

MENINGKATKAN MODEL KLASIFIKASI MENGGUNAKAN ALGORITMA *NAIVE BAYES* PADA DATA PENJUALAN DI TOKO PELANGI JAYA MOTOR

Nursyatika Berliana Susilowati¹⁾, Cep Lukman Rohman²⁾, Fadhil M.Basysyar³⁾, Muhamad Sulaeman⁴⁾

¹⁾ Jurusan Teknik Informatika, STMIK IKMI CIREBON

Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135

E-mail : ¹⁾nursyatikabeliana@gmail.com, ²⁾ceplukam.ikmi@gmail.com,

³⁾Fadhil.m.basysyar@gamil.com, ⁴⁾sulaeman.ikmi@gmail.com.

ABSTRAK

Toko Pelangi Jaya Motor menghadapi kesulitan dalam mengelola stok *sparepart* akibat ketidaksesuaian antara jumlah persediaan dan permintaan pelanggan. Untuk mengatasi hal ini, penelitian ini menggunakan algoritma *Naive Bayes* guna mengklasifikasikan barang "laris" dan "tidak laris" berdasarkan data penjualan. Dataset yang digunakan berjumlah 9.081 transaksi dengan *preprocessing* mencakup pembersihan data, pelabelan kategori, serta seleksi fitur menggunakan *Weight by Information Gain*. Hasil analisis menunjukkan bahwa atribut jumlah penjualan (*Qty*) memiliki pengaruh terbesar dalam klasifikasi dengan akurasi model mencapai 97.17%. Penelitian ini memberikan solusi berbasis data untuk meningkatkan efisiensi pengelolaan stok dan kepuasan pelanggan.

Kata kunci : Klasifikasi, *Naive Bayes*, Permintaan Pelanggan, *Sparepart* Motor

ABSTRACT

Pelangi Jaya Motor Store faces difficulties in managing spare part stock due to the mismatch between the amount of inventory and customer demand. To overcome this, this study uses the Naive Bayes algorithm to classify "hot" and "non-hot" items based on sales data. The dataset used is 9,081 transactions with preprocessing including data cleaning, category labeling, and feature selection using Weight by Information Gain. The results of the analysis show that the sales quantity (Qty) attribute has the greatest influence on classification with a model accuracy reaching 97.17%. This study provides a data-based solution to improve stock management efficiency and customer satisfaction.

Keywords: Classification, *Naive Bayes*, Customer Demand, Motor Spareparts

1. PENDAHULUAN

Sparepart merupakan sebuah barang yang berisikan berbagai komponen dalam suatu kesatuan dan memiliki fungsi tertentu. *Sparepart* banyak digunakan di berbagai jenis kendaraan, sehingga jenis-jenisnya juga sangat beragam [1]. Sebagai penyedia *sparepart*, Toko Pelangi Jaya Motor sering menghadapi tantangan dalam pengelolaan stok barang. Ketidaktepatan dalam memprediksi barang "Laris" dan "Tidak Laris" dapat menyebabkan kelebihan atau kekurangan stok, yang berujung pada

penurunan efisiensi operasional dan kepuasan pelanggan.

Pengelolaan stok yang tidak optimal dapat menimbulkan dampak signifikan, seperti meningkatnya biaya operasional akibat penumpukan barang yang tidak terjual atau hilangnya potensi penjualan karena stok habis. Untuk mengatasi masalah ini, diperlukan metode analisis data yang mampu memprediksi pola permintaan pelanggan secara akurat.

Penelitian ini bertujuan untuk mengatasi permasalahan tersebut dengan mengidentifikasi pola permintaan barang "Laris" menggunakan algoritma

Naïve Bayes. Algoritma ini dipilih karena kemampuannya dalam menganalisis data besar, memproses informasi dengan cepat, serta menghasilkan klasifikasi yang akurat meskipun dengan jumlah data yang besar dan atribut yang beragam. Dengan memahami pola permintaan pelanggan, toko dapat mengoptimalkan pengelolaan stok dan meminimalkan risiko kerugian.

Berbagai penelitian sebelumnya telah menunjukkan keberhasilan penerapan algoritma *Naïve Bayes* dalam bidang klasifikasi. Misalnya, penelitian oleh [2] menunjukkan bahwa algoritma ini dapat meningkatkan efisiensi dalam pengelolaan data penjualan. Penelitian lain oleh [3] membuktikan efektivitas metode ini dalam mendukung pengambilan keputusan strategis.

Berdasarkan hasil literatur sebelumnya, dapat disimpulkan bahwa algoritma *Naïve Bayes* memiliki keunggulan dalam mengklasifikasikan data penjualan dan ulasan produk dengan akurasi tinggi. Contohnya, dalam penelitian Nawangsih & Setyaningsih (2019) [4], algoritma *Naïve Bayes* berhasil mengklasifikasikan produk terlaris dalam penjualan voucher kuota dengan akurasi 97,50%. Sementara itu, penelitian yang dilakukan oleh Fahtu Rahman et al. (2024) [5] menggunakan *Naïve Bayes* untuk mengklasifikasikan ulasan pelanggan mengenai produk batik dengan akurasi 81,90%. Selain itu, penelitian oleh Teknika et al. (2024) [6] menunjukkan bahwa algoritma *Naïve Bayes* mampu memberikan akurasi yang sebanding dengan *K-Nearest Neighbor (KNN)*. Kelebihan utama dari *Naïve Bayes* adalah kemampuannya untuk mengklasifikasikan data ke dalam kategori tertentu dengan cepat serta kesederhanaannya dalam menangani dataset besar.

Berdasarkan hasil tersebut, penelitian ini dirancang untuk fokus pada pengelolaan stok *sparepart* di Toko Pelangi Jaya Motor dengan menerapkan algoritma *Naïve Bayes* untuk klasifikasi barang "Laris".

2. METODE PENELITIAN

Bagian ini menjelaskan pendekatan penelitian yang meliputi jenis data, metode pengumpulan data, serta tahapan dalam *Knowledge Discovery in Databases (KDD)*.

2.1 Jenis Data

Data yang digunakan pada penelitian ini bersifat sekunder [7] yaitu data diperoleh langsung dari transaksi penjualan di Toko Pelangi Jaya Motor.

2.2 Metode Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini menggunakan teknik sekunder yang diperoleh melalui kesepakatan antara peneliti dan pihak Toko Pelangi Jaya Motor di Cirebon. Data transaksi penjualan dikumpulkan secara langsung dari aplikasi toko yang digunakan untuk mencatat transaksi harian. Hasil dari pengumpulan data ini disediakan oleh pihak toko dalam format *Excel (xlsx)* berjumlah 9.081, yang kemudian diproses lebih lanjut dan diimpor ke dalam aplikasi *AI Studio 2024.1.0* untuk dianalisis sesuai dengan tujuan penelitian.

2.3 Metode Naïve Bayes

Penelitian ini menggunakan algoritma *Naïve Bayes*, salah satu algoritma klasifikasi yang umum digunakan dalam data mining. *Naïve Bayes* didasarkan pada Teorema Bayes, yang menghitung probabilitas setiap kelas berdasarkan atribut yang ada. Perhitungan probabilitas dilakukan menggunakan Persamaan (1).

$$P(c|x) = \frac{P(X|c) * P(c)}{P(X)} \quad (1)$$

Keterangan:

$P(c|x)$: Probabilitas hipotesis berdasar kondisi (*posteriori probability*)

$P(c)$: Probabilitas hipotesis (*prior probability*)

$P(x|c)$: Probabilitas berdasarkan kondisi pada hipotesis

$P(x)$: Probabilitas c

Sumber : [5]

2.4 Tahapan Perancangan KDD

1. *Data mining* merupakan bagian dari tahap proses *Knowledge Discovery in Database (KDD)*, yang bertujuan untuk mencari informasi baru dan berharga dalam suatu kumpulan data atau *database* [8]. Metode *data mining* yang digunakan dalam penelitian ini menggunakan proses tahapan *Knowledge Discovery in Database (KDD)* yang terdiri dari data *selection*, *preprocessing* data, data *transformation*, data mining, dan evaluasi dapat dilihat pada Gambar 1 [9].



Gambar 1. Tahapan Perancangan KDD

2. Data Selection

Merupakan tahapan dari pemilihan data dari sekumpulan data operasional yang perlu dilakukan sebelum tahap penggalian informasi pada KDD [10].

3. Preprocessing

Pada tahap ini adalah menyiapkan data agar siap digunakan untuk ke tahap selanjutnya seperti memperbaiki kesalahan dalam data [11].

4. Transformation

Transformasi merupakan gabungan intruksi untuk mengubah input menjadi output yang diinginkan, menyesuaikan juga dengan algoritma yang akan dipakai [10].

5. Data Mining

Tahapan yang paling utama dari proses tahapan perancangan adalah proses *data mining*, yang merupakan proses mencari bentuk atau informasi yang menarik pada sebuah data terpilih dengan menggunakan metode tertentu. Pada tahapan ini menggunakan algoritma *Naïve Bayes* untuk membangun model dan menguji model [10]. Tujuan utama dari *data mining* adalah untuk mengidentifikasi pola, hubungan, atau informasi yang mungkin tidak terlihat secara langsung pada data, sehingga dapat memberikan informasi yang lebih dalam dan valid [12].

6. Evaluation

Hasil dari *data mining*, yakni alur informasi yang muncul dari proses, perlu disajikan dalam format yang dapat dipahami dengan mudah oleh para pemangku kepentingan [13].

7. Result

Pada tahapan ini merupakan tahapan terakhir dilakukannya identifikasi barang-barang *sparepart* yang masuk dalam label "Laris".

2.5 Penanganan Data Hilang

Dalam tahap preprocessing, tidak ditemukan data duplikat karena setiap transaksi bersifat unik. Namun, terdapat nilai kosong pada atribut jumlah penjualan (*Qty*). Untuk mengatasi hal ini, nilai kosong digantikan [14] dengan 0 agar data dapat diproses dengan baik oleh algoritma *Naïve Bayes* tanpa kehilangan informasi yang relevan. Hal ini memastikan bahwa semua transaksi tetap terhitung dalam proses analisis.

2.6 Pemilihan Split Data 70%-30% Tanpa Validasi Silang

Split data adalah proses pemisahan data menjadi dua set berbeda, yaitu data latih (*training*) dan data uji (*testing*) [15]. Pemilihan rasio 70%-30% didasarkan pada keseimbangan antara pelatihan model yang cukup dan pengujian yang representatif [6]. Validasi silang tidak digunakan karena *dataset* yang dianalisis sudah memiliki distribusi yang cukup stabil dan representatif. Selain itu, pendekatan ini telah digunakan dalam penelitian sebelumnya dengan tingkat keberhasilan yang baik dalam kasus klasifikasi serupa.

3. HASIL DAN DISKUSI

Proses analisis data dilakukan menggunakan aplikasi *AI Studio 2024.1.0* dengan langkah-langkah berikut sesuai tahapan KDD:

1. Data Selection

Untuk mengelola sebuah data dengan cara menempatkan operator *Read Excel* untuk membaca *dataset* dalam format *excel*. Proses membaca *dataset* dalam format *Excel* dilakukan dengan menggunakan operator *Read Excel*, seperti yang ditunjukkan pada Gambar 2.



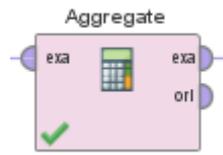
Gambar 2. Operator Read Excel

2. Preprocessing

1. Pengelompokkan Data

Data dikelompokkan berdasarkan attribute Tanggal, Kategori, Sat dan Nama Barang dengan menggunakan operator

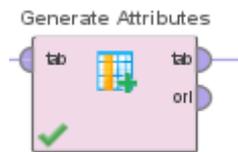
Aggregate. Prosesnya ditunjukkan pada Gambar 3.



Gambar 3. Operator *Aggregate*

2. Pembuatan Label

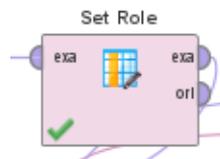
Pelabelan data dilakukan dengan operator *Generate Attribute*, dengan ketentuan: jika *Qty* (Kuantitas) lebih dari 10, maka diberi label "Laris", dan jika *Qty* (Kuantitas) Kurang dari atau sama dengan 10, maka diberi label "Tidak Laris". Penentuan label ini berdasarkan wawancara dengan pihak toko. Proses ditunjukkan pada Gambar 4.



Gambar 4. Operator *Generate Attributes*

3. Pengaturan *Attribute*

Setelah melakukan pelabelan, kemudian dilakukan pengaturan peran atribut menggunakan operator *set role* agar menjadi atribut label pada proses analisis di *AI Studio 2024.1.0*. Proses ditunjukkan pada Gambar 5.



Gambar 5. Operator *Set Role*

3. *Transformation*

a. Penilaian Atribut

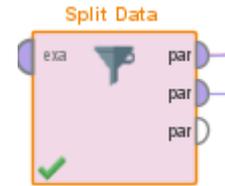
Menilai setiap atribut untuk memilih fitur yang paling berpengaruh menggunakan operator *Weight by Information Gain*.



Gambar 6. Operator *Weight by Information Gain*

b. Seleksi Fitur

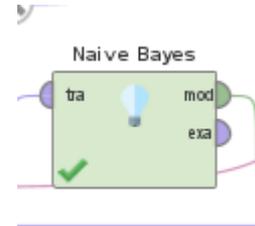
Transformation data menggunakan operator *Split Data* dengan parameter 0.7 untuk data *training* dan 0.3 untuk data *testing*. Prosesnya ditunjukkan pada Gambar 7.



Gambar 7. Operator *Split Data*

4. *Data Mining*

Data mining menggunakan operator *Naive Bayes* untuk klasifikasi. Prosesnya sebagai berikut:



Gambar 8. Operator *Naive Bayes*

5. *Evaluation*

Evaluation atau evaluasi model. Disini peneliti menggunakan operator *performance*. Prosesnya sebagai berikut:



Gambar 9. Operator *Performance*

Hasil yang didapatkan pada pengujian adalah nilai *accuracy*, *precision*, dan *recall* yang ditunjukkan pada **Tabel 1**.

Tabel 1. Hasil *Confusion Matrix*

	true tidak laris	true laris	class precision
pred. tidak laris	1938	7	99.64%
pred. laris	54	157	74.41%
class recall	97.29%	95.73%	

- a. *Accuracy* menunjukkan hasil yang benar terhadap keseluruhan data, berikut rumusnya:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Akurasi} = \frac{157 + 1938}{157 + 1938 + 54 + 7} = \frac{2095}{2156} \approx 97.17\%$$

- b. *Precision* mengukur tingkat keakuratan prediksi positif terhadap semua data yang diprediksi laris, berikut rumusnya:

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Precision} = \frac{157}{157 + 54} = \frac{157}{211} \approx 74.41\%$$

- c. *Recall* menunjukkan sejauh mana model dapat menangkap data positif yang sebenarnya, berikut rumusnya:

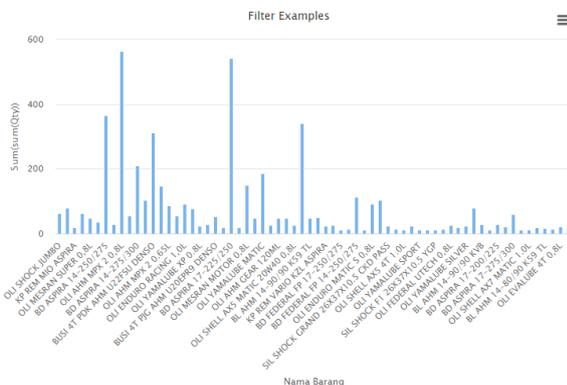
$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Recall} = \frac{157}{157 + 7} = \frac{157}{164} \approx 95.73\%$$

Interpretasi hasil *Confusion Matrix* pada Tabel 1, diperoleh bahwa *accuracy* model mencapai 97.17%, dengan *precision* kelas "Laris" sebesar 74.41% dan *recall* kelas "Laris" sebesar 95.73%. Nilai *recall* yang tinggi menunjukkan bahwa model mampu mengenali sebagian besar barang yang benar-benar laris, tetapi *precision* yang lebih rendah mengindikasikan masih adanya beberapa kesalahan prediksi dalam mengklasifikasikan barang yang seharusnya tidak laris.

6. Report

Setelah tahap *evaluation* selesai, hasil menunjukkan untuk barang-barang *sparepart* yang menunjukkan label laris sebagai berikut:



Gambar 10. Visualisasi *Report*

Gambar di atas merupakan diagram batang yang menunjukkan jumlah total penjualan barang (*Sum(Qty)*) berdasarkan nama barang di Toko Pelangi Jaya Motor. Sumbu horizontal (X) mewakili berbagai *sparepart*, seperti BD (Ban Dalam), oli, busi, dan *sparepart* motor lainnya, sementara sumbu vertikal (Y) menunjukkan total kuantitas penjualan barang dalam satuan tertentu. Dari diagram ini, dapat terlihat bahwa terdapat beberapa barang dengan penjualan yang sangat tinggi, seperti "OLI AHM MPX 2 0,8L", "BD ASPIRA 17-225/250", "BD ASPIRA 14-250/275", "BD ASPIRA 17-250/275" dan "BUSI 4T PDK AHM U22FSU DENSO".

4. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengidentifikasi barang "Laris" menggunakan algoritma Naïve Bayes, yang membantu Toko Pelangi Jaya Motor dalam memahami pola permintaan pelanggan. Dengan penerapan model ini, toko dapat mengoptimalkan pengelolaan stok, mengurangi risiko kelebihan atau kekurangan barang, serta meningkatkan kepuasan pelanggan. Model yang digunakan memiliki beberapa kelebihan, seperti akurasi tinggi dalam klasifikasi barang, efisiensi dalam komputasi, serta kemampuan menangani dataset transaksi dalam jumlah besar. Namun, model ini juga memiliki kekurangan, terutama dalam hal tidak mempertimbangkan hubungan kompleks antar fitur dan *precision* yang masih perlu ditingkatkan untuk kelas "Laris". Untuk pengembangan lebih lanjut, penelitian mendatang dapat mengeksplorasi metode ensemble seperti *Random Forest* untuk meningkatkan performa klasifikasi, menerapkan validasi silang untuk meningkatkan generalisasi model, serta mengintegrasikan sistem rekomendasi agar toko dapat mengoptimalkan stok secara otomatis berdasarkan pola permintaan pelanggan.

5. DAFTAR PUSTAKA

- [1] J. M. Sitinjak, K. Sari, and M. Yetri, "Penerapan Data Mining Dalam Penjualan Sparepart Motor Dengan Menggunakan Metode K-Means Clustering," vol. 3, no. November, pp. 862–871, 2024.
- [2] P. Ayu, W. Purnama, and T. A. Putra, "Klasifikasi Penjualan Produk Menggunakan Algoritma Naive Bayes pada Konter HP Bayu Cell," *Remik Ris. dan E-Jurnal Manaj. Inform. Komput.*, vol. 8, no. 1, pp. 286–292, 2024, [Online]. Available:

<http://doi.org/10.33395/remik.v8i1.13207>

- [3] P. S. Dewi, C. K. Sastradipraja, and D. Gustian, “Sistem Pendukung Keputusan Kenaikan Jabatan Menggunakan Metode Algoritma Naïve Bayes Classifier,” *J. Teknol. dan Inf.*, vol. 11, no. 1, pp. 66–80, 2021, doi: 10.34010/jati.v11i1.3593.
- [4] I. Nawangsih and A. Setyaningsih, “Penerapan Algoritma Naive Bayes Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Voucher Kuota Di Edi Cell,” *J. SIGMA*, vol. 3, pp. 14902–14914, 2019, [Online]. Available: <https://jurnal.pelitabangsa.ac.id/index.php/signa/article/view/2256>
- [5] M. A. Fahtu Rahman, Z. R. Mair, and D. Sartika, “Klasifikasi Ulasan Pelanggan Shopee Mall Terhadap E-Commerce Penjualan Baju Batik Metode Naïve Bayes,” *IDEALIS Indones. J. Inf. Syst.*, vol. 7, no. 2, pp. 164–177, 2024, doi: 10.36080/idealisis.v7i2.3178.
- [6] J. Teknika, K. Penjualan Produk Terlaris Pada Kedai Ira Dengan Menggunakan Algoritma Naïve Bayes Dan Algoritma K-Nearest Neighbor Intan Purwasih, K. Setiawan, F. Matheos Sarimole, J. Sistem Informasi, and S. Tinggi Ilmu Komputer Cipta Karya, “Klasifikasi Penjualan Produk Terlaris Pada Kedai Ira Dengan Menggunakan Algoritma Naïve Bayes Dan Algoritma K-Nearest Neighbor Intan,” *Ijccs*, vol. x, No.x, no. x, pp. 1–5, 2024.
- [7] A. Azis, A. T. Zy, and A. S. Sunge, “Prediksi Penjualan Obat Dan Alat Kesehatan Terlaris Menggunakan Algoritma K-Nearest Neighbor,” *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 1, pp. 117–124, 2024, doi: 10.47233/jteksis.v6i1.1078.
- [8] R. Hamonangan, R. K. Sari, S. Anwar, and T. Hartati, “Klasifikasi Algoritma KNN dalam menentukan Penerima Bantuan Langsung Tunai,” vol. 6, no. 1, pp. 198–204, 2024.
- [9] M. Sutra Safira, N. Rahaningsih, and R. Dinar Dana, “Penerapan Data Mining Untuk Klasifikasi Penjualan Obat Menggunakan Metode K-Nearest Neighbor,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 380–385, 2024, doi: 10.36040/jati.v8i1.8325.
- [10] P. Kepemilikan, A. Kelahiran, and D. I. Jawa, “IMPLEMENTASI ALGORITMA K-MEANS TERHADAP,” vol. 18, pp. 1–9, 2023.
- [11] B. N. Sari *et al.*, “PENERAPAN DATA MINING DALAM KLASIFIKASI DATA TRANSAKSI PRODUK KOPERASI DI SMK PGRI 2 KARAWANG,” vol. 9, no. 1, pp. 263–269, 2025.
- [12] Z. Fatah *et al.*, “KLASIFIKASI DATA MINING UNTUK MEMPREDIKSI KELULUSAN MAHASISWA,” vol. 2, no. 1, pp. 29–37, 2025.
- [13] D. Sajidan, N. Suarna, and T. Suprpti, “Analisis Pola Penjualan Sepatu Dengan Implementasi Algoritma Apriori Data Mining,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 2340–2347, 2024, doi: 10.36040/jati.v8i2.8818.
- [14] M. Irfan and A. A. Syukron, “IMPLEMENTASI DATA MINING UNTUK MENGLASIFIKASI PENJUALAN PRODUK TERLARIS PADA KOPERASI AL-BAEDLOWI,” vol. 9, no. 1, pp. 1423–1428, 2025.
- [15] A. A. Surya and Y. Yamasari, “Penerapan Algoritma Naïve Bayes (NB) untuk Klasifikasi Penyakit Jantung,” *J. Informatics Comput. Sci.*, vol. 5, no. 03, pp. 447–455, 2024, doi: 10.26740/jinacs.v5n03.p447-455.