

## Classification of Best-Selling Products Based at Noenaasstore Using Naïve Bayes Algorithm

Nurul Hidayah<sup>1</sup>, Paminto Agung Chrishtianto<sup>2</sup>, Nurul Amalia<sup>3\*</sup>

<sup>1</sup> Sistem Informasi, Fakultas Teknologi Informasi, Institut Widya Pratama Pekalongan, Kota Pekalongan, 51146, Indonesia

<sup>2,3</sup> Sistem Informasi, Fakultas Teknologi Informasi, Institut Widya Pratama Pekalongan, Kota Pekalongan, 51146, Indonesia

### Informasi Artikel

Diterima : 8 September 2025  
Revisi : 09 Januari 2026  
Publikasi : 31 Maret 2026

### Kata Kunci:

Naïve Bayes  
CRIPS-DM  
Confusion Matrix  
RapidMiner

### ABSTRAK

Kemajuan teknologi informasi telah mengubah perilaku belanja masyarakat ke arah penggunaan platform *e-commerce*, dengan kategori fashion menempati peringkat teratas. Situasi ini menuntut UMKM untuk mengenali produk dengan penjualan tertinggi agar strategi pemasaran dan keputusan bisnis dapat dilakukan secara lebih tepat. Penelitian ini menerapkan algoritma Naïve Bayes pada data transaksi penjualan Noenaasstore, salah satu UMKM yang bergerak di bidang fashion wanita, untuk melakukan klasifikasi produk berdasarkan tingkat penjualannya. Evaluasi model menggunakan RapidMiner menunjukkan akurasi sebesar 92,62%, *weighted mean precision* 83,81%, serta *weighted mean recall* 95,51%. Temuan ini mengindikasikan bahwa algoritma Naïve Bayes mampu mengelompokkan produk secara efektif, sehingga dapat dimanfaatkan pemilik usaha untuk merumuskan strategi promosi serta mendukung pengambilan keputusan bisnis.

### ABSTRACT

The advancement of information technology has shifted consumer shopping behavior toward e-commerce platforms, with the fashion category occupying the top position. This situation requires MSMEs to identify their best-selling products in order to design more accurate marketing strategies and business decisions. This study applies the Naïve Bayes algorithm to sales transaction data from Noenaasstore, an MSME engaged in women's fashion, to classify products based on their sales levels. Model evaluation using RapidMiner achieved an accuracy of 92,62%, a weighted mean precision of 83,81%, and a weighted mean recall of 95,51%. These findings indicate that the Naïve Bayes algorithm can effectively categorize products, thereby enabling business owners to formulate promotional strategies and support data-driven decision-making.

This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



### \*Penulis Koresponden

Email: [nurul196b@gmail.com](mailto:nurul196b@gmail.com)

Cara sitasi IEEE:

N. Hidayah, P. A Chrishtianto, & N. Amalia, "Classification of Best-Selling Products Based at Noenaasstore Using Naïve Bayes Algorithm," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 6, no. 1, pp. 08-17, Maret 2026. doi: 10.30811/jaise.v6i1.7701

## 1. PENDAHULUAN

Kemajuan teknologi informasi telah membawa pengaruh besar dalam kehidupan sehari-hari, terutama pada penggunaan *e-commerce* yang mengubah cara masyarakat berbelanja, berbisnis, dan berinteraksi di dunia digital [1]. Menurut data yang bersumber dari Jakpat tahun 2022 bahwa fashion menjadi kategori produk dengan pembelian tertinggi melalui *e-commerce* di Indonesia mencapai 58%. Dominasi ini tidak hanya menunjukkan tingginya minat konsumen, tetapi juga erat kaitannya dengan tren fashion yang terus berkembang dan cepat berganti, sehingga mendorong konsumen untuk lebih aktif berbelanja secara online. Dibandingkan kategori lain seperti gadget dan elektronik (47%), kecantikan (43%), serta bahan makanan (33%), fashion menempati posisi teratas [2].

Fashion seringkali memiliki keterkaitan yang kuat dengan tren. Gaya hidup dan kebutuhan yang terus beradaptasi dengan tren fashion turut mendorong pergeseran perilaku belanja masyarakat dari toko konvensional ke platform digital. Dalam industri fashion, tren selalu berubah secara dinamis, dipengaruhi oleh berbagai elemen seperti situasi ekonomi, faktor sosial, serta perkembangan zaman [3].

Dengan adanya tren yang berkembang pesat, klasifikasi produk menjadi sangat penting bagi perusahaan dan pengusaha. Klasifikasi produk dapat membantu menentukan strategi pemasaran yang tepat, mempermudah proses produksi dan distribusi. Namun dalam praktiknya, banyak pengusaha terutama pelaku UMKM masih menghadapi kendala dalam mengklasifikasikan produk mereka. Kendala ini disebabkan oleh kurangnya pemahaman tentang konsep klasifikasi, produk yang tidak terkelompok dengan jelas, kesalahan dalam menentukan target pasar, keterbatasan sumber daya, serta minimnya data dan analisis terkait kebutuhan konsumen [4]. Kondisi ini juga dialami oleh Noenaasstore, salah satu UMKM yang bergerak di bidang fashion wanita dengan menjual produk seperti baju dan celana. Meskipun Noenaasstore memiliki data riwayat penjualan yang cukup, pemanfaatannya untuk analisis mendalam masih belum optimal karena metode yang digunakan cenderung tradisional dan belum mengadopsi pendekatan berbasis data yang lebih modern.

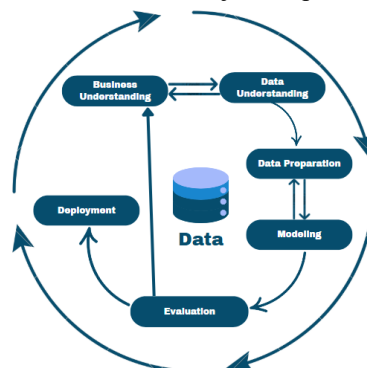
Berdasarkan tinjauan literatur yang ada, Beberapa penelitian sebelumnya telah mengkaji klasifikasi produk terlaris menggunakan algoritma Naive Bayes seperti pada penelitian yang menggunakan data penjualan pulsa di konter RA Cell. Hasilnya menunjukkan tingkat akurasi yang sangat tinggi, yakni 97,50%, dengan nilai *precision* mencapai 100% dan *recall* 93,48%, sehingga dapat mengidentifikasi bahwa pulsa Telkomsel merupakan produk paling laris [5]. Selain itu, penelitian yang mengimplementasikan algoritma Naive Bayes pada data penjualan CV Akusara Jaya Abadi. Dari hasil pengujian, diperoleh tingkat akurasi sebesar 89,33% dengan nilai *precision* 88,73% dan *recall* 90,12%, yang menunjukkan bahwa metode ini cukup efektif untuk mengklasifikasikan produk terlaris berdasarkan riwayat transaksi [6].

Kedua penelitian tersebut membuktikan algoritma Naive Bayes dapat diimplementasikan secara optimal dalam mengklasifikasikan produk terlaris. Namun, fokus penelitian terdahulu lebih banyak diarahkan pada sektor ritel umum, seperti pulsa, maupun perusahaan skala besar, bukan pada toko fashion lokal. Padahal, sektor fashion memiliki karakteristik unik yang sangat dipengaruhi oleh tren, musim, dan gaya hidup, sehingga produk yang populer dapat berubah dengan cepat.

Pada penelitian ini menghadirkan pendekatan yang berbeda dengan mengimplementasikan algoritma Naive Bayes pada data transaksi penjualan fashion wanita di Noenaasstore. Sehingga diharapkan dapat membantu pemilik usaha dalam pengambilan keputusan bisnis yang lebih efektif.

## 2. METODE

Metode yang diterapkan dalam penelitian ini yaitu menggunakan kerangka kerja data mining *Cross-Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM merupakan salah satu metode yang banyak digunakan dalam pendekatan *data mining*. Metode ini tidak hanya digunakan pada aspek teknis, tetapi juga berfungsi sebagai metodologi proyek dalam konteks formal, termasuk penelitian, karena mencakup seluruh proses yang diperlukan dalam *data mining* [7]. Fase-fase utama dalam CRISP-DM ditunjukkan pada Gambar 1.



Gambar 1. Tahapan CRISP-DM

## 2.1 Business Understanding

*Business understanding* merupakan proses untuk memperoleh pemahaman yang mendalam mengenai permasalahan yang akan diselesaikan, dampaknya terhadap organisasi, serta penetapan tujuan yang ingin dicapai dalam mengatasinya. Tahap ini merupakan penentuan tujuan proyek serta pemetaan kebutuhan bisnis maupun penelitian secara komprehensif. Tahapan ini juga dirumuskan strategi awal yang akan digunakan dalam mencapai tujuan penelitian.

## 2.2 Data Understanding

Tahap kedua dari proses *data mining* dilakukan setelah penetapan tujuan dan perumusan rencana. Ada fase ini dilakukan serangkaian aktivitas yang mencakup pengumpulan data, pendeskripsian karakteristik data, eksplorasi untuk mengidentifikasi pola awal, serta verifikasi terhadap kualitas data. Apabila ditemukan kendala atau ketidaksesuaian, maka rencana maupun pemahaman bisnis yang telah dirumuskan sebelumnya perlu dilakukan peninjauan ulang agar tetap selaras dengan kebutuhan analisis penelitian.

## 2.3 Data Preparation

Tahap ini merupakan persiapan dalam membangun dataset akhir yang akan digunakan pada fase berikutnya (*modeling*). Proses ini mencakup pembersihan data guna mengatasi duplikasi dan kekosongan data, integrasi data dari berbagai sumber, serta pemilihan data yang relevan. Selain itu, dilakukan *labeling* atau pengkategorian data berdasarkan nilai *mean* dan standar deviasi, yang dihitung dari harga produk dan jumlah total penjualan.

$$\bar{X} = \frac{\sum X}{n} \quad (1)$$

Rumus (1) merupakan rumus *Mean* ( $\bar{X}$ ) yang diperoleh dari perhitungan harga produk dan total penjualan setiap produk, sehingga dapat memberikan gambaran mengenai rata-rata harga yang ditawarkan oleh Noenaasstore sekaligus rata-rata jumlah produk yang berhasil terjual.  $\sum X$  merupakan total penjualan produk dan  $n$  merupakan jumlah produk.

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}} \quad (2)$$

Rumus (2) merupakan rumus Standar deviasi ( $\sigma$ ) yang digunakan untuk mengetahui tingkat penyebaran data terhadap nilai rata-rata ( $\bar{X}$ ). Perhitungannya dilakukan dengan melihat selisih antara total terjual per produk ( $x_i$ ) dengan rata-rata ( $\bar{X}$ ), kemudian hasil selisih kuadrat tersebut dijumlahkan dan dibagi dengan jumlah produk ( $n$ ) atau jumlah kategori produk.

## 2.4 Modeling

Tahap ini berfokus pada pemilihan serta penerapan teknik pemodelan yang sesuai dengan metode yang digunakan, yaitu algoritma Naive Bayes untuk melakukan klasifikasi produk terlaris. Pada tahap ini, peneliti menyesuaikan parameter serta melakukan proses optimasi guna memperoleh hasil perhitungan yang lebih optimal.

## 2.5 Evaluation

Tahapan ini digunakan untuk menilai kualitas serta efektivitas model yang telah dibangun. Evaluasi dilakukan dengan mempertimbangkan aspek penting dari permasalahan bisnis sehingga peneliti dapat menentukan apakah hasil prediksi dengan algoritma Naive Bayes layak digunakan. Dalam penelitian ini, pengukuran evaluasi dilakukan menggunakan *confusion matrix* 3x3 yang umum dipakai dalam konteks klasifikasi maupun prediksi pada *data mining* dengan lebih dari dua kategori.

## 2.6 Deployment

Tahap akhir berfokus pada penggunaan hasil prediksi produk terlaris sebagai acuan dalam perumusan strategi pemasaran. Model yang dihasilkan tidak sekadar ditampilkan dalam bentuk laporan, tetapi juga dimanfaatkan sebagai landasan dalam menyusun strategi promosi dan pendekatan pemasaran yang lebih optimal guna meningkatkan daya saing serta volume penjualan.

Penelitian ini menggunakan algoritma Naive Bayes Multinomial, karena sesuai untuk data kategorikal dan klasifikasi tingkat penjualan produk. Evaluasi model dilakukan dengan menggunakan *confusion matrix* 3x3, sehingga kinerja model dapat dinilai secara menyeluruh melalui pengukuran akurasi, weighted mean precision, dan weighted mean recall. Seluruh rangkaian analisis, mulai dari pra-pemrosesan hingga evaluasi diolah menggunakan perangkat lunak RapidMiner yang mendukung proses data mining secara terintegrasi.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Business Understanding

Tahapan awal dalam kerangka kerja CRISP-DM adalah memahami kebutuhan bisnis yang melatarbelakangi penelitian. Tujuan dari penelitian ini, untuk mengklasifikasikan produk pada toko Noenaasstore berdasarkan tingkat

kelarisannya ke dalam tiga kategori, yaitu laris, cukup laris, dan tidak laris. Klasifikasi ini diharapkan dapat memberikan solusi bagi pemilik usaha dalam menentukan strategi pemasaran yang lebih tepat sasaran dan pengambilan keputusan bisnis yang lebih optimal.

### 3.2 Data Understanding

Tahapan ini merupakan tahapan pemahaman data. Data yang digunakan dalam penelitian ini merupakan data primer yang bersumber dari transaksi penjualan Noenaasstore.

Gambar 2. Transaksi Penjualan Bulan Januari Sampai Juli 2025

Pada gambar 2 merupakan transaksi penjualan Noenaasstore pada periode bulan Januari sampai Juli 2025. Secara keseluruhan, dataset awal berjumlah 620 entri dan memuat sejumlah atribut, di antaranya *order ID, order Status, SKU ID, seller SKU, product name, variation, quantity, price, SKU quantity of return, created time, paid time, shipping provider name, buyer user, phone, address, payment method*. Atribut-atribut tersebut merupakan rincian dari transaksi yang terjadi di toko, namun tidak seluruh atribut memiliki keterkaitan dengan tujuan penelitian, sehingga perlu dilakukan proses seleksi untuk menentukan data yang relevan dalam pembangunan model klasifikasi.

### 3.3 Data Preparation

#### 3.3.1 Cleaning Data

Tahapan ini merupakan tahapan menyeleksi data yang bermasalah, seperti adanya duplikasi maupun kekosongan data (*missing values*).

Gambar 3. Proses Pembersihan Data / Cleaning Data

Pada gambar 3 data yang bermasalah akan dihapus dan tidak dimasukkan dalam analisis. Upaya ini dilakukan untuk memastikan kualitas data tetap terjaga serta meminimalkan potensi bias dalam proses pemodelan. Pada penelitian ini data yang semula terdiri dari 620 transaksi dibersihkan menjadi 612 data transaksi yang nantinya akan digunakan untuk analisa model Naïve Bayes.

#### 3.3.2 Integration Data

Pada tahap ini integrasi data dilakukan dengan tujuan untuk menyederhanakan proses analisis.

Gambar 4. Proses Integrasi Data

Gambar 4 merupakan integrasi data pada atribut *created time* yang semula berformat (tanggal-bulan-tahun) diubah menjadi (bulan) dengan menggunakan rumus excel.

### 3.3.3 Selection Data

Dalam proses seleksi data, hanya atribut yang memiliki relevansi dengan tujuan penelitian yang dipilih untuk digunakan. Atribut yang dipertahankan meliputi *order created time*, *product name*, *variation*, *quantity*, *Price*. Kelimanya atribut tersebut berperan dalam proses klasifikasi tingkat kelarisan produk, sedangkan atribut lain yang dinilai tidak memiliki pengaruh signifikan terhadap analisis disisihkan. Dengan langkah ini, data yang digunakan menjadi lebih terarah dan selaras dengan fokus penelitian.

Tabel 1. Data Atribut

No	Nama Atribut	Keterangan
1.	Created Time	Tanggal Pemesanan
2.	Product Name	Nama Produk
3.	Variation	Variasi Produk
4.	Quantity	Jumlah Pesanan Setiap Transaksi
5.	Price	Harga per Produk

### 3.3.4 Labeling

Tahap labeling pada penelitian ini digunakan untuk mengelompokkan produk ke dalam tiga kelas tingkat kelarisan, yaitu laris, cukup laris, dan kurang laris.

Kriteria pengelompokan ditentukan melalui perhitungan nilai rata-rata ( $\bar{X}$ ) dan standar deviasi ( $\sigma$ ). Penentuan kategori ini didasarkan pada total penjualan dan harga produk, sehingga hasil klasifikasi lebih proporsional. Data transaksi untuk perhitungan *mean* dan standar deviasi terdapat pada tabel 2.

Tabel 2. Data Transaksi Penjualan

Nama Produk	Total Transaksi	Harga	Total Penjualan
Rok Susun Jeans	25	112000	2800000
Celana Kulot	27	100000	2700000
Scuba Flare	104	58000	6032000
Vest Salur	29	55000	1595000
Oro Pants	31	55000	1705000
Cutbray Jeans	35	95000	3325000
Blouse Bordir	98	87000	8526000
Obi Pants	28	60000	1680000
Blouse Rempel	113	85000	9605000
Kemeja Wispie	122	55890	6818580
Total Penjualan Keseluruhan			44786580

$$\mu = \frac{\sum X}{n} = \frac{44786580}{10} = 4.478.658$$

$$\sigma = \sqrt{\frac{\sum_i^n (x_i - \mu)^2}{n}} = \sqrt{\frac{81.698.566.361.316}{10}} = 2.853.357$$

Kategori :

Laris =  $4.478.658 + 2.853.357 = \geq 7.332.015$

Kurang Laris =  $4.478.658 - 2.853.357 = x < 1.625.301$

Cukup Laris =  $1.625.301 \leq x < 7.332.015$

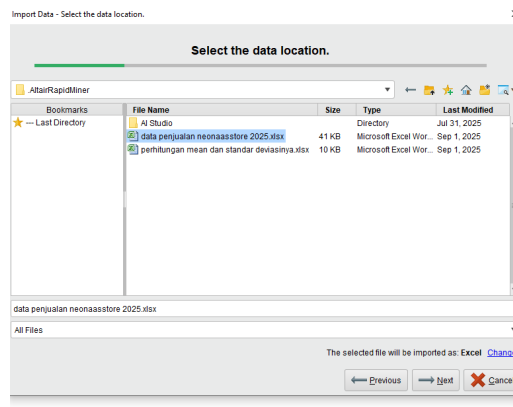
Dari perhitungan *Mean* dan Standar Deviasi yang telah dilakukan diperoleh nilai mean sebesar 4.478.658 dan standar deviasi sebesar 2.853.357 sehingga batas bawah ditetapkan pada 1.625.301 dan batas atas pada 7.332.015. Sehingga dapat dibuat kategori kelarisan sesuai dengan gambar 5.

B	C	D	E	F	G
Created time	Product Name	Variation	Quantity	total After	kategori
February	NEONA ATASAN KEMEJA WISPIE/LUNA STRIPE/KEMEJA SLIM FIT SALU	Ld 100, mahogany	1	55890	Cukup Laris
April	NEONA JEANS CUTBRAY	29, green retro	1	95000	Cukup Laris
January	CELANA PANJANG WANITA OBI PANTS	standar, coklat mahog	1	60000	Cukup Laris
July	NEONA JEANS CUTBRAY	33, navy	1	95000	Cukup Laris
March	CELANA PANJANG WANITA OBI PANTS	jumbo, coklat mahoga	1	60000	Cukup Laris
February	NEONA JEANS CUTBRAY	29, green retro	1	95000	Cukup Laris
January	NEONA JEANS CUTBRAY	27, hitam	1	95000	Cukup Laris
January	CELANA PANJANG WANITA OBI PANTS	standar, coklat mahog	1	60000	Cukup Laris
May	NEONA JEANS CUTBRAY	27, hitam	1	95000	Cukup Laris
May	NEONA ATASAN KEMEJA WISPIE/LUNA STRIPE/KEMEJA SLIM FIT SALU	navy, S/M	2	55890	Cukup Laris
June	CELANA PANJANG WANITA OBI PANTS	standar, coklat mahog	1	60000	Cukup Laris
March	NEONA ATASAN KEMEJA WISPIE/LUNA STRIPE/KEMEJA SLIM FIT SALU	hitam, L/XL	1	55890	Cukup Laris
July	NEONA JEANS CUTBRAY	27, green retro	1	95000	Cukup Laris
March	CELANA PANJANG WANITA OBI PANTS	standar, coklat mahog	1	60000	Cukup Laris
February	NEONA ATASAN KEMEJA WISPIE/LUNA STRIPE/KEMEJA SLIM FIT SALU	grey, L/XL	1	55890	Cukup Laris
March	CELANA PANJANG WANITA OBI PANTS	standar, coklat mahog	1	60000	Cukup Laris
March	CELANA PANJANG WANITA OBI PANTS	standar, hitam	1	60000	Cukup Laris
March	NEONA ATASAN KEMEJA WISPIE/LUNA STRIPE/KEMEJA SLIM FIT SALU	grey, L/XL	1	55890	Cukup Laris
May	NEONA JEANS CUTBRAY	27, green retro	1	95000	Cukup Laris
June	NEONA JEANS CUTBRAY	27, navy	1	95000	Cukup Laris
June	CELANA PANJANG WANITA OBI PANTS	standar, cream	1	60000	Cukup Laris
April	NEONA ATASAN KEMEJA WISPIE/LUNA STRIPE/KEMEJA SLIM FIT SALU	coklat mahogany, L/X	1	55890	Cukup Laris

Gambar 5. Pemberian Label Kelas

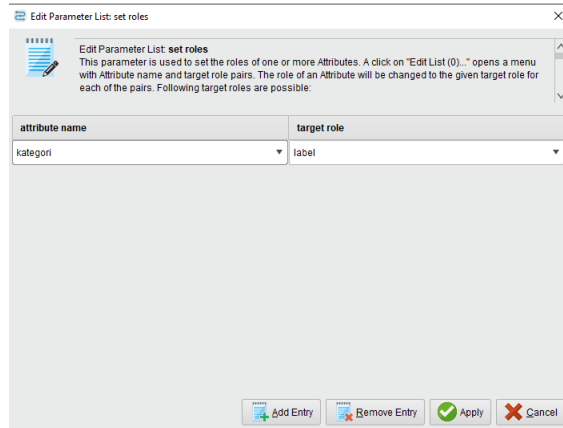
### 3.4 Modeling

Pada tahap ini, algoritma Naïve Bayes yang digunakan merupakan Naïve Bayes Multinomial untuk mengklasifikasikan produk ke dalam tiga kategori, yaitu laris, cukup laris, dan kurang laris.



Gambar 6. Proses *Import* Data ke RapidMiner

Gambar 6 merupakan proses *importing* data penjualan yang telah melalui tahapan pra-pemrosesan, kemudian diimport ke dalam RapidMiner 2025.1 untuk dilakukan pengujian model. Kemudian *drag and drop* operator *retrieve* untuk membaca data excel yang sudah diimport tadi.



Gambar 7. Parameter List Set Role

Pada Gambar 7 pengujian dilakukan dengan *drag and drop* operator *Set Role* dengan atribut kategori dan target role label. Selanjutnya, data dibagi menggunakan operator *Split Data* dengan proporsi 0.8 sebagai data latih dan 0.2 sebagai data uji. *Data training* digunakan untuk membangun model menggunakan algoritma Naïve Bayes, sedangkan *data testing* diproses melalui operator *Apply Model* untuk menghasilkan prediksi kategori produk.

Result History ExampleSet (Set Role data training)

Row No.	kategori	Created time	Product Na...	Variation	Quantity	SKU Subota...
1	Cukup Laris	June	Celana Kulot ...	RANDOM 10...	3	100000
2	Cukup Laris	May	NEONA ORO ...	standar, cream	1	55000
3	Cukup Laris	April	NEONA JEAN...	29, green reto	1	95000
4	Cukup Laris	Oktober	NEONA ORO ...	ivory, standar...	1	55000
5	Cukup Laris	July	NEONA JEAN...	33, navy	1	95000
6	Cukup Laris	March	CELANA PAN...	jumbo, coklat...	1	60000
7	Cukup Laris	August	NEONA JEAN...	29, green reto	1	95000
8	Cukup Laris	September	Celana Kulot ...	RANDOM 10...	3	100000
9	Cukup Laris	February	ELI CUTBRA...	xl to jumbo, hi...	1	58000
10	Cukup Laris	January	NEONA JEAN...	27, hitam	1	95000
11	Cukup Laris	December	CELANA PAN...	standar, cokl...	1	60000
12	Cukup Laris	July	ELI CUTBRA...	standar, hitam	1	58000
13	Cukup Laris	June	Celana Kulot ...	RANDOM 10...	3	100000
14	Cukup Laris	May	NEONA JEAN...	27, hitam	1	95000

ExampleSet (490 examples, 1 special attribute, 5 regular attributes)

(a)

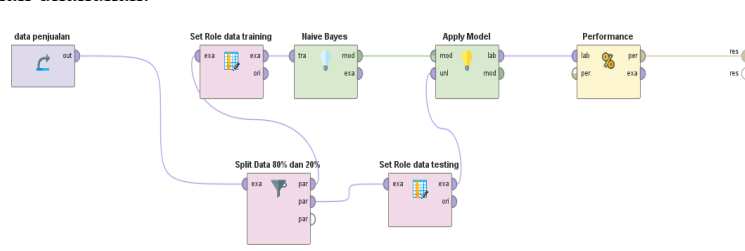
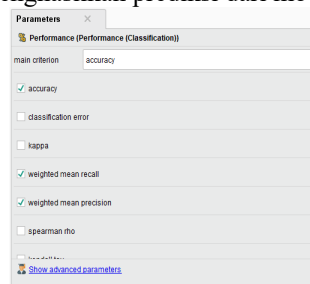
Row No.	kategori	prediction(k...	confidence(...	confidence(...	confidence(...	Created time	Product Na...	Variation	Qua...
1	Cukup Laris	Cukup Laris	1.000	0.000	0	January	ELI CUTBRA...	xl to jumbo, a...	1
2	Cukup Laris	Cukup Laris	1	0	0	September	Celana Kulot ...	RANDOM 10...	3
3	Cukup Laris	Cukup Laris	1.000	0.000	0	February	NEONA ATAS...	Ld 100, mah...	1
4	Cukup Laris	Cukup Laris	1.000	0.000	0	January	CELANA PAN...	standar, cokl...	1
5	Cukup Laris	Cukup Laris	1	0	0	March	Celana Kulot ...	RANDOM 10...	3
6	Cukup Laris	Kurang Laris	0.003	0.000	0.997	November	NEONA ORO ...	jumbo, cream	1
7	Cukup Laris	Cukup Laris	1.000	0.000	0	September	ELI CUTBRA...	xl to jumbo, hi...	1
8	Cukup Laris	Cukup Laris	1.000	0.000	0	February	NEONA ATAS...	grey, LXL	1
9	Cukup Laris	Kurang Laris	0.000	0.000	1.000	July	NEONA ORO ...	jumbo, hitam	1
10	Cukup Laris	Cukup Laris	1	0	0	July	Celana Kulot ...	RANDOM 10...	3
11	Cukup Laris	Cukup Laris	1	0	0	July	Celana Kulot ...	RANDOM 10...	3
12	Cukup Laris	Cukup Laris	1.000	0.000	0	May	NEONA ATAS...	coklat mahog...	1
13	Cukup Laris	Kurang Laris	0.000	0.000	1.000	January	NEONA ORO ...	ivory, standar...	1
14	Cukup Laris	Kurang Laris	0.000	0.000	1.000	June	NEONA ORO ...	standar, moc...	1

ExampleSet (122 examples, 5 special attributes, 5 regular attributes)

(b)

Gambar 8. Data Training untuk (a) dan (b) Data Testing

Hasil pembagian dari *Split Data* tersebut terdapat pada gambar 8a dan 8b. Gambar 8a merupakan hasil dari data training yang telah dibagi menggunakan split data, sedangkan gambar 8b merupakan data testing yang telah diuji dan menghasilkan prediksi dari model yang telah dilakukan.

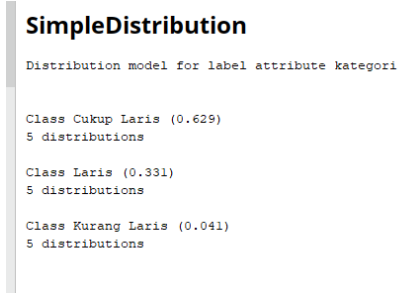


(a)

(b)

Gambar 9. Parameter Performance untuk (a) dan Proses Pengujian Model menggunakan RapidMiner (b)

Gambar 9a merupakan proses tahapan akhir adalah evaluasi performa menggunakan operator *Performance Classification* dengan menghitung nilai *accuracy*, *weighted mean precision*, dan *weighted mean recall*. Evaluasi ini berfungsi untuk menilai sejauh mana model mampu melakukan klasifikasi secara akurat sekaligus mengurangi risiko terjadinya *overfitting* maupun *underfitting*. Gambar 9b merupakan proses pengujian model dari awal sampai tahap akhir sehingga model dapat diimplementasikan.



Gambar 10. *SimpleDistribution* menggunakan RapidMiner

Pada Gambar 10 ditunjukkan hasil distribusi kelas dari model Naïve Bayes terhadap atribut kategori. Nilai probabilitas terbesar terdapat pada kelas Cukup Laris dengan proporsi 0,629, kemudian diikuti oleh kelas Laris sebesar 0,331, dan yang paling kecil adalah kelas Kurang Laris dengan nilai 0,041. Setiap kelas memiliki 5 distribusi sebagaimana dihasilkan oleh RapidMiner.

### 3.5 Evaluation

Model Naïve Bayes yang dikembangkan kemudian diuji menggunakan data testing dan dievaluasi melalui *confusion matrix 3x3*. Proses evaluasi ini dilakukan untuk menilai sejauh mana model dapat membedakan tiap kategori, yakni cukup laris, laris, dan kurang laris, dengan menghitung jumlah prediksi yang tepat maupun yang keliru pada masing-masing kelas.

**accuracy: 92.62%**

	true Cukup Laris	true Laris	true Kurang Laris	class precision
pred. Cukup Laris	65	1	0	98.48%
pred. Laris	0	39	0	100.00%
pred. Kurang Laris	8	0	9	52.94%
class recall	89.04%	97.50%	100.00%	

Gambar 11. Hasil *Accuracy* melalui RapidMiner

Gambar 11 merupakan hasil *Accuracy* yang dilakukan melalui RapidMiner. Dari pengujian tersebut diperoleh tingkat akurasi sebesar 92,62%, yang menunjukkan bahwa mayoritas data dapat diprediksi dengan benar oleh model. Berdasarkan hasil pengujian, pada kelas cukup laris sebanyak 65 data berhasil dikenali dengan benar, namun masih ada 8 data yang keliru terklasifikasi sebagai kategori kurang laris. Untuk kelas laris, sebanyak 39 data dapat diprediksi sesuai label aslinya dengan hanya 1 data yang salah teridentifikasi. Sementara itu, pada kelas kurang laris, terdapat 9 data yang berhasil diklasifikasikan dengan tepat, meskipun masih ditemukan 8 data yang salah teridentifikasi sebagai cukup laris.

**weighted\_mean\_precision: 83.81%, weights: 1, 1, 1**

	true Cukup Laris	true Laris	true Kurang Laris	class precision
pred. Cukup Laris	65	1	0	98.48%
pred. Laris	0	39	0	100.00%
pred. Kurang Laris	8	0	9	52.94%
class recall	89.04%	97.50%	100.00%	

Gambar 12. Hasil *Weighted Mean Precision* melalui RapidMiner

Gambar 12 merupakan nilai *weighted mean precision* sebesar 83,81% dengan rincian 98,48% untuk kelas cukup laris, 100% untuk kelas laris, serta 52,94% untuk kelas kurang laris.

**weighted\_mean\_recall: 95.51%, weights: 1, 1, 1**

	true Cukup Laris	true Laris	true Kurang Laris	class precision
pred. Cukup Laris	65	1	0	98.48%
pred. Laris	0	39	0	100.00%
pred. Kurang Laris	8	0	9	52.94%
class recall	89.04%	97.50%	100.00%	

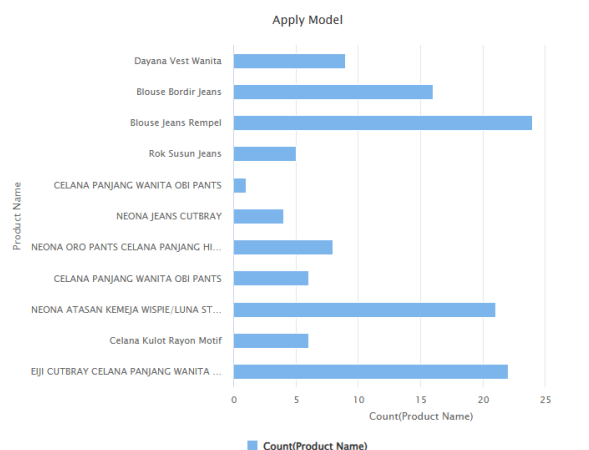
Gambar 13. Hasil *Weighted Mean Recall* melalui RapidMiner

Gambar 13 menunjukkan bahwa nilai *weighted mean recall* sebesar 95,51% dengan rincian masing-masing adalah 89,04% pada kelas cukup laris, 97,50% pada kelas laris, dan 100% pada kelas kurang laris.

Hasil pengujian menunjukkan bahwa algoritma Naïve Bayes dapat melakukan klasifikasi produk fashion wanita pada Noenaasstore ke dalam tiga kategori yaitu laris, cukup laris, dan kurang laris, dengan pemodelan yang dibangun melalui aplikasi RapidMiner. Penelitian ini menggunakan pendekatan CRISP-DM dengan pembagian data sebesar 80% untuk data latih dan 20% untuk data uji. Evaluasi model menunjukkan capaian akurasi sebesar 92,62%, *weighted mean precision* sebesar 83,81%, serta *weighted mean recall* sebesar 95,51%. Dibandingkan dengan penelitian terdahulu, capaian ini sedikit lebih rendah dari penelitian dengan judul Penerapan Algoritma Naïve Bayes Untuk Menentukan Produk Terlaris Pada Penjualan Pulsa [5] yang melaporkan akurasi 97,50%, *precision* 100%, dan *recall* 93,48%, namun lebih tinggi dibandingkan penelitian dengan judul Penerapan Algoritma Naive Bayes Untuk Prediksi Penjualan Produk Terlaris Pada CV Akusara Jaya Abadi [6] dengan akurasi 89,33%, *precision* 88,73%, dan *recall* 90,12%. Perbedaan hasil tersebut menunjukkan bahwa performa Naïve Bayes sangat dipengaruhi oleh karakteristik serta distribusi data yang digunakan. Tingkat akurasi sebesar 92,62% pada penelitian ini diperoleh dari perhitungan probabilitas dilakukan dengan memperhitungkan distribusi data pada setiap kelas berdasarkan nilai *mean* dan standar deviasi, sehingga kelas dengan perbedaan rata-rata yang jelas serta variasi yang relatif kecil yaitu dengan 10 produk sehingga dapat dikenali dengan lebih baik oleh model. Dengan memanfaatkan 612 data transaksi penjualan, model dapat menangkap pola dengan baik meskipun masih terdapat ketidakseimbangan pada kategori tertentu. Tujuan ini sejalan dengan tujuan penelitian yang diuraikan pada bagian pendahuluan, yakni menghadirkan analisis berbasis data untuk mendukung pengambilan keputusan bisnis. Lebih lanjut, model ini berpotensi dikembangkan dengan menambahkan variabel baru, melakukan perbandingan dengan algoritma lain, serta mengimplementasikannya langsung dalam sistem *e-commerce* sebagai strategi pemasaran yang lebih efektif dan sarana pengambilan keputusan yang lebih efisien.

### 3.6 Deployment

Tahap terakhir dari penelitian ini adalah deployment, yaitu penerapan model pada praktik bisnis.



Gambar 14. Hasil Penerapan Model Produk Noenaasstore melalui RapidMiner

Hasil model yang divisualisasikan pada gambar 14, menunjukkan bahwa beberapa produk, seperti Blouse Jeans Rempel, Neona Atasan Kemeja Wispie/Luna Style, dan Eiji Cutbray Celana Panjang Wanita, memiliki jumlah penjualan yang tinggi sehingga perlu diprioritaskan dalam penyediaan stok. Produk dengan penjualan sedang,

seperti Blouse Bordir Jeans dan Neona Oro Pants Celana Panjang Highwaist, dapat ditingkatkan melalui promosi yang lebih optimal, sedangkan produk dengan penjualan rendah, misalnya Celana Panjang Wanita Obi Pants dan Dayana Vest Wanita, lebih tepat diarahkan pada strategi pengembangan atau bundling produk. Selain sebagai dasar strategi bisnis, hasil penelitian ini juga disusun sesuai standar ilmiah agar dapat dijadikan rujukan untuk penelitian dan pengembangan model berikutnya.

#### 4. KESIMPULAN

Berdasarkan hasil pengujian, algoritma Naïve Bayes terbukti mampu mengklasifikasikan produk fashion wanita pada Noenaasstore ke dalam tiga kategori yaitu laris, cukup laris, dan kurang laris dengan menggunakan RapidMiner. Penelitian ini menerapkan metode CRISP-DM dengan pembagian data sebesar 80% untuk data latih dan 20% untuk data uji. Hasil evaluasi menunjukkan akurasi sebesar 92,62%, *weighted mean precision* sebesar 83,81%, dan *weighted mean recall* sebesar 95,51%. Temuan ini sejalan dengan tujuan penelitian yang disampaikan pada bagian pendahuluan, yakni menghadirkan analisis berbasis data guna mendukung pengambilan keputusan bisnis. Kedepannya, model ini berpotensi dikembangkan melalui penambahan variabel, perbandingan dengan algoritma lain, serta penerapan langsung pada sistem *e-commerce* untuk strategi pemasaran yang efektif dan pengambilan keputusan bisnis yang efisien.

#### UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Bapak Paminto Agung Christianto, M.Kom selaku Dosen Pembimbing I dan Ibu Nurul Amalia, M.Kom selaku pembimbing II atas bimbingan, arahan, serta dukungan yang diberikan selama proses penelitian hingga penulisan jurnal ini dapat diselesaikan.

#### REFERENSI

- [1] V. Kumala Sari and M. Irwan Padli Nasution, "Dampak E-commerce Terhadap Perkembangan Digital," *J. Akad. Ekon. Dan Manaj.*, vol. 1, no. 4, pp. 18–24, 2024, [Online]. Available: <https://doi.org/10.61722/jaem.v1i4.3113>
- [2] Jakpat Survey Report, "Indonesia E-Commerce Trends - 1st Semester of 2023," pp. 20–24, 2023, [Online]. Available: <https://insight.jakpat.net/indonesia-e-commerce-trends-1st-semester-of-2023/>
- [3] N. Sakinah, D. M. Nanda, and Tohiruddin, "Trend Fashion di Kalangan Mahasiswa-Mahasiswi Universitas Negeri Surabaya," *Pros. Semin. Nas. Ilmu Ilmu Sos.*, vol. 01, pp. 32–38, 2022, [Online]. Available: <https://proceeding.unesa.ac.id/index.php/sniis/article/view/37>
- [4] B. Aktavera and H. O. L. Wijaya, "Klasifikasi Produk Menggunakan Algoritma Decision Tree," *J. Teknol. Inf. Mura*, vol. 15, no. 1, pp. 24–29, 2024, doi: 10.32767/jti.v15i1.2264.
- [5] I. Nawangsih, A. Setyaningsih, P. Studi, T. Informatika, F. Teknik, and D. Mining, "ISSN 2337-6805 PENERAPAN ALGORITMA NAÏVE BAYES UNTUK MENENTUKAN KLASIFIKASI PRODUK TERLARIS PADA PENJUALAN PULSA THE APPLICATION OF THE NAÏVE BAYES ALGORITHM TO DETERMINE THE Penerapan Algoritma , Ismasari Nawangsih , Asti Setyaningsih ISSN 2337-6805 Pene," *Incomtech*, vol. 9, no. 1, pp. 39–45, 2020.
- [6] I. Pramana and I. W. Sudiarsa, "Penerapan Algoritma Naive Bayes Untuk Prediksi Penjualan Produk Terlaris Pada CV Akusara Jaya Abadi," *JATISI (Jurnal Tek. ....)*, vol. 10, no. 4, pp. 518–534, 2023, [Online]. Available: <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/6498%0Ahttps://jurnal.mdp.ac.id/index.php/jatisi/article/download/6498/1694>
- [7] E. B. Susanto, A. N. Anzila, and B. Ismanto, "Comparison Of The Effectiveness Of K-Nearest Neighbor (KNN) And Naive Bayes Algorithms In Identifying Diabetes Patients," *J. Artif. Intell. Softw. Eng.*, vol. 5, no. 1, p. 22, 2025, doi: 10.30811/jaise.v5i1.6275.