testing
Laris	1.056	457
Kurang Laris	321	132
Tidak Laris	93	41
Total	1.470	630

Selama pelatihan, model menghitung parameter distribusi untuk setiap kelas, berupa nilai rata-rata (*mean*) dan varians (*variance*) dari setiap fitur, yang digunakan untuk membentuk fungsi probabilitas Gaussian. Selain itu, model juga menghitung probabilitas awal (*prior*) dari masing-masing kelas berdasarkan proporsi data pada saat pelatihan.

Tabel 4. Probabilitas awal (*prior*) masing-masing kelas

Kategori	Probabilitas Awal
Laris	71,8%
Kurang Laris	21,8%
Tidak Laris	6,3%

Tabel 5. Nilai rata-rata (*mean*) fitur per kelas (setelah standarisasi)

Kategori	Harga Setelah Diskon	Total Harga Produk	Total Pembayaran
Laris	-0,1735	-0,1282	-0,1193
Kurang Laris	0,3905	0,2737	0,2554
Tidak Laris	0,6221	0,5108	0,4734

Tabel 6. Nilai varians (*Variance*) fitur per kelas (setelah standarisasi)

Kategori	Harga Setelah Diskon	Total Harga Produk	Total Pembayaran
Laris	0,0218	0,2864	0,6631
Kurang Laris	2,1448	1,7585	1,4481
Tidak Laris	6,9010	5,7799	2,6670

Model ini kemudian digunakan untuk memprediksi kategori kelarisan pada data pengujian berdasarkan parameter distribusi yang telah diperoleh selama proses pelatihan.

4. Evaluasi Model Naive Bayes

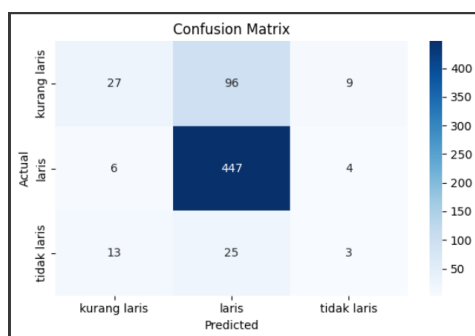
Model dievaluasi dengan metrik klasifikasi sebagai berikut:

Tabel 7. Evaluasi Kinerja Model Naive Bayes

Kelas	Precision	Recall	F1-score	Support
Kurang Laris	0.59	0.20	0.30	132
Laris	0.79	0.98	0.87	457
Tidak Laris	0.19	0.07	0.11	41
Accuracy			0.76	630
Macro Avg	0.52	0.42	0.43	630
Weighted Avg	0.71	0.76	0.70	630

Tabel tersebut menyajikan hasil evaluasi kinerja model *Gaussian Naive Bayes* dalam mengklasifikasikan tingkat kelarisan produk pada data pengujian. Model menunjukkan performa yang sangat baik pada kategori

laris, dengan nilai *precision* sebesar 0,79, *recall* 0,98, dan *F1-score* 0,87. Hal ini menunjukkan bahwa model mampu mengidentifikasi sebagian besar produk yang benar-benar laris dengan tingkat kesalahan prediksi yang rendah. Namun, kinerja model pada kategori *kurang laris* dan *tidak laris* relatif rendah, Performa buruk pada kategori minoritas disebabkan oleh distribusi data yang tidak seimbang, di mana jumlah data pada kelas *laris* jauh mendominasi dibandingkan *kurang laris* dan *tidak laris*. Hal ini mengakibatkan model lebih terlatih pada pola produk laris, sehingga cenderung memprediksi sebagian besar data ke kelas tersebut. Selain itu, distribusi fitur pada kelas minoritas yang sangat terbatas menyebabkan parameter probabilitas yang dihitung pada model kurang representatif, sehingga mengurangi akurasi prediksi pada kelas tersebut.



Gambar 3. *Confusion Matrix*

Berdasarkan hasil *confusion matrix*, model Gaussian Naive Bayes menunjukkan kinerja yang sangat baik dalam memprediksi kategori *laris*. Dari total 457 data aktual pada kategori ini, sebanyak 447 data berhasil diprediksi dengan benar, menghasilkan nilai *recall* sebesar 0,98 dan *precision* 0,79. Namun, performa model menurun pada kategori *kurang laris* dan *tidak laris*. Terdapat 96 data *kurang laris* dan 25 data *tidak laris* yang salah diklasifikasikan sebagai *laris*. Prediksi benar untuk kategori *kurang laris* hanya sebanyak 27 data dari total 132 data (*recall* 0,20), sedangkan pada kategori *tidak laris* hanya 3 data dari total 41 data (*recall* 0,07). Kondisi ini menunjukkan adanya bias model terhadap kelas mayoritas (*laris*), yang disebabkan oleh ketidakseimbangan distribusi data pelatihan. Meskipun demikian, secara keseluruhan model menghasilkan akurasi sebesar 76% dan *F1-score* 0,87 pada kategori *laris*, sehingga dapat dijadikan dasar pengambilan keputusan untuk pengelolaan stok produk dengan penjualan tinggi, meskipun perlu peningkatan akurasi pada kategori penjualan menengah dan rendah melalui teknik penyeimbangan data atau eksplorasi algoritma lain.

Hasil ini mengindikasikan bahwa model cenderung bias terhadap kelas mayoritas (*laris*), sehingga prediksi pada kelas minoritas belum optimal. Pola ini sesuai dengan hasil metrik evaluasi yang menunjukkan

kinerja lebih baik pada kategori *laris* dibandingkan kategori lainnya.

4. Kesimpulan

Penelitian ini berhasil menerapkan algoritma Gaussian Naïve Bayes untuk mengklasifikasikan tingkat kelarisan produk di Toko Jati Karebet berdasarkan data penjualan selama tahun 2024. Pemilihan Gaussian Naïve Bayes didasarkan pada sifat data penjualan yang sebagian besar bersifat numerik dan terdistribusi mendekati normal, sehingga sesuai dengan asumsi probabilistik algoritma ini. Model yang dibangun menghasilkan tingkat akurasi keseluruhan sebesar 76%, yang menunjukkan bahwa lebih dari tiga perempat data berhasil diklasifikasikan dengan benar ke dalam kategori laris, kurang laris, dan tidak laris. Evaluasi lebih lanjut menggunakan metrik *precision*, *recall*, dan *F1-score* memperlihatkan bahwa performa terbaik dicapai pada kategori *laris*, dengan nilai *precision* 0,79, *recall* 0,98, serta *F1-score* 0,87. Nilai *precision* yang tinggi berarti sebagian besar produk yang diprediksi laris memang benar-benar laris, sedangkan nilai *recall* yang mendekati 1 menunjukkan bahwa hampir semua produk laris berhasil dikenali oleh model. *F1-score* yang seimbang memperkuat bukti bahwa Gaussian Naïve Bayes sangat efektif dalam mendeteksi produk dengan tingkat penjualan tinggi. Implikasi praktis dari hasil ini adalah manajemen toko dapat menggunakan model ini sebagai acuan dalam menentukan produk prioritas untuk disediakan stok lebih banyak, merancang strategi promosi yang lebih terarah, serta mengurangi risiko kehabisan barang untuk produk dengan tingkat permintaan tinggi.

Meskipun hasil yang diperoleh cukup memuaskan, kinerja model pada kategori *kurang laris* dan *tidak laris* masih rendah. Ketidakseimbangan jumlah data antar kategori, di mana produk laris jauh lebih banyak dibandingkan produk dengan penjualan rendah, menyebabkan model cenderung bias terhadap kelas mayoritas. Kondisi ini berimplikasi pada rendahnya nilai *precision* dan *recall* untuk kategori minoritas, sehingga beberapa produk yang sebenarnya kurang laris atau tidak laris salah diklasifikasikan sebagai laris. Masalah *class imbalance* ini menjadi tantangan umum dalam implementasi *machine learning* pada data dunia nyata, terutama pada konteks UMKM yang cenderung memiliki produk unggulan dominan dan produk sampingan dengan frekuensi penjualan rendah. Untuk mengatasi hal tersebut, penelitian selanjutnya disarankan menggunakan teknik penyeimbangan data seperti SMOTE (*Synthetic Minority Oversampling Technique*) untuk memperbanyak data minoritas secara sintesis atau *undersampling* untuk mengurangi dominasi data mayoritas. Selain itu, penggunaan algoritma lain yang lebih canggih seperti *Random Forest* atau *Support Vector Machine* (SVM) dapat dipertimbangkan untuk meningkatkan akurasi pada semua kategori. Dengan langkah-langkah ini, hasil penelitian diharapkan tidak hanya bermanfaat bagi

Toko Jati Karebet, tetapi juga dapat menjadi rujukan strategis bagi UMKM lain dalam mengadopsi teknologi analitik untuk mendukung pengambilan keputusan berbasis data.

Daftar Rujukan

- [1] Hamsiah, "Implementasi Data Mining Dalam Penerapan Clustering Algoritma K-Medoid Sebaran Mahasiwa Baru Pada Stie-Sak," Jan. 2025.
- [2] D. Christine, Apriwandii, A. N. Fathonah, E. Sherlita, A. Wijaya, And E. Kartadjumen, "Analisis Penganggaran Modal Pada Usaha Mikro Kecil Dan Menengah (Ukm) Di Kabupaten Bandung Barat," *Jurnal Emt Kita*, Vol. 7, No. 7(1), Pp. 179–190, Jan. 2023.
- [3] A. Handojo, M. Christian Wibisono, A. Noertjahyana, And T. Octavia, "Inventory Control Application On Indonesia Small Medium Enterprises Using Smartphone," *Kne Life Sciences*, Pp. 1–13, Mar. 2020.
- [4] Repan, B. Ceasar Octariadi, And Sucipto, "Penerapan Algoritma (Naïve Bayes) Untuk Memprediksi Penyakit Diare," No. 15(1), Pp. 49–56, Apr. 2025.
- [5] J. Han, M. Kamber, And J. Pei, *Data Mining: Concepts And Techniques*, 3rd Ed. Elsevier, 2011.
- [6] Tukino, A. Hananto, R. A. Nanda, E. Novalia, E. Sedyono, And J. Sanjaya, "Lstm And Word Embedding: Classification And Prediction Of Puskesmas Reviews Via Twitter," In *E3s Web Of Conferences*, Edp Sciences, Mar. 2024.
- [7] Suwati, Y. Rolly, And S. Andy, "Prediksi Kelancaran Pembayaran Angsuran Pada Koperasi Dengan Metode Naive Bayes Classifier," No. 11(2), Pp. 635–644, Aug. 2022.
- [8] A. L. Hananto, A. Hananto, And B. Huda, "Analisis Dan Pemodelan Proses Bisnis Katering Pada Ukm Menggunakan Bpmn," *Internal (Information System Journal)*, Vol. 7, No. 7(1), Pp. 8–17, 2024, [Online]. Available: [Http://Jurnal.Masoemuniveristy.Ac.Id/Index.Php/Internal](http://Jurnal.Masoemuniveristy.Ac.Id/Index.Php/Internal)
- [9] D. Agustina Cahyaningrum, D. Arviana Wulan Prastika, And F. Sains Dan Teknologi, "Rancangan Sistem Informasi Pemantauan Stok Barang Dan Penentuan Manajemen Akuntansi Penggajian Karyawan Pada Ukm Bakpia," *Jurnal Teknologi Dan Manajemen Industri Terapan (Jtmit)*, Vol. 2, No. 3, Pp. 182–190, 2023.
- [10] I. Hassandi And Effiyaldi, "Analisis Pengelolaan Persediaan Bahan Baku Pada Ukm Tabib Jamu Untuk Peningkatan Efisiensi," No. 4(1), Jan. 2025.
- [11] M. Thoriq, F. Maulan, Y. Septi Eirlangga, N. Hayati, M. Ashim Madani, And F. Maulana, "Implementasi Algoritma Naïve Bayes Dalam Prediksi Penerimaan Mahasiswa Penerima Beasiswa Kip Di Universitas Adzkia," No. 15(1), Pp. 108–114, Apr. 2025.
- [12] A. L. Hananto, A. Y. Rahman, T. Paryono, B. Priyatna, A. Hananto, And B. Huda, "Classification Of Starling Images Using A Bayesian Network," *Journal Of Applied Data Sciences*, Vol. 6, No. 6(1), Pp. 34–46, 2025, [Online]. Available: <https://doi.org/10.47738/jads.v5i4.423>
- [13] Nadia Tiara Rahman, "Analisa Algoritma Decision Tree Dan Naïve Bayes Pada Pasien Penyakit Liver," No. 10(2), Pp. 144–151, Jul. 2020.
- [14] Hidayatunnisa, Kusri, And Kusnawi, "Perbandingan Kinerja Metode Naive Bayes Dan Support Vector Machine Dalam Analisis Soal," *Agustus*, Vol. 13, No. 2, Pp. 173–180, 2023.
- [15] M. Jannah, M. Arief, H. M. Kom, M. Al Fajar, And M. A. Hasan, "Perbandingan Metode Naive Bayes Dan K-Nearest Neighbor Dalam Mengklasifikasi Status Pertumbuhan Anak Stunting (Studi Kasus : Posyandu Cemara)," No. 14(1), Pp. 250–225, 2024.