

## THE PERFORMANCE OF THE NAIVE BAYES ALGORITHM IN PREDICTING CUSTOMER PRODUCT PURCHASE POTENTIAL

Muhammad Syahdan Junus<sup>1</sup>, Yuana Eka Putra Yuha<sup>2</sup>, Elkin Rilvani<sup>3</sup>

<sup>1,2,3</sup>Universitas Pelita Bangsa

[syahdanjns26@gmail.com](mailto:syahdanjns26@gmail.com)

### Abstrak

Penelitian ini mengevaluasi kinerja algoritma Naive Bayes dalam memprediksi potensi pembelian produk pelanggan. Dalam lanskap bisnis yang kompetitif, kemampuan untuk mengidentifikasi pelanggan dengan probabilitas tinggi untuk melakukan pembelian sangat krusial bagi strategi pemasaran yang efektif. Hasil penelitian menunjukkan bahwa Naive Bayes merupakan salah satu pendekatan yang umum diterapkan dalam teknik data mining. Metode ini berlandaskan pada prinsip Teorema Bayes, dengan asumsi bahwa setiap atribut dalam data bersifat independen satu sama lain. Meskipun demikian, identifikasi batasan dan area perbaikan, seperti penanganan data yang hilang atau fitur yang berkorelasi tinggi, menjadi fokus untuk pengembangan lebih lanjut.

**Kata Kunci:** Naive Bayes, Prediksi Pembelian, Potensi Pelanggan, Pembelajaran Mesin, Klasifikasi.

### Abstract

*This study evaluates the performance of the Naive Bayes algorithm in predicting potential customer product purchases. In a competitive business landscape, the ability to identify customers with a high probability of making a purchase is crucial for an effective marketing strategy. The results show that Naive Bayes is capable of classifying potential buyers with reasonable accuracy, demonstrating its potential as a valuable predictive tool. However, identifying limitations and areas for improvement, such as handling missing data or highly correlated features, are areas for further development.*

**Keywords:** Naive Bayes, Purchase Prediction, Customer Potential, Machine Learning, Classification.

## PENDAHULUAN

Di era digital saat ini, volume data yang sangat besar (big data) telah menjadi aset berharga bagi organisasi di berbagai sektor, termasuk ritel dan e-commerce. Pemanfaatan data ini secara efektif dapat memberikan wawasan mendalam mengenai perilaku konsumen, tren pasar, dan peluang bisnis. Kemampuan untuk mengidentifikasi pelanggan yang memiliki probabilitas tinggi untuk melakukan pembelian memungkinkan perusahaan untuk merumuskan

strategi pemasaran yang lebih terfokus, mengoptimalkan alokasi sumber daya, dan pada akhirnya meningkatkan pendapatan. Algoritma Naive Bayes dibangun di atas Teorema Bayes dengan anggapan bahwa tiap fitur saling bebas, menjadikannya metode yang efisien untuk digunakan dalam pembuatan model prediktif.

Studi ini telah melakukan evaluasi terhadap efektivitas algoritma Naive Bayes dalam memprediksi kemungkinan pelanggan melakukan pembelian produk. Kami akan memanfaatkan RapidMiner, sebuah perangkat lunak terkemuka untuk data mining dan pembelajaran mesin, sebagai platform implementasi. RapidMiner menawarkan antarmuka berbasis visual yang memudahkan pengguna dalam melakukan pra-pemrosesan, membangun model, serta mengevaluasi kinerjanya, sehingga mendukung proses analisis yang efisien. Teknik datamining yang digunakan adalah Algoritma Naive Bayes.

Diharapkan penelitian ini dapat memberikan masukan kepada pihak yang berkepentingan agar dapat mengetahui strategi penjualan produk. Selamat Selular yang tepat dengan menggunakan teknik datamining. RapidMiner adalah perangkat lunak mandiri yang dirancang untuk analisis data serta dapat digunakan sebagai platform data mining terintegrasi dalam berbagai sistem. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

Naive Bayes adalah metode klasifikasi yang memanfaatkan konsep probabilitas, dikembangkan berdasarkan teori dari Thomas Bayes, seorang ilmuwan asal Inggris. Yaitu memprediksi peluang di masa yang akan datang berdasarkan pengalaman di masa sebelumnya (lampau) sehingga dikenal sebagai teorema Bayes. Metode Naive Bayes merupakan komponen dasar yang digunakan dalam implementasi pembuatan perangkat lunak yang terdapat dalam sistem operasi. Dalam proses pengambilan keputusan, data historis yang tersedia dimanfaatkan untuk mengekstraksi informasi penting yang relevan. hal ini menjadikan munculnya cabang ilmu baru untuk mengatasi masalah penggalian informasi atau pola yang penting atau menarik dari data dalam jumlah yang besar, yang disebut dengan data mining. Penggunaan teknik data mining diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya.

## **METODE PENELITIAN**

### **Data Mining**

Pemilihan tugas data mining; pemilihan goal dari proses KDD misalnya klasifikasi, regresi, clustering, dll. Pemilihan algoritma data mining untuk pencarian (searching), Proses

Data mining yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Penentuan algoritma yang digunakan dalam data mining sangat ditentukan oleh sasaran serta tahapan dalam proses KDD secara menyeluruh.

Data mining merujuk pada proses eksplorasi dan ekstraksi pengetahuan yang tersembunyi dari kumpulan data berskala besar, dengan tujuan utama menemukan pola atau hubungan yang dapat dimanfaatkan untuk pengambilan keputusan. Knowledge Discovery in Databases (KDD) adalah penerapan metode saintifik pada data mining. Beberapa metode data mining, antara lain: Classification [Predictive], Clustering [Descriptive], Association Rule Discovery [Descriptive], Regression [Predictive], Deviation Detection [Predictive]

## Klasifikasi

Klasifikasi merupakan bagian dari proses pembelajaran mesin yang bertujuan mengaitkan sekumpulan atribut pada suatu objek dengan kategori tertentu. atau dengan arti lain bahwa klasifikasi adalah suatu proses pengelompokan data. Pengelompokan adalah penataan objek ke dalam kategori-kategori yang memiliki kegunaan yang sama. Pengelompokan digunakan untuk menggambarkan kumpulan data di mana setiap tipe data adalah nominal atau biner. Saat mengklasifikasikan data yang dipantau, itu dibagi menjadi dua entitas, yaitu entitas coba dan entitas uji, dimana entitas latih dikelola menggunakan metode data mining.

## Algoritma Naïve Bayes

Naïve Bayes tidak mengikuti aturan berbasis logika, namun bekerja dengan prinsip-prinsip probabilitas untuk menghitung kemungkinan suatu kategori berdasarkan data yang ada. Konsep ini digunakan untuk menemukan probabilitas tertinggi dari proses klasifikasi yang dilakukan dengan melakukan pengamatan terhadap kemunculan setiap klasifikasi dalam data pelatihan.

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Di mana:

Dalam konteks Naive Bayes,  $P(C|X)$  merepresentasikan peluang suatu kelas  $C$  berdasarkan data input  $X$ , dan menjadi fokus utama dalam perhitungan prediktif.

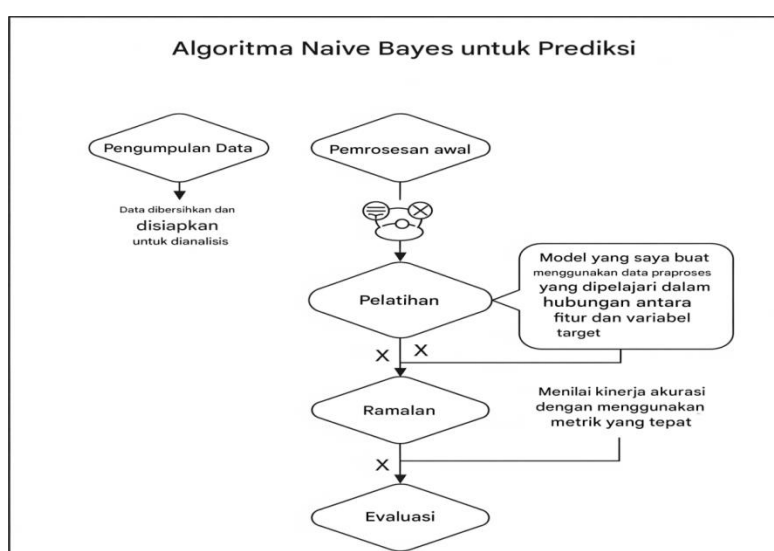
$P(X|C)$  adalah probabilitas likelihood dari observasi  $X$  diberikan kelas  $C$ .

$P(C)$  adalah probabilitas prior dari kelas  $C$ .

$P(X)$  adalah probabilitas prior dari observasi  $X$ .

### Flowchart Proses Algoritma Naïve Bayes

Dari rancangan Flowchart Algoritma Naïve Bayes diatas, dapat dijelaskan bahwa, peneliti mulai dengan menyiapkan dataset, identifikasi dan preprocessing dataset, membaca input, bobot, jumlah layer, target error dan training cycle yang diinginkan, hitung hasil bobot tiap fitur (input), nilai yang terbesar merupakan kesimpulan (kelas) dan selesai.



**Gambar 1. Tampilan Flowchart Proses Algoritma Naïve Bayes**

### Rapid Miner

RapidMiner merupakan salah satu aplikasi yang digunakan sebagai mengolah data mining. Pengolahan dilakukan pada RapidMiner text mining menganalisis teks, memisahkan pola dari data set yang besar dan menggabungkan dengan cara statistika, kecerdasan buatan, dan database. RapidMiner memiliki prosedur data mining dan machine learning, di dalamnya terdapat: ETL (extraction, transformation, loading), data visualisasi, preprocessing, modelling dan evaluasi. Pengolahan data mining terbentuk dari banyak operator yang nestable, dijelaskan dengan XML, dan digunakan dengan GUI. Tersaji dalam bahasa pemrograman Java. Rapidminer mengimplementasikan algoritme pembelajaran yang di implementasikan ke lembar entitas dari baris perintah. RapidMiner menyediakan berbagai fitur, termasuk untuk pra-pemrosesan data, klasifikasi, clustering, regresi, integrasi sistem, serta visualisasi hasil analisis.

Rapidminer dapat memproses entitas, mengintegrasikannya ke dalam rencana pembelajaran, dan menganalisis pengklasifikasi yang dihasilkan serta kerjanya.

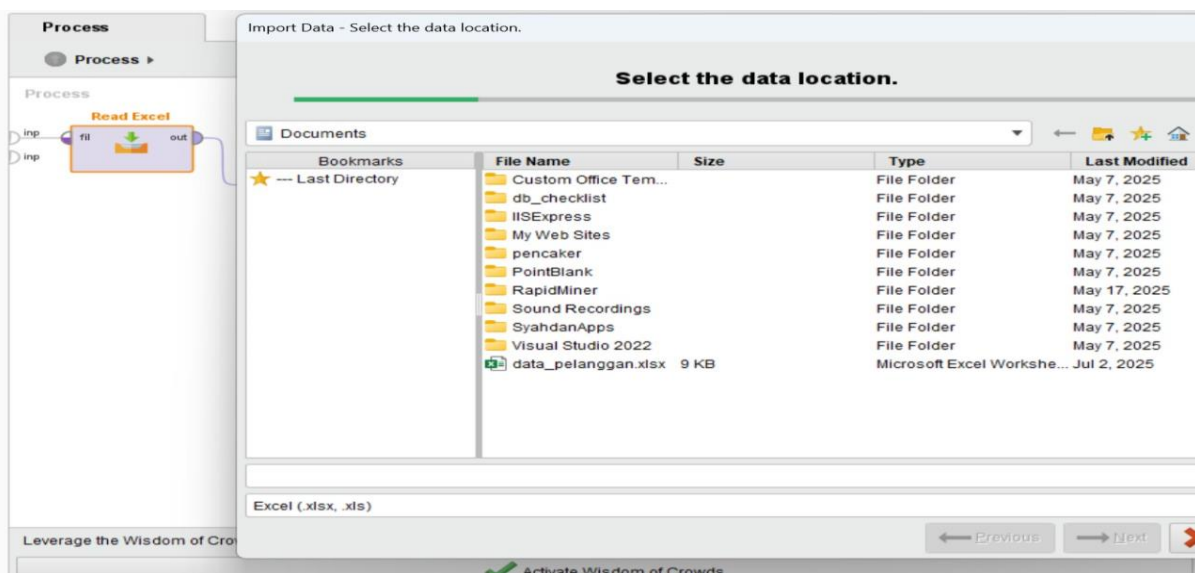
### Sumber Data

- 1) **Data primer** diperoleh dari hasil kegiatan pengumpulan data terhadap data transaksi penjualan.
- 2) **Data sekunder** dalam penelitian ini melakukan pengolahan data dari data laporan penjualan untuk menentukan kriteria status usaha, jumlah pembelian, dan lokasi. Selain itu untuk menunjang data pendukung yang diperoleh dari media pustaka tentang teori-teori yang digunakan dalam penelitian ini. Adapun data sekunder yang diantaranya Data Mining, Klafikasi, Naïve Bayes, MySQL, PHP, UML, Kajian Pustaka.

## HASIL DAN PEMBAHASAN

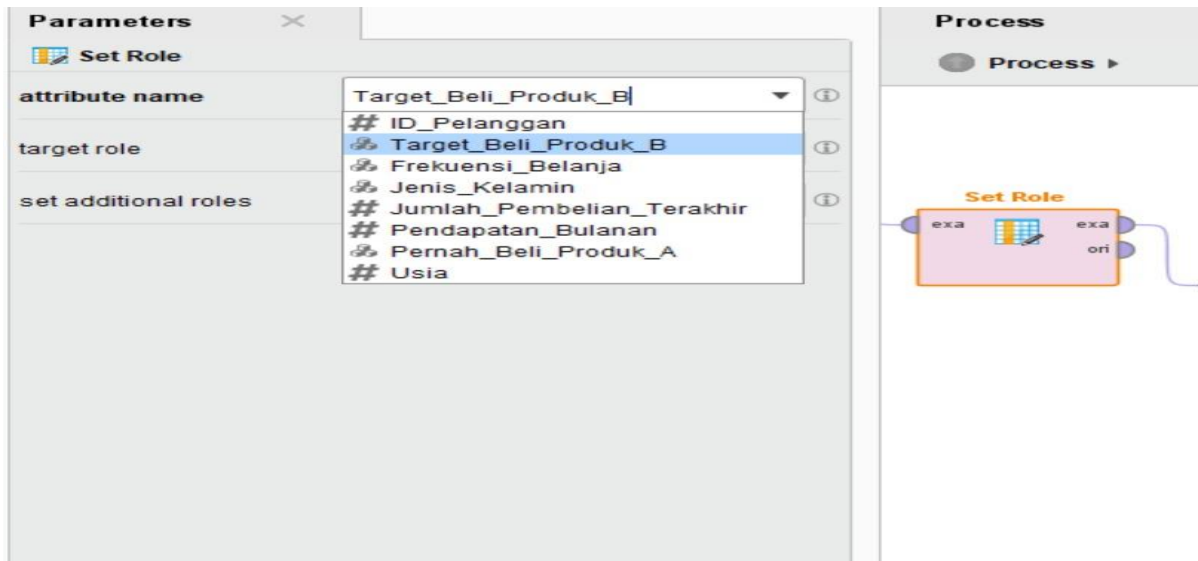
### Implementasi Dengan Rapidminer

Berikut adalah proses pengujian dengan metode naïve bayes yang menggunakan aplikasi rapid miner:



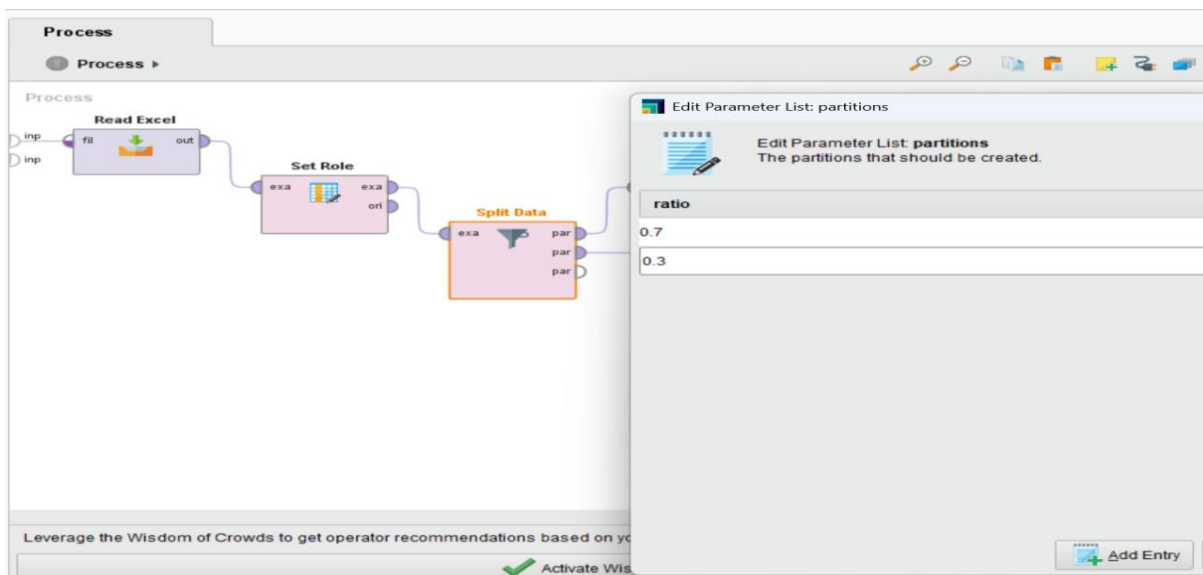
Gambar 1. Tampilan Pemilihan Data

File data\_pelanggan.xlsx dipilih sebagai sumber data utama. Tahap ini krusial untuk memastikan data yang benar diimpor ke dalam proses, menjadi fondasi bagi seluruh analisis data mining yang akan dilakukan selanjutnya.



**Gambar 2.** Tampilan Pengaturan Peran Atribut

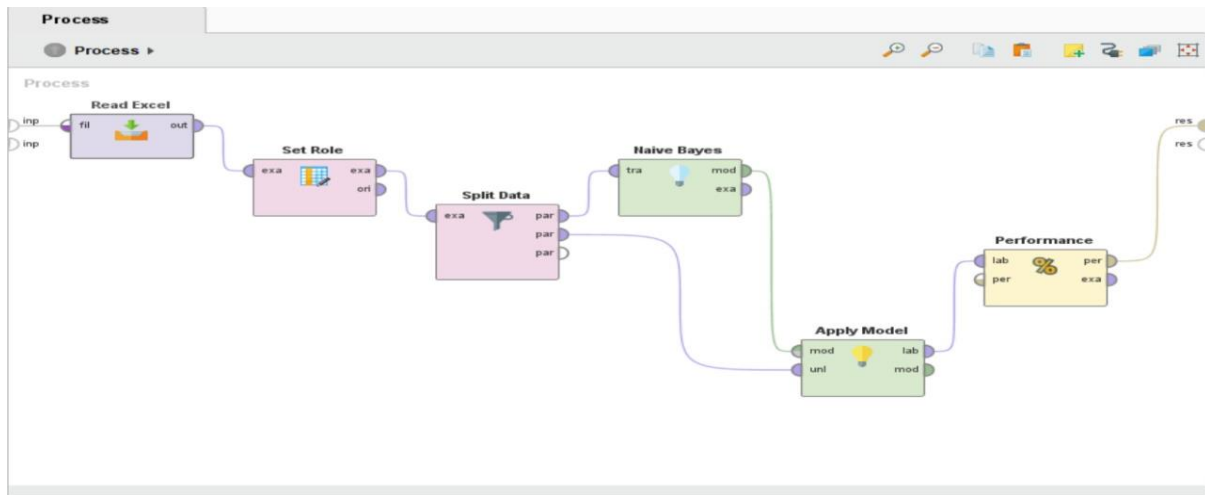
Atribut "Target\_Beli\_Produk\_B" dipilih sebagai target role atau label, yang akan diprediksi oleh model Naive Bayes. Pengaturan peran ini sangat penting agar RapidMiner memahami variabel dependen dan independen dalam dataset.



**Gambar 3.** Tampilan Pembagian Data

Data dibagi menjadi dua partisi: 70% untuk data latih (training) dan 30% untuk data uji (testing). Pembagian ini memastikan bahwa model dilatih pada satu set data dan dievaluasi

secara independen pada set data yang belum pernah dilihat, mengukur kemampuan generalisasi model.



**Gambar 4.** Alur Proses Data Mining Naïve Bayes

Dimulai dengan Read Excel, diikuti oleh Set Role untuk mendefinisikan target, lalu Split Data. Model Naive Bayes dilatih dan kemudian Apply Model digunakan pada data uji. Terakhir, Performance mengevaluasi hasil prediksi.

Result History PerformanceVector (Performance) x

Criterion: accuracy

accuracy: 80.00%

	true Tidak	true Ya	class precision
pred. Tidak	2	0	100.00%
pred. Ya	1	2	66.67%
class recall	66.67%	100.00%	

**Gambar 5.** Tampilan Hasil Kinerja Model

Akurasi model tercatat sebesar 80.00%. Matriks konfusi menunjukkan jumlah prediksi benar dan salah untuk setiap kelas ("Tidak" dan "Ya"), serta nilai presisi dan recall. Ini memberikan gambaran komprehensif tentang seberapa baik model memprediksi potensi pembelian.

## KESIMPULAN

Penelitian ini telah berhasil mengevaluasi kinerja algoritma Naive Bayes dalam memprediksi potensi pembelian produk pelanggan. Hasil implementasi menggunakan RapidMiner menunjukkan bahwa model Naive Bayes mampu mencapai akurasi sebesar 80.00% dalam mengklasifikasikan calon pembeli. Tingkat akurasi yang diperoleh menunjukkan bahwa Naive Bayes merupakan salah satu pendekatan prediktif yang layak dan bernilai untuk diterapkan dalam mengidentifikasi pelanggan dengan probabilitas tinggi untuk melakukan pembelian, mendukung strategi pemasaran yang lebih terfokus dan efisien.

Langkah-langkah penelitian ini mencakup pengumpulan informasi pelanggan, tahap pembersihan data, pemisahan antara data latih dan uji, pelatihan model klasifikasi Naive Bayes, serta analisis performanya. Pengaturan peran atribut "Target\_Beli\_Produk\_B" sebagai variabel target sangat krusial dalam proses ini. Meskipun demikian, identifikasi batasan dan area perbaikan, seperti penanganan data yang hilang atau fitur yang berkorelasi tinggi, tetap menjadi fokus untuk pengembangan lebih lanjut. Secara keseluruhan, studi ini menyimpulkan bahwa Naive Bayes menawarkan solusi yang praktis dan efisien untuk membantu perusahaan dalam merumuskan strategi penargetan pelanggan yang lebih cerdas dan berbasis data.

## DAFTAR PUSTAKA

- S. M. Dewi, A. P. Windarto, and D. Hartama, "Penerapan Datamining Dengan Metode Klasifikasi Untuk Strategi Penjualan Produk Di Ud.Selamat Selular," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 617–621, 2019, doi: 10.30865/komik.v3i1.1669.
- N. W. Wardani, P. G. S. C. Nugraha, and G. S. Mahendra, "Implementasi Naïve Bayes Pada Data Mining Untuk Mengklasifikasikan Penjualan Barang Terlaris Pada Perusahaan Ritel," *JST (Jurnal Sains dan Teknol.)*, vol. 12, no. 3, pp. 656–668, 2024, doi: 10.23887/jstundiksha.v12i3.38605.
- N. Ayuningtyas, N. R., and F. M. Basysyar, "Penerapan Data Mining pada Penjualan Produk MS Glow Menggunakan Metode Naive Bayes untuk Strategi Pemasaran," *J. Account. Inf. Syst.*, vol. 5, no. 2, pp. 157–166, 2022, doi: 10.32627/aims.v5i2.503.
- Y. Niar, K. Komariah, A. Surip, R. Saputra, and I. Ali, "Implementasi Algoritma Naïve Bayes Untuk Prediksi Persediaan Barang Rotan," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 4, no. 1, pp. 28–34, 2022, doi: 10.32485/kopertip.v4i1.112.

- Susanto, E. Dewi Sri Mulyani, and I. Ratnasari Nurhasanah, "Penerapan Data Mining Classification Untuk Prediksi Perilaku Pola Pembelian Terhadap Waktu Transaksi Menggunakan Metode Naïve Bayes," *Konf. Nas. Sist. Inform. 2015*, pp. 313–318, 2015.
- K. Subang and M. Algoritma, "Penerapan Data Mining Pada Penjualan Barang Menggunakan Metode Naive Bayes Classifier untuk Optimasi Strategi Pemasaran," *J. Teknol. Inf. dan Komun.*, vol. 11, no. April, pp. 1–12, 2018.
- R. A. Anggraini, G. Widagdo, A. S. Budi, and M. Qomaruddin, "Penerapan Data Mining Classification untuk Data Blogger Menggunakan Metode Naïve Bayes," *J. Sist. dan Teknol. Inf.*, vol. 7, no. 1, p. 47, 2019, doi: 10.26418/justin.v7i1.30211.
- P. Ajun Isawal, "Penerapan Algoritma Naive Bayes Untuk Menentukan Produk Terlaris Di Outlet Penjualan Pulsa 'Athalla Cell,'" *J. Inform. SIMANTIK*, vol. 6, no. 2, pp. 67–73, 2021.
- S. M. Tafonao, "Implementasi Algoritma Naïve Bayes Untuk Mengetahui Minat Beli Konsumen Terhadap Sarang Burung Walet," *JIKTEKS J. Ilmu Komput. dan Teknol. Inf.*, vol. 2, no. 03, pp. 11–19, 2024, doi: 10.70404/jikteks.v2i03.108.
- S. Rokhanah, A. Hermawan, and D. Avianto, "Pengaruh Principal Component Analysis Pada Naïve Bayes dan K-Nearest Neighbor Untuk Prediksi Dini Diabetes Melitus Menggunakan Rapidminer," *EVOLUSI J. Sains dan Manaj.*, vol. 11, no. 1, 2023, doi: 10.31294/evolusi.v11i1.14728.
- D. I. Mulyana and A. M. Fadhil, "Penerapan Data Mining untuk Menentukan Produk Terlaris pada PT Joynare Mitra Teknologi menggunakan Metode Naive Bayes," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 2, pp. 1206–1222, 2024, doi: 10.35870/jimik.v5i2.638.
- A. Julianto and S. Andayani, "Penerapan Data Mining Untuk Klasifikasi Produk Terlaris Menggunakan Algoritma Naive Bayes Pada Bengkel Motor," *JuSiTik J. Sist. dan Teknol. Inf. Komun.*, vol. 7, no. 2, pp. 50–58, 2024, doi: 10.32524/jusitik.v7i2.1148.
- E. Miranda and Julisar, "Data MINING DENGAN METODE KLASIFIKASI NAÏVE BAYES UNTUK MENGLASIFIKASIKAN PELANGGAN Eka Miranda , Julisar Program Sistem Informasi , Program Studi Sistem Informasi , Universitas Bina Nusantara," *Infotech*, vol. 4, no. 9, pp. 6–12, 2018.

- G. Asri, F. Doni, G. Putri Asri, and F. Roma Doni, “Klasifikasi Produk Persediaan pada PT HMS Kompresindo Sukses Menggunakan Algoritma Naive Bayes,” *J. Inform. Multimed.*, vol. 16, no. 2, pp. 1–9, 2024.
- H. F. Putro, R. T. Vlandari, and W. L. Y. Saptomo, “Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan,” *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, 2020, doi: 10.30646/tikomsin.v8i2.500.