

Analisis Data Mining dengan Naive Bayes Classifier untuk Peningkatan Penjualan Produk

Diterima:

10 Juni 2024

¹ Cinta Azzaria, ² Mieta Silvia Aviva, ³ Ela Esti Susanti, ⁴ Erna Daniati, ⁵ Aidina Ristyawan

Revisi:

10 Juli 2024

¹⁻³ Universitas Nusantara PGRI Kediri

Terbit:

1 Agustus 2024

¹ twentyn22@gmail.com, ² mieta.via.av@gmail.com, ³ elaesti366@gmail.com,⁴ ernadaniati@unpkediri.ac.id, ⁵ adinaristi@unpkediri.ac.id

Abstrak— Segala bidang kehidupan sekarang dipengaruhi oleh kemajuan teknologi informasi yang semakin pesat. Kemajuan ini memungkinkan jumlah data yang sangat besar dan beragam diakses dari berbagai bidang kehidupan, termasuk industri, ekonomi, ilmu pengetahuan, dan teknologi. Banyak hal yang mendorong kemajuan luar biasa dalam data mining. Tujuan dari dilakukan penelitian ini adalah untuk memudahkan mengetahui hasil prediksi dengan metode klasifikasi menerapkan data mining pada data penjualan dan transaksi penjualan dan mengetahui sejauh mana algoritma naive bayes untuk menghasilkan informasi atau pengetahuan baru yang dapat membantu peningkatan penjualan produk yang baik, efisien, dan efektif dalam menghadapi pesaing bisnis lainnya dan menambah omset penjualan untuk tiap bulannya. Berdasarkan hasil Analisa pada penelitian yang dilakukan, didapatkan hasil prediksi dalam menentukan penjualan produk, dari pengujian yang dilakukan dengan membandingkan data training dengan data testing menggunakan tools RapidMiner menunjukkan bahwa algoritma Naive Bayes menghasilkan akurasi yang sangat tinggi, terbukti dengan akurasi sebesar 98,98%.

Kata Kunci—Teknologi; Peningkatan Penjualan Produk; Naive Bayes

Abstract— All areas of life are now influenced by increasingly rapid advances in information technology. These advances enable enormous and diverse amounts of data to be accessed from many areas of life, including industry, economics, science, and technology. There are many things that have driven extraordinary progress in data mining. The purpose of this research is to make it easier to find out prediction results using classification methods applying data mining to sales data and sales transactions and to find out to what extent the Naive Bayes algorithm can produce new information or knowledge that can help development of good, efficient and effective product promotion strategies in dealing with other business competitors and increasing sales turnover for each month. Based on the results of the analysis of the research carried out, prediction results were obtained in determining promotional strategies. From tests carried out by comparing training data with data testing using the RapidMiner tool shows that the Naive Bayes algorithm produces very high accuracy, proven by an accuracy of 98.98%.

Keywords—Technology; increase in Product Sales; Naive Bayes

This is an open access article under the CC BY-SA License.



Penulis Korespondensi:

Erna Daniati,

Sistem Informasi,

Universitas Nusantara PGRI Kediri,

Email: ernadaniati@unpkediri.ac.idID Orcid: [<https://orcid.org/0009-0008-9471-4421>]

Handphone: 0813-3524-2202

I. PENDAHULUAN

Segala bidang kehidupan sekarang dipengaruhi oleh kemajuan teknologi informasi yang semakin pesat. Kemajuan ini memungkinkan jumlah data yang sangat besar dan beragam diakses dari berbagai bidang kehidupan, termasuk industri, ekonomi, ilmu pengetahuan, dan teknologi [1]. Banyak hal yang mendorong kemajuan luar biasa dalam data mining (Larose, 2005). Salah satunya adalah pertumbuhan kumpulan data yang cepat, serta kemajuan besar dalam kemampuan komputasi dan kapasitas media penyimpanan[2]. Dalam dunia bisnis, terutama dalam industri penjualan, banyak persaingan, sehingga pengembang harus menemukan cara untuk meningkatkan penjualan dan pemasaran produk mereka. Salah satu cara untuk melakukan ini adalah dengan menggunakan data penjualan harian yang semakin meningkat. Data ini tidak hanya berguna sebagai catatan bagi perusahaan, tetapi juga dapat diolah menjadi informasi yang berguna untuk peningkatan penjualan dan strategi promosi produk[3].

Menurut Tony Suryaudin (2018) Data mining merupakan suatu proses penggalian data atau penyaringan data dengan memanfaatkan kumpulan data dengan ukuran yang cukup besar melalui serangkaian proses untuk mendapatkan informasi yang berharga dari data tersebut. Data mining bisa digunakan oleh perusahaan besar untuk menggali data untuk mendapatkan informasi yang dapat menunjang dan meningkatkan proses bisnis perusahaan tersebut[4] [5].

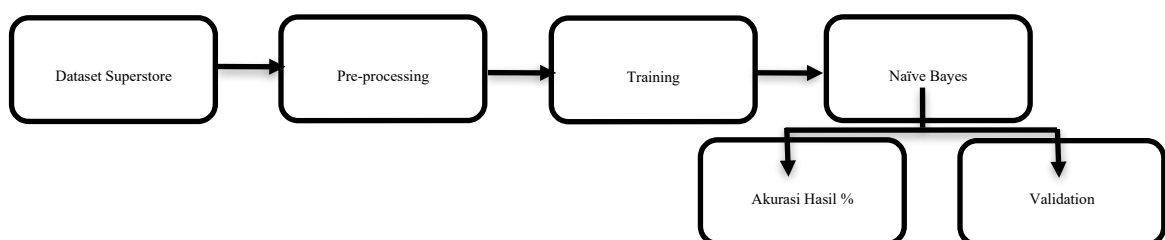
Peningkatan penjualan merupakan jumlah penjualan bersih yang tercatat dalam laporan laba perusahaan, yang mencerminkan seluruh transaksi penjualan selama periode tertentu yang mengalami kenaikan. Dalam konteks ini, peningkatan penjualan menunjukkan kinerja perusahaan yang berhasil meningkatkan pendapatannya dari waktu ke waktu. Laporan laba perusahaan mencatat data ini secara rinci, memungkinkan manajemen untuk menganalisis tren penjualan dan merumuskan strategi yang tepat untuk mempertahankan atau bahkan meningkatkan pertumbuhan tersebut di masa mendatang. Dengan demikian, pemahaman yang mendalam mengenai peningkatan penjualan sangat penting bagi keberhasilan jangka panjang perusahaan[6].

Tujuan dari dilakukan penelitian ini adalah untuk memudahkan mengetahui hasil prediksi dengan metode klasifikasi menerapkan data mining pada data penjualan dan transaksi penjualan dan mengetahui sejauh mana algoritma *naive bayes* untuk menghasilkan informasi atau pengetahuan baru yang dapat membantu peningkatan penjualan produk yang baik, efisien, dan efektif dalam menghadapi pesaing bisnis lainnya dan menambah omset penjualan untuk tiap bulannya.

II. METODE

Metode yang digunakan dalam penelitian ini adalah *Naive Bayes Classifier*. Klasifikasi merupakan bentuk dasar dari analisis data yang mempunyai teknik untuk menentukan keanggotaan kelompok berdasarkan data-data yang sudah ada [7]. Konsep dasar dari klasifikasi adalah beberapa data yang memiliki struktur data yang mirip akan memiliki klasifikasi yang mirip pula. *Naive Bayes Classifier* (NBC) merupakan pengklasifikasi probabilitas sederhana untuk menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari kumpulan data yang diberikan[8][9]. Ciri utama *Naive Bayes Classifier* adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi/kejadian (Harahap et al. 2021)[10].

Metode data mining pada pengklasifikasian dataset Superstore dilakukan dengan beberapa tahapan. Adapun alur proses pengklasifikasian dapat dilihat dibawah ini :



Gambar 1. Alur Proses Klasifikasi Naive Bayes

Adapun alur data mining menggunakan metode Naive Bayes Classifier sebagai berikut :

1. Dataset Superstore

Dataset superstore terdiri dari beberapa atribut antara lain *Row ID, Order ID, Order Date, Customer ID, Customer Name, City, State, Region, Product ID, Category, Sub-Category, Product Name, Quantity Date*[11].

2. Pre-processing

Langkah selanjutnya adalah prapemrosesan, yang bertujuan untuk menghilangkan data yang tidak sempurna dan dapat menyebabkan inkonsistensi. *Pre-processing* dilakukan untuk memastikan data siap digunakan dengan menghapus elemen yang dianggap tidak relevan / *Missing value* [12]. *Missing value* adalah informasi yang tidak tersedia untuk sebuah objek (kasus). *Missing value* terjadi karena informasi untuk sesuatu tentang objek tidak diberikan, sulit dicari atau memang informasi tersebut tidak ada. *Missing value* pada dasarnya tidak bermasalah bagi keseluruhan data, apalagi jika jumlahnya hanya sedikit, misal hanya 1% dari seluruh data. Namun jika persentase data yang hilang tersebut cukup besar, maka perlu dilakukan pengujian apakah data yang mengandung banyak *missing* tersebut masih layak diproses lebih lanjut ataukah tidak [13]. Dengan demikian, data yang diproses tidak akan mengalami kesalahan atau masalah. Selain itu, *pre-processing* juga melakukan penghapusan atribut yang tidak digunakan pada tahap selanjutnya, sehingga hanya atribut yang relevan yang tersisa dan digunakan pada tahap modeling.

3. *Naïves Bayes Classifier*

Pada tahap ini, tools *RapidMiner* studio digunakan untuk melakukan proses pengklasifikasian data pelatihan terhadap dataset Superstore yang sudah diuji menggunakan model *Classifier* yang dibentuk.

4. Akurasi Hasil

Pengertian Akurasi adalah ukuran seberapa dekat hasil pengukuran dengan nilai sebenarnya dari suatu angka [14] Akurasi hasil menunjukkan seberapa jauh hasil pengukuran berbeda dengan nilai sebenarnya.

5. *Validation*

Teknik validasi model untuk menilai keakuratan hasil analisis melibatkan proses *cross-validation*, di mana data yang sudah diproses dibagi menjadi dua bagian: data latih dan data uji. Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk menguji kinerja model dalam proses klasifikasi [15].

III. HASIL DAN PEMBAHASAN

Dataset yang digunakan untuk penelitian ini adalah dataset Superstore yang dapat diunduh di website *kaggle.com* dengan data sebanyak 9667 data. Di bawah ini adalah langkah-langkah untuk menerapkan Algoritma *Naive Bayes* :

1. *Pre-processing Data*

a. *Cleaning data*

Pada penelitian ini dataset masih memiliki variabel yang tidak digunakan atau *missing value* sehingga harus dilakukan *pre-processing* agar ketika diproses tidak muncul masalah atau kesalahan. Pada proses ini, awalnya data berjumlah 9.667 selanjutnya dilakukan penghapusan duplikasi yang ternyata data berubah hasilnya menjadi 9.632. Penghapusan atribut.

b. Penghapusan atribut

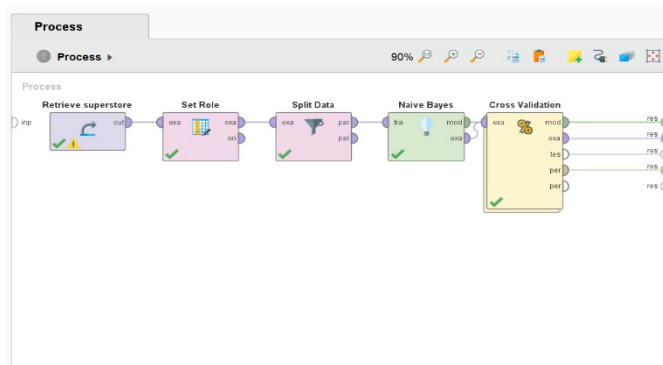
Pada tahap ini, akan dihapus beberapa atribut yang tidak digunakan sehingga hanya atribut yang relevan yang akan digunakan ditahap modeling. Berikut adalah daftar atribut setelah dilakukan penghapusan :

- a. Row ID berfungsi untuk identifikasi unik untuk setiap baris data, penting untuk pelacakan dan pengelolaan data.
- b. Order ID berfungsi untuk identifikasi unik untuk setiap pesanan, penting untuk analisis pola pembelian.
- c. Order Date berfungsi untuk analisis tren waktu, musim, dan pola pembelian.
- d. Customer ID berfungsi untuk identifikasi unik pelanggan, penting untuk segmentasi dan analisis perilaku pelanggan.

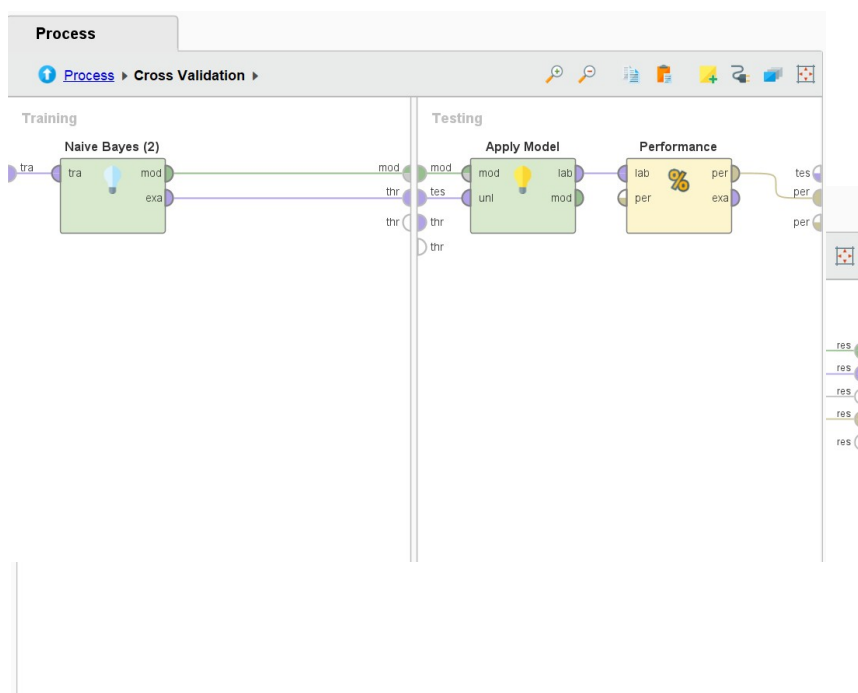
- e. Customer Name berguna untuk personalisasi dan analisis hubungan pelanggan.
- f. Category berfungsi untuk mengelompokkan produk, membantu memahami preferensi berdasarkan kategori.
- g. Sub-Category berfungsi untuk memberikan rincian lebih lanjut tentang jenis produk, membantu dalam analisis sub-segmen.
- h. Product ID berfungsi untuk identifikasi unik untuk setiap produk, penting untuk analisis kinerja produk.
- i. Product Name berfungsi untuk memberikan konteks tambahan, berguna dalam pelaporan dan visualisasi data.
- j. Quantity berfungsi untuk analisis penjualan dan permintaan produk, penting untuk peramalan inventaris

2. Kinerja Klasifikasi

Tujuan pengujian ini adalah untuk mengenali hasil klasifikasi, menilai performa Algoritma *Naive Bayes Classifier* dalam mengklasifikasikan data ke dalam kelas yang sudah ditentukan, serta memilih model terbaik untuk prediksi. Data diacak dengan menggunakan random seed, lalu dibagi menjadi data training dan data testing melalui metode *K-Fold Cross Validation*. Adapun prosesnya dapat dilihat pada Gambar 2 dan untuk memperoleh akurasi dari dataset penjualan yang telah melalui proses klasifikasi, peneliti menggunakan operator cross validation dapat dilihat pada Gambar 3.



Gambar 2. Processing data menggunakan tools RapidMiner



Gambar 3. Proses *Cross Validation*

Setelah dilakukan proses-proses diatas, pada Gambar 4 menunjukkan bahwa algoritma Naïve Bayes menghasilkan akurasi yang sangat tinggi, terbukti dengan akurasi sebesar 98,98%. Ini menandakan bahwa model Naïve Bayes sangat efektif dalam mengklasifikasikan data penjualan.

Table View Plot View

accuracy: 98.98% +/- 0.36% (micro average: 98.98%)

	true Furniture	true Office Supplies	true Technology	class precision
pred. Furniture	1338	9	16	98.17%
pred. Office Supplies	11	4102	17	99.32%
pred. Technology	9	7	1258	98.74%
class recall	98.53%	99.61%	97.44%	

Gambar 4. Hasil Akurasi

Pada Gambar 4 menunjukkan bahwa hasil jumlah precision "*Furniture*" sebesar 98,17% dan jumlah recallnya sebesar 98,53%, jumlah precision "*Office Supplies*" sebesar 99,32% dan jumlah recallnya sebesar 99,61%, jumlah precision "*Technology*" sebesar 96,74% dan jumlah recallnya sebesar 97,44%.

IV. KESIMPULAN

Berdasarkan hasil dan pembahasan, dapat disimpulkan bahwa proses klasifikasi berdasarkan atribut Category menghasilkan pendekatan yang optimal untuk meningkatkan penjualan produk, yaitu: penjualan "Teknologi" menunjukkan minat yang lebih rendah dengan hanya 1258 transaksi dibandingkan dengan kategori lainnya. Oleh karena itu, pendekatan untuk meningkatkan penjualan produk akan difokuskan pada penjualan "Teknologi" dengan memberikan diskon, penawaran khusus, dan bundling produk yang diharapkan dapat merangsang peningkatan penjualan produk tersebut..

UCAPAN TERIMA KASIH

Kami ingin mengucapkan rasa terima kasih yang tulus kepada semua pihak yang telah turut serta dalam mendukung dan memfasilitasi penyelesaian penelitian ini. Pertama-tama, kami ingin mengungkapkan penghargaan yang dalam kepada Universitas Nisantara PGRI Kediri atas fasilitas dan dukungan administratif yang diberikan. Kami juga ingin mengucapkan terima kasih kepada Bapak/Ibu Dosen Prodi Sistem Informasi atas bimbingan dan arahan yang berharga selama proses penelitian. Serta kepada seluruh anggota tim penelitian yang telah berkontribusi dengan semangat kerjasama dan dedikasi yang tinggi.

DAFTAR PUSTAKA

- [1] "PERAN TEKNOLOGI INFORMASI DAN KOMUNIKASI DALAM PENDIDIKAN", doi: HYPERLINK "http://dx.doi.org/10.24042/atjpi.v8i1.2095" <http://dx.doi.org/10.24042/atjpi.v8i1.2095>.
- [2] R. Alhapizi, M. Nasir, and I. Effendy, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang," 2020. doi: 10.51519/journalsea.v1i1.10.
- [3] S. Kanti *et al.*, "IMPLEMENTASI DATA MINING PENJUALAN HANDPHONE OPPO STORE SDC TANGGERANG DENGAN ALGORITMA APPRIORI."
- [4] S. Rahmatullah *et al.*, "DATA MINING UNTUK MENENTUKAN PRODUK TERLARIS MENGGUNAKAN METODE NAIVE BAYES." doi: <https://doi.org/10.35959/jik.v7i2.150>.
- [5] Y. Asriningtias and R. Mardiyah, "APLIKASI DATA MINING UNTUK MENAMPILKAN INFORMASI TINGKAT KELULUSAN MAHASISWA," 2014. doi: <http://dx.doi.org/10.26555/jifo.v8i1.a2082>.
- [6] "ANALISIS PERSEPSI HARGA, KUALITAS PELAYANAN, CUSTOMER RELATIONSHIP MARKETING, DAN KEPERCAYAAN TERHADAP PENINGKATAN PENJUALAN DIMEDIASI LOYALITAS PELANGGAN PADA UMKM AYAM POTONG ONLINE ELMONSU", doi: <https://doi.org/10.36778/jesyva.v3i1.66>.
- [7] P. B. N. Setio, D. R. S. Saputro, and B. Winarno, "Klasifikasi dengan Pohon Keputusan Berbasis Algoritma C4.5," vol. 3, pp. 64–71, 2020, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [8] E. Apriliyani and Y. Salim, "Analisis performa metode klasifikasi Naïve Bayes Classifier pada Unbalanced Dataset," *Indonesian Journal of Data and Science (IJODAS)*, vol. 3, no. 2, pp. 47–54, 2022, doi: 10.56705/ijodas.v3i2.45.
- [9] M. Fahmi Nugraha and S. B. Rahayu, "Penerapan Naïve Bayes dalam Mengklasifikasi Calon Penerima Bantuan Pangan Non Tunai di Desa Nanjung Mekar," *INTERNAL (Information System Journal)*, vol. 5, no. 2, pp. 137–146, doi: 10.32627.
- [10] O. Nurdawati and N. Salim, "PENERAPAN DATA MINING PADA PENJUALAN BARANG MENGGUNAKAN METODE METODE NAIVE BAYES CLASSIFIER UNTUK OPTIMASI STRATEGI PEMASARAN," *Jurnal Teknologi Informasi dan Komunikasi STMIK Subng*, 2018.
- [11] A. Fauzi Sistem Informasi, F. H. Universitas Buana Perjuangan Karawang Jl Ronggowaluyo, T. Timur, and K. priati, *Data Mining dengan Teknik Clustering Menggunakan Algoritma K-Means pada Data Transaksi Superstore*. 2017. [Online]. Available: <http://community.tableau.com>.
- [12] S. Mujahidin and M. C. Cahyo Utomo, "Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian naïve bayes", doi: 10.24036/voteteknika.v10i3.118299.
- [13] B. Rahmat *et al.*, "IMPLEMENTASI K-MEANS CLUSTERING PADA RAPIDMINER UNTUK ANALISIS DAERAH RAWAN KECELAKAAN."
- [14] S. Wibowo, "Penerapan Logika Fuzzy Dalam Penjadwalan Waktu Kuliah." doi: <https://doi.org/10.26877/jiu.v1i1%20Juni.809>.
- [15] R. Darmawan and A. Surahmat, "Optimalisasi Support Vector Machine (SVM) Berbasis Particle Swarm Optimization (PSO) Pada Analisis Sentimen Terhadap Official Account Ruang Guru Di Twitter," 2022. [Online]. Available: <http://ejournal.ubharajaya.ac.id/index.php/JKI>