

Penerapan Algoritma Klasifikasi Naïve Bayes Dalam Pengambilan Keputusan Preferensi Konsumen Terhadap Produk Sprocket Dan Chain

Yuliani Sinta^a, Alyauma Hajjah^b

^aInstitut Bisnis dan Teknologi Pelita Indonesia, Jl. Jend. Ahmad Yani No. 77-88 Pekanbaru, yuliani.sinta@student.pelitaindonesia.ac.id

^bInstitut Bisnis dan Teknologi Pelita Indonesia, Jl. Jend. Ahmad Yani No. 77-88 Pekanbaru, alyaumahajjah@lecturer.pelitaindonesia.ac.id

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 20 April 2026

Revisi Akhir: 30 Maret 2026

Diterbitkan Online: 30 April 2026

KATA KUNCI

Preferensi Konsumen, Naïve Bayes, Data Penjualan, Klasifikasi

KORESPONDENSI

E-mail: yuliani.sinta@student.pelitaindonesia.ac.id

ABSTRACT

Penelitian ini bertujuan membangun model klasifikasi untuk mengidentifikasi preferensi pembelian konsumen terhadap produk sprocket dan chain berdasarkan data transaksi penjualan PT. Buana Rantai Berkat Abadi Cabang Pekanbaru periode 2022–2024. Dataset awal berjumlah 1.818 transaksi dan setelah melalui tahap preprocessing menghasilkan 1.787 transaksi valid. Variabel target yang digunakan adalah Preferensi dengan dua kategori, yaitu Diminati dan Tidak Diminati, yang ditentukan berdasarkan jumlah pembelian, yaitu $QTY \geq 5$ sebagai Diminati dan $QTY < 5$ sebagai Tidak Diminati. Adapun atribut prediktor yang digunakan meliputi delivery, jenis produk, merek, ukuran, dan kategori harga. Pemodelan diimplementasikan menggunakan bahasa pemrograman Python pada Google Colab dengan algoritma Categorical Naïve Bayes serta Ordinal Encoding. Evaluasi model dilakukan menggunakan metode train-test split dengan rasio 80:20 dan 10-fold cross-validation. Hasil pengujian menggunakan train-test split menunjukkan nilai akurasi sebesar 88,55%. Sementara itu, hasil evaluasi menggunakan 10-fold cross-validation menghasilkan rata-rata akurasi sebesar 89,14%, precision sebesar 80,62%, recall sebesar 73,75%, dan F1-score sebesar 76,40% (macro average). Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes memiliki performa yang baik dalam mengklasifikasikan preferensi pembelian konsumen dan dapat dimanfaatkan sebagai dasar awal dalam mendukung analisis kecenderungan permintaan produk serta penyusunan rekomendasi produk berdasarkan data historis penjualan.

1. PENDAHULUAN

Perkembangan teknologi informasi pada era digital telah membawa transformasi signifikan terhadap praktik pengelolaan data dan pengambilan keputusan di berbagai perusahaan[1]. Data mining merupakan salah satu pendekatan analitis yang banyak diterapkan untuk mengekstrak pengetahuan dari data historis penjualan. Teknik ini memungkinkan peneliti mengungkap pola atau hubungan antar variabel melalui metode statistik, matematika, dan kecerdasan buatan[2]. PT. Buana Rantai Berkat Abadi merupakan perusahaan yang bergerak dalam distribusi komponen mechanical power transmission[3]. Meskipun perusahaan memiliki volume transaksi yang besar dan data penjualan yang terdokumentasi dengan baik, pemanfaatan data tersebut pada cabang Pekanbaru masih terbatas pada penyusunan laporan penjualan dan pencatatan administratif. Penelitian ini akan menganalisis dataset penjualan menggunakan algoritma Naïve Bayes untuk mengidentifikasi preferensi konsumen secara sistematis.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining merupakan suatu proses eksplorasi dan ekstraksi pengetahuan dari kumpulan data berskala besar untuk mengidentifikasi pola, hubungan, serta informasi yang sebelumnya tersembunyi[4]. Proses ini merupakan bagian integral dari tahapan Knowledge Discovery in Databases (KDD) yang meliputi pemahaman domain dan data, pembersihan dan transformasi data, seleksi fitur, penerapan metode analisis, hingga interpretasi dan evaluasi hasil[5].

2.2. Algoritma Klasifikasi

Klasifikasi merupakan salah satu teknik dalam data mining yang digunakan untuk membangun suatu model prediktif yang mampu mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan atribut yang dimilikinya[6]. Proses klasifikasi terdiri atas dua tahap utama : tahap pelatihan

(training phase) dan tahap pengujian (testing/classification phase)[7].

2.3. Algoritma Naive Bayes

Algoritma Naive Bayes merupakan salah satu metode klasifikasi yang didasarkan pada pendekatan probabilistik menggunakan Teorema Bayes[8]. Keunggulan utama algoritma ini meliputi efisiensi komputasi yang tinggi, struktur perhitungan yang relatif sederhana, serta kemampuannya dalam menangani data dengan atribut kategorikal[9]. Secara matematis, algoritma ini dinyatakan sebagai berikut[10] :

$$P(X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (2.1)$$

Dimana :

Y = kelas yang diprediksi

X = atribut atau data pengamatan

P (X|Y) = probabilitas likelihood

P (Y) = probabilitas prior

P (X) = probabilitas evidence

P (Y|X) = probabilitas posterior

Untuk multiple atribut ($X_1, X_2, \dots, X_n$), probabilitas posterior dihitung dengan[11] :

$$\frac{P(Y|X_1, X_2, \dots, X_n)}{P(Y) \prod_{i=1}^n P(X_i|Y)} \quad (2.2)$$

Karena penyebut bersifat konstan, perhitungan dapat disederhanakan menjadi [12]:

$$P(Y|X_1, X_2, \dots, X_n) \propto P(Y) \prod_{i=1}^n P(X_i|Y) \quad (2.3)$$

Untuk mengatasi probabilitas nol, digunakan teknik Laplace Smoothing [13]:

$$P(X_i|Y) = \frac{\text{count}(X_i, Y) + 1}{\text{count}(Y) + k} \quad (2.4)$$

Dimana :

count(X_i, Y) = jumlah kemunculan nilai atribut X_i pada kelas Y

count(Y) = jumlah total data pada kelas Y

k = jumlah kategori unik pada atribut X_i

2.4. Metrik Evaluasi Model

Evaluasi kinerja model klasifikasi dilakukan menggunakan confusion matrix. Tahapan ini diperlukan untuk memastikan bahwa model yang dikembangkan memiliki tingkat keandalan yang memadai sebelum digunakan dalam pengambilan keputusan yang bersifat operasional maupun strategis [14]. Confusion matrix terdiri atas empat komponen utama, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN), yang menjadi dasar dalam perhitungan berbagai metrik evaluasi kinerja model[15]. Beberapa metrik evaluasi yang digunakan adalah :

A. Accuracy (Akurasi)

Akurasi mengukur proporsi prediksi yang benar dari seluruh prediksi :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.5)$$

B. Precision (Presisi)

Presisi mengukur ketepatan prediksi kelas positif :

$$Precision = \frac{TP}{TP+FP} \quad (2.6)$$

C. Recall

Recall mengukur kemampuan model mendeteksi kelas positif :

$$Recall = \frac{TP}{TP+FN} \quad (2.7)$$

D. F1-Score

F1-Score merupakan rata-rata harmonik antara precision dan recall :

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.8)$$

Keterangan Komponen Confusion Matrix :

TP (True Positive) = Data positif yang diprediksi dengan benar

TN (True Negative) = Data negatif yang diprediksi dengan benar

FP (False Positive) = Data negatif yang salah diprediksi sebagai positif

FN (False Negative) = Data positif yang salah diprediksi sebagai negatif

E. Cross-Validation

Cross-validation merupakan teknik evaluasi model yang dilakukan dengan membagi dataset ke dalam beberapa subset (fold), kemudian melatih dan menguji model secara bergantian pada setiap subset[16]. Dengan cara ini, hasil evaluasi mencerminkan performa model secara lebih menyeluruh[17]. Nilai rata-rata metrik evaluasi dihitung dengan :

$$M = \frac{1}{k} \sum_{i=1}^k M_n \quad (2.9)$$

Dimana :

M_i = nilai metrik evaluasi pada fold ke-i

k = jumlah fold yang digunakan

2.5. Preferensi Konsumen

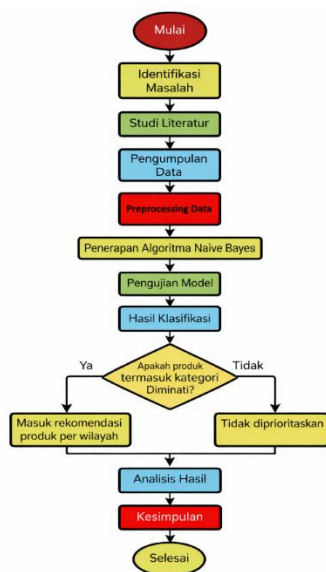
Preferensi konsumen didefinisikan sebagai kecenderungan atau pilihan konsumen terhadap suatu produk dibandingkan alternatif lain[18]. Preferensi bersifat dinamis dan dipengaruhi oleh faktor internal (persepsi, pengalaman) maupun eksternal (tren pasar, lokasi geografis)[19]. Dalam konteks penelitian ini, preferensi dikategorikan menjadi Diminati ($QTY \geq 5$) dan Tidak Diminati ($QTY < 5$).

2.6. Sprocket dan Chain

Sprocket dan chain merupakan komponen utama dalam sistem transmisi daya mekanis[20]. Sprocket umumnya memiliki spesifikasi teknis berdasarkan jumlah gigi, diameter, dan profil gigi. Pemilihan ukuran sprocket yang tepat diperlukan agar torsi dan kecepatan putar yang ditransmisikan dapat disesuaikan dengan kebutuhan aplikasi, sehingga kinerja sistem mekanis tetap optimal[21]. chain harus memiliki kekuatan tarik yang memadai, ketahanan terhadap gesekan dan keausan, serta kemampuan bekerja dalam kondisi beban berulang tanpa mengalami kerusakan struktural yang signifikan[22]

3. METODOLOGI

3.1 Kerangka Penelitian



Gambar 1. Kerangka Penelitian

Kerangka penelitian yang disusun telah menggambarkan alur penelitian secara sistematis dan terarah, dimulai dari identifikasi masalah bahwa data penjualan historis perusahaan belum dimanfaatkan secara optimal untuk memberikan preferensi atau rekomendasi produk kepada pelanggan.

Tahap berikutnya adalah studi literatur sebagai landasan teoritis, kemudian pengumpulan data penjualan historis produk sprocket dan chain periode 2022–2024.

Data selanjutnya dipersiapkan melalui tahap preprocessing yang meliputi data cleaning, data selection, dan data transformation, termasuk kategorisasi harga serta pelabelan preferensi dengan ketentuan $QTY \geq 5$ sebagai Diminati dan $QTY < 5$ sebagai Tidak Diminati.

Setelah itu dilakukan penerapan algoritma Naïve Bayes untuk membangun model klasifikasi, dilanjutkan dengan pengujian model menggunakan train-test split dan cross-validation melalui metrik akurasi, presisi, recall, dan F1-score berbasis confusion matrix.

Hasil pengujian kemudian dianalisis untuk menilai efektivitas model, lalu dimanfaatkan dalam penyusunan rekomendasi produk per wilayah, sebelum akhirnya ditarik kesimpulan

sebagai dasar pengambilan keputusan berbasis data di perusahaan.

3.2 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah algoritma klasifikasi Naïve Bayes. Data yang telah melalui tahap preprocessing (pembersihan, seleksi atribut, dan transformasi) akan digunakan sebagai input dalam pembentukan model. Proses pelatihan (training) dan pengujian (testing) model dilakukan dengan teknik train-test split menggunakan bahasa pemrograman Python melalui Google Colab. Performa model dievaluasi berdasarkan metrik akurasi, presisi, recall, dan F1-score, yang dihitung dari confusion matrix.

3.3 Perhitungan Manual

Dataset awal penelitian ini terdiri dari 1.818 transaksi penjualan produk sprocket dan chain periode 2022–2024 yang diperoleh dari catatan historis PT. Buana Rantai Berkat Abadi Cabang Pekanbaru. Setelah dilakukan preprocessing (pembersihan data), jumlah data berkurang menjadi 1.787 transaksi valid karena terdapat 31 data duplikat yang dihapus. Selanjutnya dataset dibagi menggunakan skema train-test split 80:20, yaitu 1.429 data sebagai data latih dan 358 data sebagai data uji untuk mengevaluasi kinerja model klasifikasi.

Evaluasi performa model dilakukan menggunakan confusion matrix untuk menghitung nilai accuracy, precision, recall, dan F1-score. Komponen confusion matrix terdiri dari True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN).

Tabel 1 : Confusion Matrix Hasil Prediksi Model

	Aktual Tidak Diminati	Aktual Diminati
Prediksi Tidak Diminati	286	17
Prediksi Diminati	24	31

Berdasarkan nilai-nilai tersebut, diperoleh metrik evaluasi sebagai berikut :

$$\begin{aligned}
 \text{Akurasi :} \\
 &= \frac{31 + 286}{286 + 17 + 24 + 31} \\
 &= \frac{317}{358} \\
 &= 0,8855 = 88,55\%
 \end{aligned}$$

$$\begin{aligned}
 \text{Presisi (kelas Diminati) :} \\
 &= \frac{31}{31 + 17} \\
 &= \frac{31}{48} \\
 &= 0,6458 = 64,58\%
 \end{aligned}$$

Recall (kelas Diminati) :

$$\begin{aligned}
 &= \frac{31}{31 + 24} \\
 &= \frac{31}{55} \\
 &= 0,5636 = 56,36\%
 \end{aligned}$$

F1-Score (kelas Diminati) :

$$\begin{aligned}
 &= 2 \times \frac{0,6458 \times 0,5636}{0,6458 + 0,5636} \\
 &= \frac{0,7280}{1,2094} \\
 &= 0,6019 = 60,19\%
 \end{aligned}$$

Presisi (kelas Tidak Diminati) :

$$\begin{aligned}
 &= \frac{286}{286 + 24} \\
 &= \frac{286}{310} \\
 &= 0,9226 = 92,26\%
 \end{aligned}$$

Recall (kelas Tidak Diminati) :

$$\begin{aligned}
 &= \frac{286}{286 + 17} \\
 &= \frac{286}{303} \\
 &= 0,9439 = 94,39\%
 \end{aligned}$$

F1-Score (kelas Tidak Diminati) :

$$\begin{aligned}
 &= 2 \times \frac{0,9226 \times 0,9439}{0,9226 + 0,9439} \\
 &= \frac{1,7420}{1,8665} \\
 &= 0,9331 = 93,31\%
 \end{aligned}$$

Tabel 2 : Ringkasan Metrik Evaluasi Model (Train-Test Split)

Metrik	Nilai (%)
Akurasi	88,55
Precision Diminati	64,58
Precision Tidak Diminati	92,26
Recall Diminati	56,36
Recall Tidak Diminati	94,39
F1-Score Diminati	60,19
F1-Score Tidak Diminati	93,31

Selain evaluasi menggunakan train-test split, model juga diuji menggunakan 10-fold cross-validation untuk memperoleh estimasi performa yang lebih stabil dan mengurangi ketergantungan pada satu kali pembagian data.

Pada setiap fold, model menghasilkan prediksi pada data uji, kemudian dihitung metrik evaluasi berupa akurasi, precision (macro), recall (macro), dan F1-score (macro).

Tabel 3 : Hasil Evaluasi Model Naïve Bayes pada Setiap Fold (10-Fold Cross-Validation)

Fold	Accuracy	Precision (macro)	Recall (macro)	F1-score (macro)
1	0.9106	0.8572	0.7646	0.8004
2	0.8994	0.8278	0.7428	0.7755
3	0.905	0.8373	0.7613	0.7918

4	0.8883	0.7984	0.721	0.7506
5	0.8771	0.7706	0.7381	0.7525
6	0.9106	0.8623	0.7725	0.8075
7	0.8827	0.781	0.7559	0.7675
8	0.9045	0.8483	0.746	0.7838
9	0.8876	0.7874	0.7513	0.7673
10	0.8483	0.6914	0.6217	0.6429
TOTAL	8.9142	8.0615	7.3752	7.6398

Berdasarkan nilai pada Tabel, dilakukan perhitungan manual untuk memperoleh nilai rata-rata setiap metrik sebagai berikut :

Akurasi :

$$\begin{aligned}
 &= \frac{8.9142}{10} \\
 &= 0,8914 = 89,14\%
 \end{aligned}$$

Presisi (macro) :

$$\begin{aligned}
 &= \frac{8.0615}{10} \\
 &= 0,8062 = 80,62\%
 \end{aligned}$$

Recall (kelas Diminati) :

$$\begin{aligned}
 &= \frac{7.3752}{10} \\
 &= 0,7375 = 73,75\%
 \end{aligned}$$

F1-Score (kelas Diminati) :

$$\begin{aligned}
 &= \frac{7.6398}{10} \\
 &= 0,7640 = 76,40\%
 \end{aligned}$$

Tabel 4 : Ringkasan Metrik Evaluasi (10-Fold Cross-Validation)

Metrik	Nilai (%)
Akurasi	89,14
Precision (macro)	80,62
Recall (macro)	73,75
F1-Score (macro)	76,40

4. HASIL DAN PEMBAHASAN

4.1. Evaluasi Model Naive Bayes

Tahap pemodelan menggunakan Categorical Naïve Bayes dengan variabel target Preferensi dan variabel prediktor Delivery, Jenis Produk, Merek, Ukuran, serta Kategori Harga. Seluruh prediktor bersifat kategorikal sehingga dilakukan Ordinal Encoding agar dapat diproses oleh model. Evaluasi performa dilakukan menggunakan 10-fold cross-validation dengan StratifiedKFold (shuffle=True, random_state=42) untuk menjaga proporsi kelas pada setiap fold.

```
Mounted at /content/drive
Kolon dataset: ['NO', 'TGL', 'DELIVERY', 'JENIS_PRODUK', 'MERK', 'UKURAN', 'QTY', 'UNIT_PRICE', 'U
{'fit_time': array([0.01394749, 0.01269054, 0.00858463, 0.0081234, 0.00800562,
0.00768232, 0.0073248, 0.00790477, 0.00798464, 0.00781798]),
'score_time': array([0.02101421, 0.01919746, 0.01395035, 0.01374793, 0.01306963,
0.0129807, 0.01316643, 0.01292968, 0.01956058, 0.01274514]),
'test_accuracy': array([0.61452514, 0.70949721, 0.59217872, 0.7158838, 0.70949721,
0.67597765, 0.65921788, 0.71348315, 0.6741573, 0.66292135]),
'test_precision_macro': array([0.60260903, 0.70135629, 0.57937435, 0.70816573, 0.70333881,
0.66613966, 0.64816305, 0.70639835, 0.66536284, 0.65389577]),
'test_recall_macro': array([0.60134615, 0.69794872, 0.57839744, 0.69903846, 0.69851282,
0.66166667, 0.64166667, 0.70711974, 0.66487767, 0.65581559]),
'test_f1_macro': array([0.60179901, 0.6992763, 0.5787149, 0.70163742, 0.69341238,
0.66309709, 0.64313495, 0.70673558, 0.66463898, 0.65436893])}
```

Gambar 1 : Output Proses Cross-Validation pada Pemodelan Naïve Bayes

Metrik evaluasi yang dihitung pada setiap fold meliputi accuracy, precision (macro), recall (macro), dan F1-score (macro). Berdasarkan hasil tersebut, diperoleh nilai akurasi sebesar 89,14%, precision sebesar 80,62%, recall sebesar 73,75%, dan F1-score sebesar 76,40%.

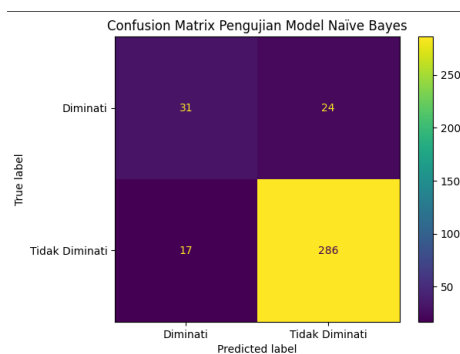
	Metrik	Nilai (%)
0	Akurasi	89.14
1	Precision (macro)	80.62
2	Recall (macro)	73.75
3	F1-Score (macro)	76.40

Gambar 2 : Output hasil evaluasi model Naïve Bayes

Nilai metrik yang relatif berdekatan menunjukkan performa model yang cukup konsisten. Namun, variasi kategori yang tinggi pada data (terutama atribut ukuran) dapat menimbulkan tumpang tindih pola antar kelas preferensi sehingga mempengaruhi ketepatan klasifikasi pada sebagian data.

4.2. Hasil Pengujian

Pengujian dilakukan menggunakan skema train–test split 80:20 dan dari total 1.787 data valid, sebanyak 1.429 data digunakan sebagai data latih dan 358 data digunakan sebagai data uji. Hasil prediksi pada data uji kemudian dievaluasi menggunakan confusion matrix.



Gambar 3 : Confusion Matrix Pengujian Model Naïve Bayes

Berdasarkan confusion matrix, diperoleh nilai TP = 31, TN = 286, FP = 24, dan FN = 17. Hasil evaluasi menunjukkan bahwa model mencapai nilai akurasi sebesar 88,55%. Pada kelas Diminati, model memperoleh nilai precision sebesar 64,58%, recall sebesar 56,36%, dan F1-score sebesar 60,19%. Sementara itu, pada kelas Tidak Diminati, model memperoleh nilai precision sebesar 92,26%, recall sebesar 94,39%, dan F1-score sebesar 93,31%.

Tabel 5 : Metrik Evaluasi Pengujian Model Naïve Bayes

Kelas	Precision (%)	Recall (%)	F1-Score (%)
Diminati	64.58	56.36	60.19
Tidak Diminati	92.26	94.39	93.31

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian dan pembahasan yang telah dilakukan, maka dapat diambil kesimpulan sebagai berikut :

1. Penerapan algoritma Naïve Bayes mampu mengklasifikasikan preferensi konsumen produk sprocket dan chain ke dalam dua kelas, yaitu Diminati dan Tidak Diminati, berdasarkan atribut delivery (wilayah), jenis produk, merek, ukuran, dan kategori harga.
2. Evaluasi model menggunakan metode train–test split (80:20) menghasilkan nilai akurasi sebesar 88,55%, sedangkan pengujian menggunakan 10-fold cross-validation menghasilkan rata-rata akurasi sebesar 89,14%, precision sebesar 80,62%, recall sebesar 73,75%, dan F1-score sebesar 76,40% (macro average)ss.
3. Hasil klasifikasi preferensi konsumen dapat dimanfaatkan sebagai dasar penyusunan rekomendasi produk per wilayah untuk mendukung pengambilan keputusan perusahaan, terutama dalam strategi distribusi dan pengelolaan persediaan.

Adapun saran untuk penelitian ini yaitu :

1. Perusahaan disarankan melakukan pembaruan data secara berkala agar model tetap sesuai dengan perubahan pola permintaan, sehingga akurasi prediksi dan relevansi rekomendasi tetap terjaga.
2. Penelitian selanjutnya disarankan menambahkan atribut yang lebih beragam serta membandingkan Naïve Bayes dengan metode klasifikasi lain untuk meningkatkan performa, khususnya pada kelas Tidak Diminati.

DAFTAR PUSTAKA

- [1] R. Rizkiyanto and A. Rahmayana, "Proceedings of the 3rd International Conference on Pesantren (3rd ICOP) 2025 26 May 2025, Darunnajah University, Jakarta, Indonesia," no. May, 2025.
- [2] R. P. M. Rosidi and K. Setiawan, "Implementasi Algoritma Naïve Bayes Terhadap Data Penjualan untuk Mengetahui Pola Pembelian Konsumen pada Kantin," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 1, pp. 120–126, 2024, doi: 10.35870/jimik.v5i1.407.
- [3] Stefvy and D. Salim, "Penerapan Metode Penyusutan Menurut Ketentuan Pajak Dan Pengaruhnya Terhadap Pajak Penghasilan Badan Di Pt Buana Rantai Berkat Abadi Medan," *JUSAPAK J. Stud. Akunt. Pajak*

Keuang., vol. 2, no. 3, pp. 146–151, 2024.

- [4] P. M. S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, and R. Winanjaya, "IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA APRIORI DALAM MENENTUKAN PERSEDIAAN BARANG (STUDI KASUS: TOKO SINAR HARAHAHAP)," *J. Tek. Inf. dan Komput.*, vol. 7, no. 1, p. 144, 2022, doi: 10.37600/tekinkom.v7i1.1270.
- [5] F. Alghifari and D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *J. Ilm. Inform.*, vol. 9, no. 2, pp. 75–81, 2021.
- [6] Hizbul Izzi, Arief Setyanto, and Anggit Dwi Hartanto, "Optimalisasi Akurasi Algoritma Naïve Bayes Dengan Metode Syntetic Minority Oversampling Technique (Smote) Pada Data Numerik," *Infotek J. Inform. dan Teknol.*, vol. 8, no. 1, pp. 217–227, 2025, doi: 10.29408/jit.v8i1.28340.
- [7] Heliyanti Susana, "Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet," *J. Ris. Sist. Inf. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2022, doi: 10.52005/jursistekni.v4i1.96.
- [8] 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.
- [9] S. Nuraeni, H. Harliana, and T. Prabowo, "Analisis Akurasi Naïve Bayes Dan Knn Dalam Penentuan Penerima Pkh Di Lombok Utara," *J. Inf. Syst. Manag.*, vol. 5, no. 2, pp. 121–126, 2024, doi: 10.24076/joism.2024v5i2.1205.
- [10] S. S. Prasetyowati and Y. Sibaroni, "Unlocking the potential of Naive Bayes for spatio temporal classification: a novel approach to feature expansion," *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-00958-x.
- [11] S. Ridho, "Analisis Preferensi Konsumen dalam Memilih Produk Hortikultura Menggunakan Metode Algoritma C45 dan Naive Bayes," *Emit. J. Tek. Elektro*, pp. 66–77, 2024, doi: 10.23917/emit.v24i1.2401.
- [12] J. P. Tanjung, F. C. Tampubolon, A. W. Panggabean, and M. A. A. Nandrawan, "Customer Classification Using Naive Bayes Classifier With Genetic Algorithm Feature Selection," *Sinkron*, vol. 8, no. 1, pp. 584–589, 2023, doi: 10.33395/sinkron.v8i1.12182.
- [13] A. C. Dewi, A. Hermawan, and D. Avianto, "Classification of Customers' Repeat Order Probability Using Decision Tree, Naïve Bayes and Random Forest," *J. Pilar Nusa Mandiri*, vol. 20, no. 1, pp. 52–59, 2024, doi: 10.33480/pilar.v20i1.5243.
- [14] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, pp. 1–14, 2024, doi: 10.1038/s41598-024-56706-x.
- [15] S. Sathyanarayanan and B. R. Tantri, "Confusion Matrix-Based Performance Evaluation Metrics," vol. 27, no. 4, 2024.
- [16] J. Allgaier and R. Pryss, "Cross-Validation Visualized: A Narrative Guide to Advanced Methods," pp. 1378–1388, 2024.
- [17] V. W. Lumumba, D. Kiprotich, N. G. Makena, M. D. Kavita, and M. L. Mpaine, "Comparative Analysis of Cross-Validation Techniques: LOOCV, K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models," vol. 13, no. 5, pp. 127–137, 2024.
- [18] C. Marceline Gunawan and T. Melinda, "Analisis Atribut Yang Menjadi Preferensi Konsumen Dalam Memilih Produk Ud Sumber Urip," *J. Indones. Sos. Sains*, vol. 2, no. 10, pp. 1655–1662, 2021, doi: 10.36418/jiss.v2i10.432.
- [19] M. S. Rohman, M. Lukitasari, D. A. Nugroho, N. Widodo, I. Kusumastuty, and N. I. P. Nugrahini, "Consumer Perceptions and Preferences of Beverage Products in Malang City," pp. 2614–2623, 2022.
- [20] A. Mubarak, "Analisis Perbandingan Kekuatan Material Sprocket Belakang Sepeda Motor Menggunakan Metode Elemen Hingga," *Pap. Knowl. . Towar. a Media Hist. Doc.*, vol. 3, no. April, pp. 49–58, 2021.
- [21] M. S. T. Raharjo, Basuki, M. A. Irfa'i, and M. M. Rosadi, "ANALISIS PENGARUH VARIASI UKURAN SPROCKET TERHADAP DAYA DAN TORSI MESIN SEPEDA MOTOR HONDA SUPRA X 125 TAHUN 2008," vol. 02, pp. 0–3, 2024.
- [22] H. Darmadi, M. Muttaqin, D. Paranita, and T. F. Simbolon, "ANALISA NILAI TEGANGAN PATAH RANTAI JENIS ROLLER AKIBAT ADANYA DEFORMASI," vol. 9, no. 2, pp. 50–55, 2023.