



KLASIFIKASI KATEGORI PRODUK TERLARIS PADA *E-COMMERCE* MENGGUNAKAN ALGORITMA NAIVE BAYES

Immanuel Marcell Sumual¹, Jonathan Supriadi², Ellena Effendy³, Andri Wijaya^{4(*)}

¹Universitas Katolik Musi Charitas, Palembang

²Universitas Katolik Musi Charitas, Palembang

³Universitas Katolik Musi Charitas, Palembang

⁴Universitas Katolik Musi Charitas, Palembang

Abstract

Naive Bayes algorithm used to classify the best-selling product categories in e-commerce. The data used comes from a public Kaggle dataset, comprising 250,000 transactions during the 2020–2023 period. The analysis process follows the CRISP-DM model, including stages such as business understanding, data preparation, modeling, and model evaluation using a confusion matrix. Evaluation results show that the model achieved an accuracy of 92.64%, precision of 91.51%, and recall of 96.68%. The analysis revealed that the best-selling product category is Clothing, followed by Books, Electronics, and Home. This study demonstrates that the Naive Bayes algorithm can be effectively implemented to support stock management and data-driven marketing strategies in e-commerce.

Kata Kunci: Classification, Naive Bayes, data mining, best-selling products

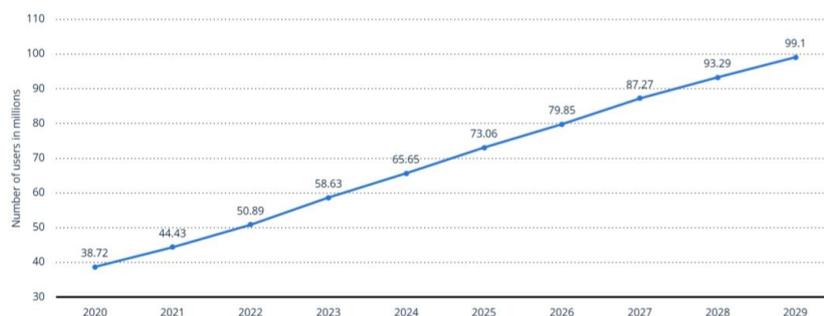
Juli – Desember 2024, Vol 5 (2) : hlm 51-60
©2024 Institut Teknologi dan Bisnis Ahmad Dahlan.
All rights reserved.

(*) Korespondensi: andri_wijaya@ukmc.ac.id (Andri Wijaya).

PENDAHULUAN

Perkembangan teknologi digital telah mengubah pola konsumsi masyarakat secara signifikan. Salah satu wujud transformasi ini adalah meningkatnya peran perdagangan elektronik atau *e-commerce* dalam kehidupan sehari-hari. *E-commerce* memungkinkan bisnis untuk menjual produk dan layanan secara *online*, memperluas jangkauan pasar, dan menghadirkan peluang baru melalui internet. Model bisnis *e-commerce* yang terus berkembang telah mendorong banyak pelaku usaha dari berbagai sektor untuk beradaptasi dan memanfaatkan potensi besar di dunia digital (Wijaya, 2023).

Dalam beberapa tahun terakhir, *e-commerce* telah menjadi salah satu sektor utama yang mendukung transaksi jual beli secara daring. Kemudahan akses, variasi produk, dan metode pembayaran yang fleksibel membuat *e-commerce* semakin diminati oleh konsumen. Jumlah pengguna *e-commerce* Indonesia meningkat setiap tahunnya dan diperkirakan pada tahun 2029, jumlah pengguna *e-commerce* di Indonesia dapat mencapai 99,1 juta pengguna (Indonesia, 2024).



Sumber: Kementerian Perdagangan Republik Indonesia, 2024

Gambar 1. Jumlah Pengguna E-Commerce di Indonesia periode 2020-2029

Di tengah pertumbuhan pesat ini, pelaku bisnis menghadapi tantangan besar untuk terus bertahan. Untuk tetap kompetitif, analisis terhadap data perusahaan menjadi salah satu cara efektif. Salah satu contohnya adalah analisis kategori produk terlaris. Analisis ini efektif membantu perusahaan untuk mengidentifikasi kebutuhan bisnis seperti preferensi konsumen, peningkatan kapasitas produk terlaris, serta mengambil keputusan yang lebih tepat dalam pengelolaan stok dan pemasaran (Harahap, et al., 2023). Untuk melakukan analisis ini, dibutuhkan sebuah teknik yang dapat mengubah tumpukan data perusahaan tersebut menjadi sebuah informasi berharga. Teknik seperti *data mining* menjadi alat yang sangat efektif untuk mengolah data perusahaan dan mengubahnya menjadi informasi yang berharga (Rahmatullah, et al., 2019).

Seiring perkembangan teknologi, analisis berbasis data menjadi alat yang semakin penting untuk mengatasi tantangan ini. *Data mining*, sebagai cabang ilmu komputer, menawarkan solusi dengan mengeksplorasi kumpulan data besar untuk mendapatkan wawasan yang berharga (Rofiani, et al., 2024). Dalam *data mining* terdapat beberapa fungsi, salah satu fungsi utama *data mining* adalah klasifikasi, yang digunakan untuk mengelompokkan data berdasarkan karakteristik tertentu. Dalam konteks *e-commerce*, klasifikasi dapat diterapkan untuk mengidentifikasi kategori produk yang paling diminati. Salah satu algoritma yang banyak digunakan untuk tugas klasifikasi adalah algoritma Naive Bayes.

Algoritma Naive Bayes dikenal memiliki keunggulan dalam menangani data yang bersifat kategorikal dan memiliki efisiensi dalam proses klasifikasi (Julianto & Andayani, 2024). Meskipun algoritma ini mengasumsikan independensi antar fitur, yang sering kali tidak sepenuhnya akurat dalam dunia nyata, hasilnya tetap memuaskan dalam berbagai implementasi (Haq, et al., 2024). Dengan memanfaatkan *dataset e-commerce* yang tersedia di platform *open*

source Kaggle, penelitian ini bertujuan untuk mengklasifikasikan kategori produk terlaris berdasarkan pola konsumsi digital.

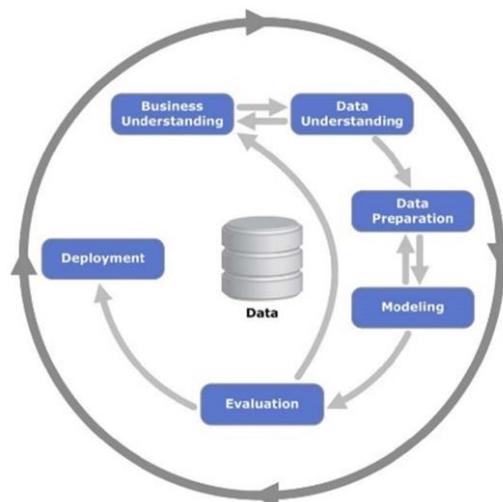
Pada penelitian yang dilakukan oleh Alvin Julianto dan Sri Andayani pada tahun 2024 dengan judul “Penerapan Data Mining untuk Klasifikasi Produk Terlaris Menggunakan Algoritma Naive Bayes Pada Bengkel Motor” menghasilkan model berbasis algoritma Naive Bayes. Penelitian ini berfokus pada pengelolaan stok produk di Bengkel Aldo Motor dengan menganalisis data penjualan untuk menentukan produk terlaris. Hasil penelitian menunjukkan bahwa sistem membagi produk berdasarkan status penjualannya dengan tingkat akurasi 86,77%, presisi 85,39%, dan *recall* 76%. Sistem ini mampu membantu pengelola bengkel dalam menganalisis pola penjualan dan mengurangi masalah kelebihan atau kekurangan stok. Kelemahan penelitian ini terletak pada keterbatasan data, yang hanya mencakup satu bengkel dan periode waktu tertentu.

Berdasarkan uraian tersebut, peneliti tertarik untuk melakukan penelitian yang berjudul “Klasifikasi Kategori Produk Terlaris pada *E-Commerce* Menggunakan Algoritma Naive Bayes”. Tujuan utama pada penelitian ini, yaitu menemukan pola konsumsi digital berdasarkan data transaksi di platform *e-commerce*, mengklasifikasikan kategori produk terlaris menggunakan algoritma Naive Bayes, serta memberikan wawasan kepada pelaku bisnis untuk mendukung strategi pemasaran berbasis data berdasarkan pola konsumsi digital.

METODE

Dalam penelitian ini dilakukan pengumpulan data dengan cara mengunduh melalui platform *open-source* Kaggle. *Dataset* dapat diakses melalui tautan berikut: <https://www.kaggle.com/datasets/shriyashjagtap/e-commerce-customer-for-behavior-analysis>. *Dataset* ini merupakan data transaksi yang terdiri dari 250.000 baris dengan rentang tahun tahun 2020-2023. Seluruh data dikumpulkan secara legal dan sesuai dengan kebijakan privasi sumber data.

Selanjutnya, pengolahan data dilakukan menggunakan model CRISP-DM. CRISP-DM (*Cross-Industry Standard Process for Data Mining*) adalah sebuah metodologi standar yang diterapkan lintas industri untuk aktivitas *data mining*.



Sumber: Pramana, et al, 2024

Gambar 2. Model CRISP-DM

Seperti yang diilustrasikan pada Gambar 2, alur kerja CRISP-DM terdiri dari beberapa fase yang saling terhubung, di mana setiap tahapan memiliki peran dan tujuan spesifik. Konektivitas antar tahap digambarkan melalui tanda panah, menunjukkan bahwa urutan pelaksanaan bersifat fleksibel dan dapat disesuaikan berdasarkan hasil yang diperoleh pada

setiap fase. Lingkaran yang mengelilingi setiap fase melambangkan sifat siklus dari proses *data mining* (Denni Kurniawan, 2022). Enam tahapan utama model ini adalah sebagai berikut, *business understanding, data understanding, data preparation, modeling, evaluation, dan deployment* (Sastya & Nugraha, 2023).

Business Understanding

Tahap ini menjelaskan tujuan bisnis untuk menentukan pola yang dicari dalam proses *data mining*. Kemudian untuk menjawab permasalahan penelitian, maka dibuat laporan klasifikasi kategori produk terlaris agar pihak *e-commerce* dapat menentukan manajemen stok produk maupun meningkatkan nilai penjualannya

Data Understanding

Tahap ini penulis melakukan proses pengumpulan data dan menganalisis data. Dalam hal ini, data yang digunakan penulis adalah data yang berasal dari website *Kaggle*, diperoleh data detail penjualan selama 3 tahun sebanyak 250.000 *records* dan 13 atribut (*features*) dimulai dari tahun 2020-2023. Adapun atribut-atribut yang terdapat dalam *dataset* seperti: *Customer ID, Purchase Date, Product Category, Product Price, Quantity, Total Purchase Amount, Payment Method, Customer Age, Returns, Customer Name, Age, Gender, dan Churn*

Data Preparation

Merupakan tahap di mana kita menyiapkan data untuk pemodelan. Tahap ini termasuk *data cleansing, construct data, integrate data, dan format data*.

Modeling

Merupakan tahap di mana kita menggunakan algoritma Naive Bayes sebagai model klasifikasi produk terlaris pada *e-commerce*.

Evaluation

Pada tahap ini akan dilakukan evaluasi *Confusion Matrix* untuk data yang telah divisualisasikan dan model. *Confusion Matrix* merupakan sebuah teknik evaluasi yang umumnya dipakai untuk mengkalkulasi tingkat akurasi pada *data mining*. Evaluasi dengan menggunakan metode ini menghasilkan nilai *accuracy, precision, dan recall*. Nilai *accuracy* adalah rasio dari jumlah total perkiraan yang benar. Semakin tinggi nilai *accuracy* maka semakin tinggi perkiraan kebenaran lebih banyak. Nilai *precision* adalah nilai yang digunakan untuk menilai seberapa baik sebuah model dalam memperkirakan kelas. Nilai *recall* adalah nilai yang dihitung untuk menilai kemampuan model dalam memprediksi kelas tertentu (Julianto & Andayani, 2024).

Contoh penerapan data parsial pada metode *Confusion Matrix* diberikan pada Tabel 1 di bawah ini.

Tabel 1. Contoh Data Parsial pada *Confusion Matrix*

<i>Confusion Matrix</i>		Kelas Hasil Prediksi (<i>Predicted</i>)	
		Kelas (+)	Kelas (-)
Kelas Asli (Actual)	Kelas (+)	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
	Kelas (-)	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

Adapun rumus untuk metode *Confusion Matrix* sebagai berikut (Pramana, et al., 2023):

$$1. \text{ Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$2. \text{ Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$3. \text{ Recall} = \frac{TP}{TP + FN} \quad (3)$$

Adapun keterangan untuk rumus di atas sebagai berikut:

1. *True Positive* (TP) adalah jumlah baris positif yang diklasifikasikan sebagai positif.
2. *False Positive* (FP) adalah jumlah baris negatif yang diklasifikasikan sebagai positif.
3. *False Negative* (FN) adalah jumlah baris positif yang diklasifikasi sebagai negatif.
4. *True Negative* (TN) adalah jumlah baris negatif yang diklasifikasikan sebagai negatif.

Nilai *accuracy* dapat dilihat berdasarkan beberapa kategori yang disajikan pada Tabel 2.

Tabel 2 Kategori Nilai AUC

NILAI AUC	Kategori Klasifikasi
0.90-1.00	Excellent
0.80-0.90	Good
0.70-0.80	Fair
0.60-0.70	Poor
0.50-0.60	Failure

Sumber: (Irawan, 2018)

Untuk nilai *precision* dan *recall* direpresentasikan ke dalam tiga kategori kelas yang dapat dilihat pada Tabel 3 di bawah ini.

Tabel 3 Kategori *Precision* dan *Recall*

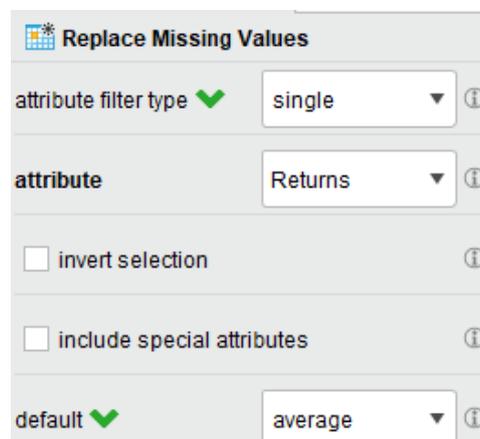
Nilai Precision dan Recall	Kategori Klasifikasi
0.67-1.00	Tinggi
0.34-0.66	Sedang
0.00-0.33	Rendah

Sumber: (Julianto & Andayani, 2024)

HASIL DAN PEMBAHASAN

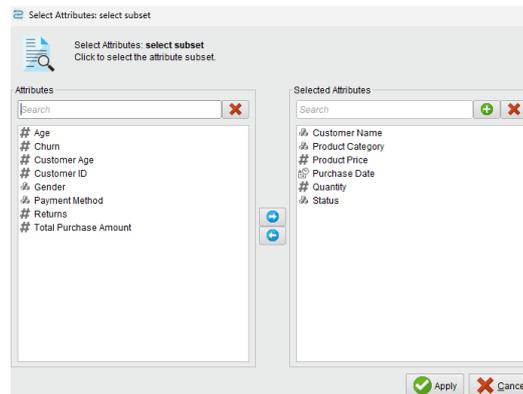
Data Preparation

Fase ini menjelaskan proses persiapan data yang siap untuk dianalisis dengan algoritma *Naive Bayes*. Dalam hal ini, terdapat *missing values* pada atribut *Returns*. Maka untuk langkah awal adalah mengisi data yang hilang dengan operator *Replace Missing Values* dalam *Parameters Replace Missing Values* pada *attribute filter type*-nya, yaitu *single*, pada *attribute*-nya yaitu *Returns*, dan pada *default*-nya yaitu *average* yang disajikan pada Gambar 3.



Gambar 3. *Replace Missing Values* dengan *RapidMiner*

Setelah itu adalah pemilihan atribut (*select attributes*) sebanyak 6 atribut yang termasuk dengan *class*-nya pada Gambar 4.



Gambar 4. *Select Attributes* dengan *RapidMiner*

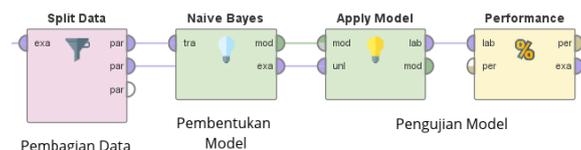
Selanjutnya tahap terakhir dari *data preparation*, yaitu melakukan transformasi data beserta menentukan *class*/label-nya, yaitu Status. Dalam hal ini, sebelum melanjutkan ke fase *modeling*, diperlukan normalisasi atribut pada *Quantity* yang berkorelasi dengan *class* Status. Adapun ketentuan atribut *Quantity*, yaitu sebagai berikut:

1. Jika *Quantity* yang terjual kurang dari dari 2, maka dinyatakan *Quantity* tidak laris.
2. Jika *Quantity* yang terjual lebih dari dari 2, maka dinyatakan *Quantity* laris.

Modeling

Proses *modeling* bertujuan untuk membentuk pola dan menguji model *Naive Bayes* dengan cara membagi *dataset* menjadi *data training* dan *data testing* dengan masing-masing bobot 80:20, dimana 80% untuk *data training* dan 20% untuk *data testing*. Di mana pembagian *dataset* menggunakan teknik *stratified sampling*. Adapun teknik tersebut sudah tersedia melalui *RapidMiner* dalam *Parameters Split Data* pada *sampling type*-nya, yaitu *stratified sampling*. *Stratified sampling* bertujuan untuk membagi set data ke k partisi secara seimbang (*balance*), juga hal ini berkaitan dengan distribusi kelas di setiap partisi harus sama dengan distribusi kelas di set data originalnya. Maka dari itu, teknik *sampling* sangat penting untuk kasus klasifikasi dalam menghasilkan model yang konsisten.

Setelah melakukan pembagian data, selanjutnya membentuk model algoritma *Naive Bayes*. *RapidMiner* memiliki operator bernama *Naive Bayes* untuk membentuk model serta melatih *dataset* yang siap untuk diolah dan dianalisis. Kemudian untuk pengujian model menggunakan operator *Apply Model* yang berfungsi untuk menguji model *Naive Bayes* berdasarkan *connection operator Split Data*. Selain *Apply Model*, diperlukan operator yang dapat mengukur kinerja dari suatu pola dari model yang telah dihasilkan, yaitu dengan *confusion matrix*. Dalam hal ini, *RapidMiner* memiliki operator untuk menampilkan evaluasi kinerja model dengan operator *Performance (Classification)*. Perihal proses pembagian data, pembentukan model dan pengujian model pada *RapidMiner* dapat dilihat pada Gambar 5.



Gambar 5 Pembagian Data, Pembentukan dan Pengujian Model dengan *RapidMiner*

Setelah membentuk model melalui proses pembagian *data training* sebesar 80% dan *data testing* sebesar 20%. Implementasi *RapidMiner* dalam *dataset* yang digunakan untuk *data*

training menghasilkan sebanyak 200.000 records pada Gambar 6 dan data testing sebanyak 50.000 records pada Gambar 7.

Row No.	Status	Purchase D...	Product Category	Product Price	Quantity	Customer Name
1	Laris	Sep 8, 2020 ...	Electronics	12	3	Christine Hernandez
2	Laris	Mar 5, 2022 1...	Home	468	4	Christine Hernandez
3	Tidak Laris	May 23, 2022 ...	Home	288	2	Christine Hernandez
4	Tidak Laris	Nov 12, 2020 ...	Clothing	196	1	Christine Hernandez
5	Laris	Mar 7, 2023 2...	Home	250	4	James Grant
6	Tidak Laris	Apr 15, 2023 ...	Electronics	73	1	James Grant
7	Tidak Laris	Mar 27, 2021 ...	Books	337	2	James Grant
8	Tidak Laris	May 5, 2020 8...	Clothing	182	2	James Grant
9	Tidak Laris	Sep 13, 2023...	Electronics	394	2	Jose Collier
10	Tidak Laris	Jan 18, 2021 ...	Books	348	1	Jose Collier
11	Laris	Jan 7, 2020 1...	Books	103	4	Jose Collier
12	Tidak Laris	Apr 25, 2021 ...	Books	30	1	James Stein
13	Laris	Jan 13, 2020 ...	Books	153	5	James Stein
14	Tidak Laris	Jun 18, 2023 ...	Clothing	259	1	James Stein

Gambar 6 Hasil Data Training sebanyak 200.000 Records/Examples

Row No.	Status	prediction S...	confidence...	confidence...	Purchase D...	Product Cat...	Product Price	Quantity	Customer N...
1	Tidak Laris	Tidak Laris	0.000	1.000	Nov 27, 2020...	Home	449	1	James Grant
2	Tidak Laris	Tidak Laris	0.001	0.999	Mar 31, 2021...	Clothing	356	1	Jose Collier
3	Tidak Laris	Tidak Laris	0.001	0.999	Feb 12, 2021...	Books	240	1	Jose Collier
4	Tidak Laris	Tidak Laris	0.001	0.999	Jul 2, 2020 2...	Clothing	368	1	Jose Collier
5	Laris	Laris	0.974	0.026	Sep 10, 2021...	Books	489	3	James Stein
6	Laris	Laris	0.985	0.015	Jun 20, 2020...	Books	227	3	Sonia Moreno
7	Tidak Laris	Tidak Laris	0.112	0.888	Aug 21, 2020...	Electronics	382	2	Sonia Moreno
8	Laris	Laris	1.000	0.000	May 10, 2020...	Books	434	4	Jam Mendosa
9	Laris	Laris	0.985	0.015	Jan 20, 2022...	Clothing	101	3	Cori Zuniga
10	Laris	Laris	0.963	0.037	Mar 10, 2020...	Electronics	276	3	Shelly Higgins
11	Tidak Laris	Tidak Laris	0.039	0.961	Oct 3, 2022 1...	Home	291	2	Steve Patton
12	Tidak Laris	Tidak Laris	0.001	0.999	May 19, 2021...	Clothing	124	1	Steve Patton
13	Laris	Laris	1.000	0.000	Aug 17, 2022...	Books	18	5	Steve Patton
14	Tidak Laris	Tidak Laris	0.168	0.832	Nov 30, 2020...	Home	450	2	Kimberly Bur...

Gambar 7 Hasil Data Testing sebanyak 50.000 Records/Examples

Setelah mengetahui hasil pembentukan model *Naive Bayes*. Terdapat hal yang penting dalam analisis *data mining*, yaitu mengetahui tingkat probabilitas dari suatu atribut dengan *class*-nya (*prior probability*). Maka distribusi kelas sangat diperlukan dalam analisis prediksi *class*/label dengan data atributnya. Distribusi kelas pada penelitian ini menggunakan operator *RapidMiner*, yaitu *Simple distribution* yang diperoleh dari proses *data training* dengan *Naive Bayes* pada *RapidMiner* memperoleh hasil distribusi nilai *Class* “Laris” sebesar 0.599 dan *Class* “Tidak Laris” sebesar 0.401 pada Gambar 8. Dalam hal ini, nilai tersebut muncul dari nilai *prior probability* pada *class*/label *Status*, beserta probabilitas atribut umum (*likelihood*) yang digunakan sebagai *predictor*. Dengan begitu, hasil nilai distribusi tersebut berasal dari *data training* sebanyak 200.000 records.

SimpleDistribution

Distribution model for label attribute Status

Class Laris (0.599)

5 distributions

Class Tidak Laris (0.401)

5 distributions

Gambar 8 Simple Distribution pada Class/Label Status

Evaluation

Fase yang paling terakhir, yaitu *evaluation* yang bertujuan untuk mengevaluasi model *Naive Bayes* yang telah dianalisis melalui pola-pola data yang telah dilakukan proses *data training* dan *data testing* pada tahap *modeling*. Dengan demikian, untuk pengujian performa model *Naive Bayes* dilakukan pada hasil *data testing* menggunakan *confusion matrix*. Adapun hasil data uji (*data testing*) pada *RapidMiner* diperoleh data sebanyak 50.000 records, yaitu:

1. Sebanyak 28.980 kelas positif dinyatakan produk laris (TP), dan 17.338 kelas negatif dinyatakan produk tidak laris (TN).
2. Kelas positif yang prediksinya laris dan kenyataannya (actual) tidak laris sebanyak 2688 (FP).
3. Sedangkan kelas negatif yang prediksinya tidak laris dan kenyataannya (actual) laris sebanyak 994 (FN).

Sehingga perhitungan tersebut dirumuskan beberapa pertanyaan seperti:

- a. *Accuracy*: “Berapa persen jumlah kelas Status yang Laris dan Tidak Laris yang diprediksi dengan benar?”
- b. *Precision*: “Berapa persen kelas Status yang benar kategori produk Laris dari keseluruhan kelas Status yang diprediksi Laris?”
- c. *Recall*: “Berapa persen kelas Status yang diprediksi kategori produk Laris dibandingkan keseluruhan kelas Status yang sebenarnya Tidak Laris?”

Setelah itu, dilakukan pembentukan struktur tabulasi *confusion matrix* seperti Tabel 4 di bawah ini.

Tabel 4 Struktur Tabulasi *Confusion Matrix*

Confusion Matrix		Kelas Hasil Prediksi (Predicted)	
		Laris (+)	Tidak Laris (-)
Kelas Asli (Actual)	Laris (+)	28.980 (TP)	994 (FN)
	Tidak Laris (-)	2688 (FP)	17338 (TN)

Perihal hasil perhitungan *accuracy*, *precision*, *recall* secara manual dan dengan *tools RapidMiner* dijelaskan sebagai berikut:

1. Perhitungan *Accuracy*

$$\begin{aligned}
 Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\
 &= \frac{28.980 + 17.338}{28.980 + 17.338 + 994 + 2688} = \frac{46.318}{50.000} = 0,9263 \\
 &= 0,9263 * 100\% \\
 &= 92,64\%
 \end{aligned}$$

Dari hasil *accuracy* di atas dapat dikatakan bahwa jumlah kelas Status yang Laris dan Tidak Laris yang diprediksi dengan benar adalah sebesar 92,64%. Selain itu, hasil formulasi *accuracy* yang dilakukan secara manual dengan *RapidMiner* memperoleh hasil yang sama. Hasil *accuracy* di *RapidMiner* disajikan pada Gambar 9.

accuracy: 92.64%

	true Laris	true Tidak Laris	class precision
pred. Laris	28980	2688	91.51%
pred. Tidak Laris	994	17338	94.58%
class recall	96.68%	86.58%	

Gambar 9. Hasil Evaluasi *Naive Bayes Model Accuracy* pada *RapidMiner*

2. Perhitungan *Precision*

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \\
 &= \frac{28.980}{28.980 + 2688} = \frac{28.980}{31.668} = 0,9151 \\
 &= 0,9151 * 100\% \\
 &= 91,51\%
 \end{aligned}$$

Dari hasil *precision* di atas dapat dikatakan bahwa kelas Status yang benar produk laris dari keseluruhan kelas Status yang diprediksi laris sebesar 91,51%. Selain itu, hasil formulasi *precision* yang dilakukan secara manual dengan *RapidMiner* memperoleh hasil yang sama. Hasil *precision* di *RapidMiner* disajikan pada Gambar 10.

precision: 91.51% (positive class: Laris)

	true Tidak Laris	true Laris	class precision
pred. Tidak Laris	17338	994	94.58%
pred. Laris	2688	28980	91.51%
class recall	86.58%	96.68%	

Gambar 10. Hasil Evaluasi *Naive Bayes Model Precision* pada *RapidMiner*

3. Pehitungan *Recall*

$$\begin{aligned}
 \text{Recall} &= \frac{TP}{TP + FN} \\
 &= \frac{28.980}{28.980 + 994} = \frac{28.980}{29.974} = 0,9668 \\
 &= 0,9668 * 100\% \\
 &= 96,68\%
 \end{aligned}$$

Dari hasil *recall* di atas dapat dikatakan bahwa kelas Status yang diprediksi produk laris dibandingkan keseluruhan kelas Status yang sebenarnya tidak laris sebesar 96,68%. Selain itu, hasil formulasi *recall* yang dilakukan secara manual dengan *RapidMiner* memperoleh hasil yang sama. Hasil *recall* di *RapidMiner* disajikan pada Gambar 11.

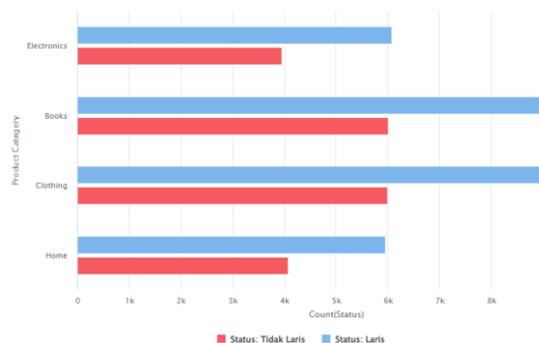
recall: 96.68% (positive class: Laris)

	true Tidak Laris	true Laris	class precision
pred. Tidak Laris	17338	994	94.58%
pred. Laris	2688	28980	91.51%
class recall	86.58%	96.68%	

Gambar 11 . Hasil Evaluasi *Naive Bayes Model Recall* pada *RapidMiner*

Analisis Prediksi

Kategori Produk Terlaris terbanyak dari 4 kategori, yaitu *Clothing* yang merupakan kaetgori produk terlaris dari keseluruhan kategori produk yang dijual oleh *e-commerce*, karena kategori produk tersebut memiliki probabilitas transaksi penjualan produk yang tertinggi, yaitu sebanyak 8981 kali transaksi, Kemudian kategori produk terlaris berikutnya diperoleh *Books* sebanyak 8961 kali transaksi, *Electronics* sebanyak 6072 kali transaksi, dan *Home* sebanyak 5960 kali transaksi. Adapun hasil visualisasi kategori produk terlaris tersebut disajikan pada Gambar 12.



Gambar 12. Kategori Produk Terlaris *e-commerce*

KESIMPULAN

Berdasarkan pembahasan di yang telah dijabarkan dalam penelitian ini, maka dapat ditarik kesimpulan bahwa algoritma *Naive Bayes* dapat dipakai untuk penelitian klasifikasi produk terlaris karena memiliki hasil sebagai berikut:

1. Nilai *accuracy* jumlah kelas Status yang Laris dan Tidak Laris yang diprediksi dengan benar adalah sebesar 92,64% atau setara dengan 0,9264 yang termasuk ke dalam kategori klasifikasi ‘Excellent’.
 2. Nilai *precision* jumlah kelas Status yang diprediksi laris sebesar 91,51% atau setara dengan 0,9151 yang termasuk ke dalam kategori ‘Tinggi’.
 3. Nilai *recall* jumlah kelas Status yang diprediksi produk laris dibandingkan produk tidak laris sebesar 96,68% atau setara dengan 0,9668 yang termasuk ke dalam kategori ‘Tinggi’.
- Berdasarkan hasil penelitian yang dikerjakan dan kesimpulan yang didapat, maka saran yang ditujukan kepada peneliti adalah melakukan perbandingan dengan algoritma klasifikasi lain seperti C45, Neural Network, dan Support Vector Machine (SVM).

DAFTAR PUSTAKA

- Denni Kurniawan, M. Y., 2022. **Optimization Sentiment Analysis Using CRISP-DM and Naive Bayes Methods Implemented on Social Media**. *Cyberspace: Jurnal Pendidikan Teknologi Informasi*, pp. 74-85.
- Haq, M. Z. et al., 2024. **Algoritma Naive Bayes untuk Mengidentifikasi Hoaks di Media Sosial**. *Jurnal Minfo Polgan*, pp. 1079-1084.
- Harahap, F. et al., 2023. **Implementasi Data Mining dalam Memprediksi Produk AC Terlaris untuk Meningkatkan Penjualan Menggunakan Metode Naive Bayes**. *Jurnal Unitek*, pp. 41-51.
- Indonesia, K. P. R., 2024. **Perdagangan Digital (E-Commerce) Indonesia Periode 2023**, s.l.: Pusat Data dan Sistem Informasi Sekretariat Jenderal Kementrian Perdagangan.
- Irawan, F., 2018. **A Comparative Assessment of Random Forest and SVM Algorithms, Using Combination of Principal Component Analysis and SMOTE for Accounts Receivable Seamless Prediction**. Bali, s.n.
- Julianto, A. & Andayani, S., 2024. **Penerapan Data Mining Untuk Klasifikasi Produk Terlaris Menggunakan Algoritma Naive Bayes pada Bengkel Motor**. *Jurnal Sistem & Teknologi Informasi Komunikasi*, pp. 50-58.
- Pramana, I. M. A. A., Sudiarsa, I. W. & Nugraha, P. G. S. C., 2023. **Penerapan Algoritma Naive Bayes untuk Prediksi Penjualan Produk Terlaris pada CV Akusara Jaya Abadi**. *Jurnal Teknik Informatika dan Sistem Informasi*, pp. 518-534.
- Rahmatullah, S., M. & Pramitha, M. N., 2019. **Data Mining untuk Menentukan Produk Terlaris Menggunakan Metode Naive Bayes**. *Jurnal Informasi dan Komputer*, pp. 57-64.
- Rofiani, R., Oktaviani, L., Vernanda, D. & Hendriawan, T., 2024. **Penerapan Metode Klasifikasi Decision Tree dalam Prediksi Kanker Paru-paru Menggunakan Algoritma C4.5**. *Jurnal Teknokompak*, pp. 126-139.
- Sastya, N. C. & Nugraha, I., 2023. **Penerapan Metode CRISP-DM dalam Menganalisis Data untuk Menentukan Customer Behavior di MeatSolution**. *Jurnal Pendidikan dan Aplikasi Industri*, pp. 103-115.
- Wijaya, O. T. H., 2023. **E-Commerce: Perkembangan, Tren, dan Peraturan Perundang-undangan**. *Jurnal Ilmiah Ekonomi dan Bisnis*, pp. 41-47.