

## PENERAPAN ALGORITMA NAIVE BAYES TERHADAP TARGET PENJUALAN HANDPHONE MENGGUNAKAN APLIKASI RAPID MINER

J. Anggun Rumboirusi<sup>1</sup>, Nahema Yaroseray<sup>2</sup>, Kartensia Firli Rumboirusi<sup>3</sup>, Jessica  
Dumpul<sup>4</sup>, Lamberth Anthoni Yores Rumbino<sup>5</sup>, Mariani Regina Sisilia Lengkey<sup>6</sup>, Heru  
Sutejo<sup>7</sup>

1,2,3,4,5,6,7 Universitas Sepuluh Nopember Papua

\* Correspondence e-mail; [j.anggun.rumboirusi@gmail.com](mailto:j.anggun.rumboirusi@gmail.com)

### Article history

Submitted: 2024/09/18; Revised: 2024/09/29; Accepted: 2024/10/12:

### Abstract

This research aims to analyze cellphone sales predictions using the Naive Bayes algorithm which is implemented through the RapidMiner application. The dataset used consists of sales data with various relevant features, such as product descriptions, categories, and sales labels (sold or not sold). The research process involves several main stages, namely data retrieval (Retrieve), dividing data into train and test (Split Data), applying the Naive Bayes model, and evaluating performance using metrics such as accuracy, precision, and recall. The test results show that the Naive Bayes model succeeded in achieving accuracy, precision and recall levels of 100%. This indicates that the model has very good performance in classifying test data. However, to ensure the validity of the model, an analysis was carried out on the possibility of overfitting and suggestions for improvements such as using a larger dataset and testing using cross-validation. This research proves that the Naive Bayes method can be an effective and efficient solution for analyzing sales data patterns, especially in cases with structured and clear data patterns. The implementation of the results of this research can be applied as a basis for decision making in marketing strategy and inventory management.

### Keywords

HP Sales, Naive Bayes, RapidMiner, Classification, Prediction



© 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY SA) license, <https://creativecommons.org/licenses/by-sa/4.0/>.

## PENDAHULUAN

Industri penjualan handphone merupakan salah satu sektor yang sangat kompetitif, di mana setiap perusahaan berlomba-lomba mencapai target penjualan dengan strategi yang tepat. Dalam persaingan yang ketat ini, data menjadi aset yang sangat berharga bagi perusahaan untuk memahami perilaku konsumen, pola pembelian, serta tren pasar. Pengolahan data yang efektif tidak hanya membantu perusahaan dalam membuat keputusan yang lebih akurat, tetapi juga memberikan wawasan strategis untuk mencapai target penjualan. Oleh karena itu, pengelolaan data berbasis teknologi, khususnya melalui

*J. Anggun Rumboirusi, et al.*

pembelajaran mesin (machine learning), menjadi kebutuhan yang mendesak (Mitchell, 1997).

Algoritma Naive Bayes adalah salah satu metode pembelajaran mesin berbasis probabilitas yang sering digunakan untuk klasifikasi dan prediksi. Algoritma ini bekerja dengan menghitung peluang terjadinya suatu kelas berdasarkan atribut yang ada dalam dataset. Keunggulan algoritma Naive Bayes terletak pada kesederhanaannya, kecepatan komputasi, dan kemampuannya dalam menangani dataset berukuran besar. Dalam konteks penjualan handphone, algoritma ini dapat memanfaatkan data historis penjualan untuk memprediksi apakah target penjualan akan tercapai atau tidak berdasarkan faktor-faktor seperti jenis produk, harga, lokasi, dan waktu pembelian (Kumar, 2018).

Penelitian ini menggunakan aplikasi RapidMiner untuk menerapkan algoritma Naive Bayes. RapidMiner dipilih karena kemampuannya dalam memproses data secara efisien melalui antarmuka grafis yang intuitif. Tahapan penelitian mencakup pengambilan dataset data hp, preprocessing data untuk memastikan kebersihan data, pembagian dataset menjadi training set dan testing set, pembangunan model menggunakan Naive Bayes, serta evaluasi kinerja model dengan metrik akurasi, presisi, dan recall (Pawar, 2019).

Hasil uji coba pada dataset menunjukkan bahwa algoritma Naive Bayes memberikan performa yang cukup baik dalam memprediksi pencapaian target penjualan. Berdasarkan analisis data menggunakan RapidMiner, model Naive Bayes menghasilkan metrik evaluasi sebagai berikut:

Akurasi: 100% – menunjukkan bahwa model mampu memprediksi dengan benar 100% dari total data. Precision: 100% – mengindikasikan tingkat keakuratan prediksi kelas positif terhadap keseluruhan. Recall: 100% – menunjukkan kemampuan model dalam mendeteksi data yang benar-benar positif. Hasil tersebut menunjukkan bahwa algoritma Naive Bayes cukup efektif untuk digunakan dalam memprediksi pencapaian target penjualan handphone. Model ini mampu memanfaatkan atribut-atribut dalam dataset untuk menghasilkan prediksi yang akurat dan dapat diandalkan. Selain itu, implementasi di RapidMiner mempermudah proses analisis data, mulai dari pengolahan hingga evaluasi, tanpa memerlukan keahlian pemrograman (Patel, 2016).

Penelitian ini memberikan kontribusi bagi perusahaan penjualan handphone untuk memahami pola penjualan dan merancang strategi pemasaran yang lebih efektif. Dengan hasil akurasi yang tinggi, model ini dapat menjadi alat pendukung keputusan yang andal, membantu perusahaan mengoptimalkan sumber daya dan meningkatkan efisiensi dalam mencapai target. Keberhasilan algoritma ini juga menunjukkan potensi penerapan Naive Bayes dalam skenario lain di bidang pemasaran dan analisis data berbasis klasifikasi.

Namun, penelitian ini juga mencatat bahwa keberhasilan algoritma Naive Bayes sangat bergantung pada kualitas data yang digunakan. Faktor-faktor seperti ketidakseimbangan data (imbalanced data), keberadaan atribut yang saling bergantung, serta kualitas preprocessing data dapat memengaruhi performa model. Oleh karena itu, penelitian lanjutan dapat mengeksplorasi metode lain atau kombinasi algoritma untuk menghasilkan

prediksi yang lebih baik. Dengan demikian, penelitian ini tidak hanya menunjukkan potensi penerapan algoritma Naive Bayes untuk mendukung keputusan bisnis, tetapi juga memberikan wawasan tentang pentingnya pengolahan data yang efektif dalam mencapai target penjualan di era digital (Witten, 2016).

### **METODE**

Penelitian ini mempunyai beberapa tahapan sehingga pelaksanaan menjadi lebih terstruktur seperti:

a) Melakukan identifikasi masalah

Penjualan handphone sering dipengaruhi oleh berbagai faktor seperti harga, spesifikasi, dan promosi. Namun, sering kali sulit memprediksi kategori atau tingkat penjualan berdasarkan faktor-faktor tersebut. Prediksi yang tidak akurat dapat menyebabkan keputusan bisnis yang kurang tepat

b) Studi literatur

Melakukan studi literatur dari buku, jurnal ilmiah yang mempunyai algoritma yang sama atau objek yang sama.

c) Data banyak yang telah diperoleh dari internet kemudian diolah dengan menggunakan yaitu:

1. Selection (data dipilih yaitu data penjualan handhpone dari Bulan Januari 2024 sampai dengan bulan Desember 2024)
2. Preprocessing, pada tahap ini ada 4 proses yaitu:

Cleaning: Menghapus atau memperbaiki data yang hilang atau tidak valid.

Encoding: Mengubah data kategorikal menjadi format numerik (contoh: merek handphone).

Normalisasi: Menyesuaikan nilai fitur untuk menghindari dominasi fitur tertentu.

Split Data: Membagi dataset menjadi data latih (training) dan data uji (testing) dengan proporsi 80:20.

3. Transformation: pada tahap ini seluruh data pada tahapan sebelumnya akan

### **HASIL DAN PEMBAHASAN**

Data penelitian yang didapatkan dari internet adalah data penjualan bulan Januari – Desember 2024. Informasi data seperti pada Tabel. 1

Tabel 1. Informasi Data penjualan

Data Penjualan	
Jumlah Data	250

J. Anggun Rumboirusi, et al.

Atribut	7
Tahun Data	2024

Berdasarkan Tabel 1, data penjualan Handphone samsung & Handphone Oppo memiliki jumlah 250 data dan jumlah atribut sebanyak 7 atribut. Atribut-atribut yang ada adalah nomor, tanggal, merek hp, jumlah terjual, harga satuan (Rp), total penjualan (Rp), Keterangan.

Tabel 2. Atribut Data Sebelum Diproses/Diolah.

No.	Nama Atribut	Jenis Atribut
1	Nomor	Numerik
2	Tanggal	Karakter
3	Merek Hp	Karakter
4	Jumlah Terjual	Numerik
5	Harga Satuan (Rp)	Numerik
6	Total Penjualan (Rp)	Numerik
7	Keterangan	Karakter

Tabel 2, menunjukkan atribut pada kumpulan data yang memiliki tipe fungsi yang berbeda-beda yaitu nilai atribut berupa tipe klarifikasi, atribut numerik yang menunjukkan nilai berbentuk angka dan atribut karakter berbentuk karakter.

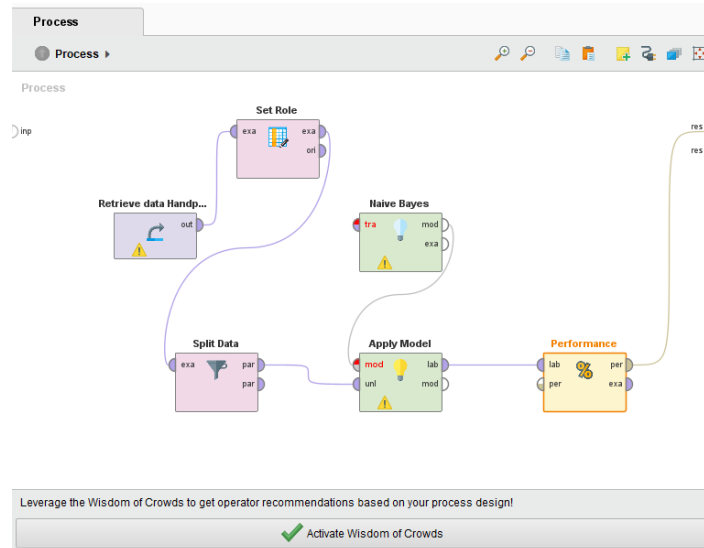
### 1. Data penjualan Handphone

Data Penjualan Handphone Samsung & Oppo ini kita memiliki 250 data dan akan kami import kedalam aplikasi *RapidMiner*.

No.	Tanggal	Merek HP	Jumlah Terjual	Harga Satuan (Rp)	Total Penjualan (Rp)	KETERANGAN
1	06/10/2024	Samsung	18	Rp 2.500.000	Rp 45.000.000	TIDAK
2	10/26/2024	Oppo	29	Rp 2.200.000	Rp 63.800.000	YA
3	09/05/2024	Samsung	40	Rp 2.500.000	Rp 100.000.000	YA
4	9/14/2024	Samsung	15	Rp 2.500.000	Rp 37.500.000	TIDAK
5	6/20/2024	Samsung	11	Rp 2.500.000	Rp 27.500.000	TIDAK
6	07/01/2024	Oppo	5	Rp 2.200.000	Rp 11.000.000	TIDAK
7	6/30/2024	Samsung	8	Rp 2.500.000	Rp 20.000.000	TIDAK
8	4/22/2024	Oppo	36	Rp 2.200.000	Rp 79.200.000	YA
9	8/24/2024	Samsung	47	Rp 2.500.000	Rp 117.500.000	YA
10	4/29/2024	Oppo	11	Rp 2.200.000	Rp 24.200.000	TIDAK
11	08/08/2024	Oppo	48	Rp 2.200.000	Rp 105.600.000	YA
12	10/27/2024	Oppo	28	Rp 2.200.000	Rp 61.600.000	YA
13	8/21/2024	Samsung	36	Rp 2.500.000	Rp 90.000.000	YA
14	02/07/2024	Oppo	23	Rp 2.200.000	Rp 50.600.000	YA
15	10/11/2024	Samsung	45	Rp 2.500.000	Rp 112.500.000	YA
16	12/03/2024	Samsung	30	Rp 2.500.000	Rp 75.000.000	YA
17	7/22/2024	Oppo	43	Rp 2.200.000	Rp 94.600.000	YA
18	4/25/2024	Samsung	34	Rp 2.500.000	Rp 85.000.000	YA
19	04/08/2024	Oppo	46	Rp 2.200.000	Rp 101.200.000	YA
20	05/01/2024	Samsung	26	Rp 2.500.000	Rp 65.000.000	YA
....	...	...	...	...	...	...
250	1/22/2024	Samsung	1	Rp 2.500.000	Rp 2.500.000	TIDAK

Tabel 3. Data penjualan Handphone Sambung & Oppo

2. Koneksi Data Import dengan Metode Naive Bayes  
pada koneksi di bawah ini kami akan menjelaskan alur proses pada aplikasi *RapidMiner* pada gambar 1 di bawah ini.



Gambar 1. Koneksi Data Import dengan Metode Naive Bayes

Alur Proses:

Proses analisis dan klasifikasi penjualan produk HP menggunakan metode Naive Bayes di RapidMiner melalui beberapa tahap berikut:

1. Retrieve Data Handphone
  - a. Data penjualan diimpor ke RapidMiner menggunakan operator Retrieve. Dataset memuat informasi seperti stok, jumlah penjualan, omset, dan kategori penjualan ("laris" atau "tidak laris").
  - b. Data ini merupakan dasar untuk pelatihan dan evaluasi model klasifikasi.
2. Set Role
  - a. Operator Set Role digunakan untuk menentukan atribut target (label). Dalam konteks ini, kolom KETERANGAN ditetapkan sebagai variabel label, yaitu status "laris" atau "tidak laris".
  - b. Proses ini memastikan model memahami variabel yang akan diprediksi.
3. Split Data
  - a. Operator Split Data membagi dataset menjadi dua bagian:
    - 1) 80% data training: digunakan untuk membangun model.
    - 2) 20% data testing: digunakan untuk menguji kinerja model.



### Confusion Matrix:

1. Prediksi "TIDAK" (Negatif):
  - a. True Negative (TN): 17
  - b. False Negative (FN): 0
2. Prediksi "YA" (Positif):
  - a. True Positive (TP): 33
  - b. False Positive (FP): 0

Kesimpulan Awal: Precision dan Recall pada kedua kelas sempurna (100%).

### KESIMPULAN

Berdasarkan hasil pengujian model Naive Bayes pada dataset penjualan HP menggunakan RapidMiner, diperoleh akurasi 100%, dengan precision dan recall juga mencapai 100%. Hasil ini menunjukkan bahwa model berhasil mengklasifikasikan semua data uji dengan benar tanpa kesalahan. Beberapa poin penting adalah:

#### 1. Efektivitas Model:

Metode Naive Bayes sangat cocok untuk dataset ini, terutama jika fitur yang digunakan relevan dan mendukung klasifikasi.

#### 2. Distribusi Data:

Tidak ada ketidakseimbangan dalam dataset yang digunakan, sehingga model bekerja optimal tanpa bias.

#### 3. Potensi Overfitting:

Hasil sempurna ini dapat mengindikasikan kemungkinan overfitting, terutama jika pola data uji terlalu mirip dengan data latih.

#### 4. Faktor Data:

Kinerja sempurna ini juga menunjukkan bahwa dataset memiliki pola yang jelas, sehingga mudah dipelajari oleh model.

### DAFTAR PUSTAKA

- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann. Referensi utama untuk konsep dasar data mining dan algoritma Naive Bayes.
- Kumar, V., & Poonam. (2018). *Application of Naive Bayes Algorithm in Predictive Analytics: A Review*. *International Journal of Advanced Research in Computer Science*,

J. Anggun Rumboirusi, et al.

9(3), 123-129.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

Patel, H., & Thakur, M. (2016). *A Survey on the Implementation of Data Mining Algorithms Using RapidMiner*. *International Journal of Computer Applications*, 144(8), 22-28.

Pawar, S. V., & Bhalchandra, P. (2019). *Comparative Analysis of Machine Learning Algorithms Using RapidMiner*. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(7), 1455-1459.

RapidMiner Documentation. (n.d.). *Getting Started with RapidMiner*. Retrieved from <https://docs.rapidminer.com>

Rish, I. (2001). *An Empirical Study of the Naive Bayes Classifier*. In *IJCAI 2001 Workshop on Empirical Methods in AI* (pp. 41-46).

Statology. (2021). *Naive Bayes Explained*. Retrieved from <https://www.statology.org>

Towards Data Science. (2020). *Understanding Naive Bayes Classifier*. Retrieved from <https://towardsdatascience.com>

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Morgan Kaufmann.