PENERAPAN DATA MINING DALAM KLASIFIKASI DATA TRANSAKSI PRODUK KOPERASI DI SMK PGRI 2 KARAWANG

M Yazid Zidane, Betha Nurina Sari, Iqbal Maulana, Aji Primaya, Garno
Informatika, Universitas Singaperbangsa Karawang
Universitas Singaperbangsa Karawang, Karawang Jawa Barat
1910631170202@student.unsika.ac.id

ABSTRAK

Koperasi SMK PGRI 2 Karawang menghasilkan data transaksi harian yang sering kali hanya terdokumentasi tanpa pemanfaatan lebih lanjut, meskipun data tersebut dapat digunakan untuk mengoptimalkan penjualan dan mengurangi kerugian. Penelitian ini bertujuan menerapkan algoritma Naive Bayes untuk mengklasifikasikan data transaksi koperasi dalam memprediksi keuntungan dan kerugian. Data terdiri dari 774 transaksi dari Januari 2023 hingga Agustus 2024, mencakup atribut seperti tanggal, nama produk, kategori, jumlah terjual, harga, dan total penjualan. Metodologi penelitian menggunakan Knowledge Discovery in Database (KDD), yang meliputi pemilihan data, preprocessing, transformasi, data mining, dan evaluasi model. Hasil evaluasi menunjukkan akurasi model Naive Bayes sebesar 98,97% pada test size 0,5 dengan F1-score 0,99 dalam mengklasifikasikan produk laku dan kurang laku, sehingga meningkatkan efisiensi pengelolaan transaksi koperasi.

Kata kunci : *Koperasi, Data Mining, Naïve Bayes, Klasifikasi, Transaksi*

1. PENDAHULUAN

Seiring perkembangan teknologi dan kebutuhan akan informasi, pengelolaan data menjadi aspek penting dalam mendukung kegiatan bisnis, termasuk di lingkungan koperasi sekolah. Koperasi sekolah, sebagai lembaga ekonomi yang beranggotakan siswa, memiliki peran ganda, yaitu memberikan pelatihan praktis tentang pengelolaan usaha sekaligus menyediakan kebutuhan siswa. Di SMK PGRI 2 Karawang, koperasi menghasilkan data transaksi harian yang terus meningkat seiring waktu, namun data ini belum dimanfaatkan secara optimal. Banyak transaksi yang tercatat hanya sebagai dokumentasi tanpa analisis lanjut yang dapat memberikan manfaat strategis bagi pengelolaan koperasi [1].

Data transaksi harian yang besar dan terus bertambah ini mengandung informasi penting yang dapat digunakan untuk mengoptimalkan penjualan dan menekan kerugian, terutama dalam pengelolaan stok produk yang efektif. Penggunaan teknologi data mining, khususnya metode klasifikasi, berpotensi untuk membantu koperasi dalam mengidentifikasi pola dan tren transaksi. Dengan menerapkan algoritma yang tepat, koperasi dapat melakukan prediksi terhadap produk yang memiliki kecenderungan laku atau tidak, sehingga dapat mengambil langkah strategis dalam pengelolaan produk dan stok.

Data mining adalah teknik yang menggunakan data dalam jumlah besar untuk menemukan informasi berharga yang sebelumnya tidak terlihat dan dapat digunakan dalam pengambilan keputusan penting. Dalam penelitian ini, data mining diterapkan pada transaksi koperasi SMK PGRI 2 Karawang untuk memprediksi apakah suatu produk akan laku atau tidak, sehingga dapat membantu koperasi mengoptimalkan strategi penjualan dan pengelolaan produk secara lebih efektif [2].

Algoritma Naive Bayes, yang menggunakan pendekatan probabilistik dalam klasifikasi, telah banyak digunakan dalam pengolahan data yang melibatkan klasifikasi kelas. Algoritma ini terkenal dengan kesederhanaan dan efisiensinya, terutama dalam mengolah data besar dengan banyak atribut. Dalam penelitian ini, algoritma Naive Bayes dipilih untuk mengklasifikasikan data transaksi di koperasi SMK PGRI 2 Karawang [3]. Melalui pendekatan Knowledge Discovery in Database (KDD), penelitian ini melibatkan serangkaian proses mulai dari pemilihan data, preprocessing, transformasi, hingga evaluasi model untuk memperoleh hasil prediktif yang dapat digunakan dalam pengambilan keputusan di koperasi [4].

Tujuan utama penelitian ini adalah untuk membantu koperasi dalam meningkatkan efisiensi pengelolaan transaksi, dengan cara memanfaatkan data transaksi yang selama ini hanya menjadi dokumentasi, sehingga menghasilkan informasi yang dapat diandalkan untuk mendukung strategi penjualan dan pengelolaan produk. Diharapkan hasil penelitian ini dapat memberikan kontribusi praktis bagi koperasi sekolah dalam pengelolaan data transaksi mereka.

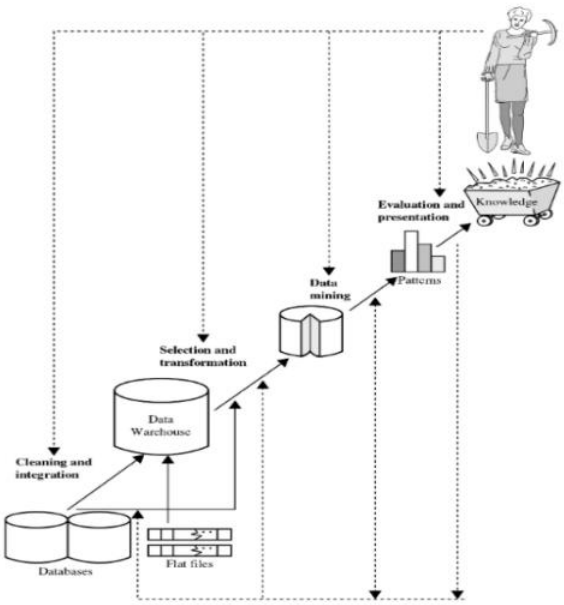
2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Dalam penelitian yang berjudul "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes (Studi Kasus: Makan Barbeque Sepuasnya)" penulis mengklasifikasikan penjualan makanan dan minuman menggunakan metodologi KDD dan algoritma Naïve Bayes. Dari sembilan skenario pengujian dengan cross-validation, performa terbaik diperoleh dengan menggunakan 3-fold, yang menghasilkan nilai akurasi sebesar 88,73%, precision sebesar 64,42%, recall sebesar 45,41%, dan nilai kappa sebesar 0,451 [5].

Dalam penelitian yang berjudul "Penerapan Data Mining Analisa Data Penjualan Obat Menggunakan Metode Random Forest" algoritma Random Forest digunakan untuk membangun pohon keputusan berdasarkan atribut data yang tidak teratur. Hasil penelitian menunjukkan bahwa obat yang paling cepat habis adalah Neurobion putih dengan persentase penjualan sebesar 60%, diikuti oleh Viks F44 dengan persentase 0,44% [6].

Penerapan algoritma Naive Bayes dalam asesmen calon debitur Koperasi Mitra Sejahtera memudahkan rekomendasi calon debitur yang berpotensi lunas atau tidak lunas, menggunakan tujuh parameter: usia, jenis kelamin, jumlah pinjaman, pekerjaan, penghasilan, jangka waktu pengembalian, dan status pelunasan. Hasil pengujian dengan 35 data testing menunjukkan akurasi 86%, berdasarkan perbandingan dengan data riil nasabah Desember 2019. Saran pengembangan selanjutnya adalah menambah jumlah data training untuk meningkatkan presisi dan mengkombinasikan metode klasifikasi lain dengan Naive Bayes [7].



Gambar 1. Tahapan Knowledge Discovery in Database

2.2. Koperasi

Koperasi sekolah adalah koperasi yang anggotanya terdiri dari siswa-siswa di berbagai jenjang pendidikan, mulai dari sekolah dasar hingga pendidikan menengah atas, serta lembaga pendidikan setara lainnya. Koperasi sekolah termasuk dalam kategori koperasi khusus, di mana meskipun tidak berbadan hukum, koperasi ini tetap dapat melakukan kegiatan ekonomi seperti jual beli. Tujuan utama dari koperasi sekolah adalah untuk mendukung pembelajaran teoretis yang diterima siswa dengan memberikan pengalaman praktis secara langsung. Melalui pengalaman tersebut, diharapkan siswa dapat memperoleh pemahaman yang lebih mendalam [8].

2.3. Data Mining

Data Mining adalah proses pengumpulan dan pengolahan data besar untuk menghasilkan informasi berguna, dikenal juga sebagai knowledge discovery in database (KDD), yang melibatkan langkah-langkah interaktif dan berulang. Teknik ini mencakup statistik, pembelajaran mesin, dan manajemen basis data [9].

Metode dalam data mining meliputi clustering, untuk mengelompokkan data berdasarkan kesamaan tanpa label. association, yang mencari pola hubungan antar item, seperti dalam analisis keranjang belanja, dan regression, untuk memprediksi nilai numerik berdasarkan hubungan antar variabel [10].

2.4. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah metode untuk mengolah atau menggali data guna memperoleh informasi dari basis data yang ada. Proses KDD melibatkan beberapa tahapan sebagai berikut:

2.4.1. Data Selection

Proses pemilihan data dari dataset yang relevan untuk analisis. Data yang tidak sesuai dengan kebutuhan analisis akan disaring sebelum tahap pengolahan lebih lanjut [11].

2.4.2. Data Preprocessing (Cleaning)

Tahap untuk membersihkan data dengan menghapus data yang mengganggu, memverifikasi data yang berubah-ubah, dan memperbaiki kesalahan dalam data [11].

2.4.3. Data Transformation

Proses mengubah data agar sesuai dengan format yang diperlukan oleh algoritma, seperti mengonversi data teks menjadi numerik, agar dapat diproses dalam data mining [11].

2.4.4. Data Mining

Proses inti untuk menggali pengetahuan dan menghasilkan model yang dapat memberikan informasi yang berguna [11].

2.4.5. Evaluasi

Tahap evaluasi model bertujuan untuk menilai kinerja model dalam memprediksi data yang belum terlihat, dengan mengukur akurasi serta menggunakan metrik lain seperti precision, recall, AUC, dan F1-score. Evaluasi ini juga membantu mengidentifikasi potensi bias atau overfitting. Jika hasil evaluasi belum optimal, tahap sebelumnya, seperti pemilihan algoritma atau preprocessing data, dapat diulang untuk meningkatkan performa model [11].

2.5. Algoritma Naïve Bayes

Naive Bayes adalah metode klasifikasi berbasis probabilitas yang sederhana, yang menghitung

probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari data yang tersedia [12].

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

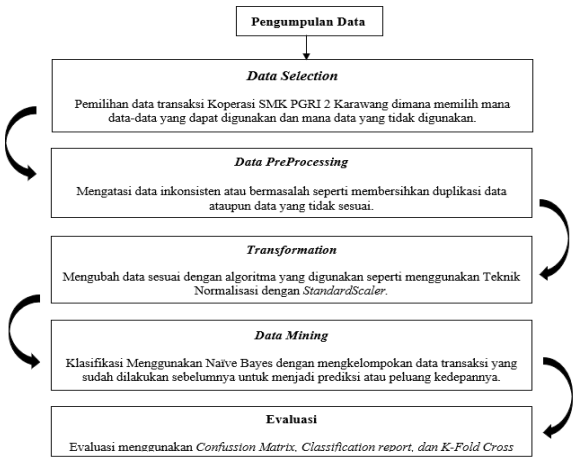
Keterangan :
X : Bukti data class yang belum diketahui
H : Hipotesis
P(H|X) : Probabilitas bahwa hipotesis H benar untuk bukti X
P(H) : Probabilitas prior hipotesis H
P(X|H) : Probabilitas bahwa hipotesis X benar untuk bukti H
P(X) : Probabilitas X

2.6. Google Colaboratory

Colab, yang dikembangkan oleh Google Research, adalah platform yang memungkinkan pengguna menulis dan menjalankan kode Python langsung dari browser tanpa perlu menginstal perangkat lunak. Platform ini gratis dan mudah diakses oleh siapa saja dengan koneksi internet. Colab menyediakan hosting untuk Jupyter Notebook dan akses GPU tanpa biaya tambahan, menjadikannya sangat berguna bagi pengguna dengan komputer berperforma rendah. Fitur ini memungkinkan komputasi berat dan proyek pembelajaran mesin tanpa memerlukan perangkat keras canggih [13].

3. METODE PENELITIAN

Metodologi penelitian yang diterapkan adalah Knowledge Discovery in Database (KDD). Tahapan dalam metodologi KDD meliputi pemilihan data, pra-pemrosesan, transformasi, penambangan data, dan evaluasi. Proses-proses dalam KDD dapat dijelaskan sebagai berikut:



Gambar 2. Metode Penelitian

3.1. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan melalui wawancara langsung dengan narasumber, yaitu guru yang bertanggung jawab atas koperasi di SMK PGRI 2 Karawang, yang dipastikan memiliki semua data terkait transaksi di koperasi.

3.2. Data Selection

Pada tahap ini, dilakukan pemilihan data yang akan digunakan dalam penelitian. Data diperoleh melalui wawancara dengan guru yang bertanggung jawab atas koperasi, kemudian dilakukan seleksi atribut sebelum melanjutkan ke tahap preprocessing.

3.3. Data Preprocessing

Tahap preprocessing bertujuan untuk mengubah data mentah menjadi data yang siap untuk diproses lebih lanjut. Proses ini melibatkan pembersihan data dengan menghapus data yang tidak konsisten atau data ganda (double), serta menghilangkan data yang tidak relevan atau tidak dapat digunakan.

3.4. Data Transformation

Pada tahap ini, data diubah menjadi format yang sesuai untuk pemrosesan lebih lanjut. Teknik yang digunakan adalah normalisasi, yang dilakukan untuk mengubah nilai fitur numerik ke dalam skala dengan rata-rata 0 dan standar deviasi 1 menggunakan StandardScaler dari scikit-learn.

3.5. Data Mining

Pada tahap ini, data yang telah diproses siap untuk diklasifikasikan. Penelitian ini menggunakan Google Colab dan algoritma Naive Bayes untuk mengklasifikasikan data transaksi koperasi di SMK PGRI 2 Karawang. Hasil klasifikasi ini nantinya digunakan untuk memprediksi transaksi produk di masa depan berdasarkan data transaksi sebelumnya.

3.6. Evaluasi

Tahap evaluasi menggunakan metode confusion matrix untuk mengukur akurasi, precision, dan recall. Akurasi mengukur sejauh mana data diklasifikasikan dengan benar. Precision menunjukkan proporsi prediksi positif yang benar, sementara recall mengukur proporsi kasus positif yang benar-benar diprediksi positif.

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Tahap ini dimulai dengan pemilihan data transaksi yang dianggap relevan untuk penelitian dari koperasi SMK PGRI 2 Karawang.

Tabel 1. Dataset Awal

Produk	Terjual	Harga	Laku
Topi	2	30000	Kurang laku
Susu	3	5000	Kurang laku
Kue	13	2500	Laku
Pulpen	14	4000	Laku
Pensil	1	3000	Kurang laku
Kopi	8	5000	Laku
Roti	11	3000	Laku
Penggaris	3	4000	Kurang Laku
Jus	10	12000	Laku

Data yang tidak relevan atau mengandung kesalahan telah dihapus untuk memastikan kualitas

data yang digunakan. Semua atribut yang ada dianggap relevan dengan penelitian, sehingga tidak ada atribut yang dihapus.

Tabel 2. Deskripsi Dataset

No.	Atribut	Keterangan	Tipe
1	Produk	Produk yang dibeli dalam transaksi (Topi, Pulpen, Roti)	Kategorik
2	Jumlah	Jumlah unit produk yang terjual dalam transaksi	Numerik
3	Harga	Harga per unit dari produk yang terjual	Numerik
4	Laku	Kategori laku atau tidaknya suatu produk.	Kategorik

4.2. Data Preprocessing

Pada tahap ini, data yang telah dipilih dibersihkan dan dinormalisasi untuk mengatasi inkonsistensi atau kesalahan, seperti nilai yang hilang, duplikasi, dan penyesuaian format agar sesuai dengan analisis menggunakan algoritma Naive Bayes.

4.2.1. Tipe Dataset

Pemilihan tipe dataset yang tepat sangat penting karena metode analisis dan pemrosesan data yang digunakan akan bervariasi tergantung pada tipe data yang disajikan. Misalnya, data numerik sering kali memerlukan teknik statistik tertentu atau transformasi matematis, sementara data kategorikal mungkin memerlukan pendekatan seperti encoding atau pembobotan. Memahami tipe data memungkinkan pemilihan teknik yang tepat untuk analisis, memastikan hasil yang lebih akurat dan relevan.

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 774 entries, 0 to 773
Data columns (total 7 columns):
 #   Column           Non-Null Count  Dtype 
---  ------           --------------  ----- 
 0   Tanggal          774 non-null    object
 1   Nama Produk      774 non-null    object
 2   Kategori Produk  774 non-null    object
 3   Jumlah Terjual   774 non-null    int64 
 4   Harga per Unit   774 non-null    int64 
 5   Total Penjualan  774 non-null    int64 
 6   laku             774 non-null    object
dtypes: int64(3), object(4)
memory usage: 42.5+ KB
```

Gambar 4. Tipe Dataset

4.2.2. Data Duplikat

Pada tahap ini, dilakukan pengecekan terhadap data duplikat dalam dataset, yaitu data yang memiliki nilai identik di seluruh atribut pada satu atau lebih baris. Kehadiran data duplikat dapat mempengaruhi hasil analisis dan model yang dibangun, berisiko menyebabkan bias atau overfitting. Proses ini dilakukan menggunakan fungsi ‘duplicated()’ untuk memeriksa apakah ada baris yang memiliki nilai yang sama di semua kolom.

```
data.duplicated().sum()
```

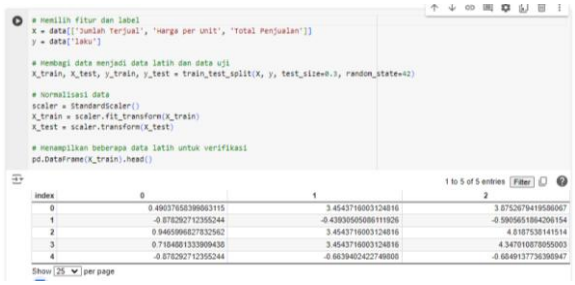
```
4
```

Gambar 5. Data Duplikat

Hasil pemeriksaan menunjukkan adanya 4 data duplikat dalam dataset, yang tercatat pada transaksi dengan tanggal, nama produk, dan jumlah yang sama. Duplikasi ini disebabkan oleh transaksi serupa yang terjadi pada hari yang sama, yang dapat terjadi secara alami dalam operasi koperasi. Karena duplikasi ini mencerminkan transaksi yang sah dan umum, tidak dilakukan penghapusan atau penanganan khusus terhadap data tersebut. Dengan demikian, dataset dibiarkan seperti aslinya untuk mempertahankan keutuhan data yang relevan bagi analisis selanjutnya.

4.3. Data Transformation

Pada tahap ini, data yang telah dibersihkan dinormalisasi agar fitur numerik (Jumlah Terjual, Harga per Unit, dan Total Penjualan) memiliki skala seragam, dengan rata-rata 0 dan standar deviasi 1, menggunakan StandardScaler dari scikit-learn. Ini mencegah fitur dengan rentang nilai lebih besar mendominasi pemodelan.



Gambar 6. Normalisasi Data

Hasil normalisasi menunjukkan bahwa data kini memiliki skala seragam, sehingga siap digunakan dalam tahap pemodelan. Proses ini penting agar setiap fitur memiliki bobot yang setara, menghindari dominasi dari fitur dengan skala lebih besar. Dengan data yang telah dinormalisasi, algoritma Naive Bayes dapat bekerja lebih efektif dan menghasilkan prediksi yang lebih akurat.

4.4. Data Mining

Pada tahap ini, model Naive Bayes diterapkan untuk mengelompokkan data transaksi koperasi SMK PGRI 2 Karawang, dengan memanfaatkan tiga fitur utama: Jumlah Terjual, Harga per Unit, dan Total Penjualan. Untuk menguji performa model, dataset dibagi menjadi data latih dan data uji dengan berbagai ukuran test size, yaitu 0.1, 0.2, 0.3, 0.4, dan 0.5. Setiap skenario test size diuji untuk menilai keakuratan model melalui metrik-metrik seperti akurasi, classification report, dan visualisasi confusion matrix. Evaluasi ini bertujuan untuk mengetahui sejauh mana model Naive

Bayes mampu mengklasifikasikan data transaksi dengan akurat berdasarkan variasi data uji yang digunakan.

4.4.1. Skenario Test Size 0,1

Pada skenario ini, dataset dibagi menjadi 90% data pelatihan dan 10% data pengujian. Model yang diterapkan menghasilkan akurasi sebesar 93,59%, menunjukkan kemampuan model yang baik dalam mengklasifikasikan data pengujian dengan benar.

```
Accuracy with test size 0.1: 93.59%
Classification Report with test size 0.1:
              precision    recall  f1-score   support

 kurang laku       0.91      0.98      0.94        41
       laku        0.97      0.89      0.93        37

    accuracy                           0.94        78
   macro avg       0.94      0.93      0.94        78
weighted avg       0.94      0.94      0.94        78
```

Gambar 7. Hasil Skenario 1

4.4.2. Skenario Test Size 0,2

Dengan test size 0.2, dataset dibagi menjadi 80% data pelatihan dan 20% data pengujian. Model menghasilkan akurasi sebesar 92,90%, menunjukkan performa yang baik dalam mengklasifikasikan data pengujian. Ini menunjukkan bahwa model dapat mempelajari pola dari data pelatihan dan menerapkannya dengan efektif pada data yang belum dilihat sebelumnya.

```
Accuracy with test size 0.2: 92.90%
Classification Report with test size 0.2:
              precision    recall  f1-score   support

 kurang laku       0.88      0.99      0.93        76
       laku        0.99      0.87      0.93        79

    accuracy                           0.93       155
   macro avg       0.93      0.93      0.93       155
weighted avg       0.94      0.93      0.93       155
```

Gambar 8. Hasil Skenario 2

4.4.3. Skenario Test Size 0,3

Dengan test size 0.3, dataset dibagi menjadi 70% data pelatihan dan 30% data pengujian. Model menghasilkan akurasi sebesar 97,85%, menunjukkan kinerja yang sangat baik dalam mengklasifikasikan data pengujian.

```
Accuracy with test size 0.3: 97.85%
Classification Report with test size 0.3:
              precision    recall  f1-score   support

 kurang laku       0.97      0.98      0.98       111
       laku        0.98      0.98      0.98       122

    accuracy                           0.98       233
   macro avg       0.98      0.98      0.98       233
weighted avg       0.98      0.98      0.98       233
```

Gambar 9. Hasil Skenario 3

4.4.4. Skenario Test Size 0,4

Pada skenario ini, dataset dibagi menjadi 60% data training dan 40% data testing. Model mencapai akurasi sebesar 97,42%, menunjukkan bahwa model mampu mengklasifikasikan data pengujian dengan sangat baik.

```
Accuracy with test size 0.4: 97.42%
Classification Report with test size 0.4:
              precision    recall  f1-score   support

 kurang laku       0.97      0.97      0.97       151
       laku        0.97      0.97      0.97       159

    accuracy                           0.97       310
   macro avg       0.97      0.97      0.97       310
weighted avg       0.97      0.97      0.97       310
```

Gambar 10. Hasil Skenario 4

4.4.5. Skenario Test Size 0,5

Pada skenario ini, dataset dibagi menjadi 50% data pelatihan dan 50% data pengujian. Model menghasilkan akurasi sebesar 98,97%, menunjukkan performa yang sangat baik dalam mengklasifikasikan data pengujian. Pembagian yang seimbang ini memungkinkan model untuk belajar secara optimal dan menguji kemampuannya dengan jumlah data yang cukup besar.

```
Accuracy with test size 0.5: 98.97%
Classification Report with test size 0.5:
              precision    recall  f1-score   support

 kurang laku       1.00      0.98      0.99       190
       laku        0.98      1.00      0.99       197

    accuracy                           0.99       387
   macro avg       0.99      0.99      0.99       387
weighted avg       0.99      0.99      0.99       387
```

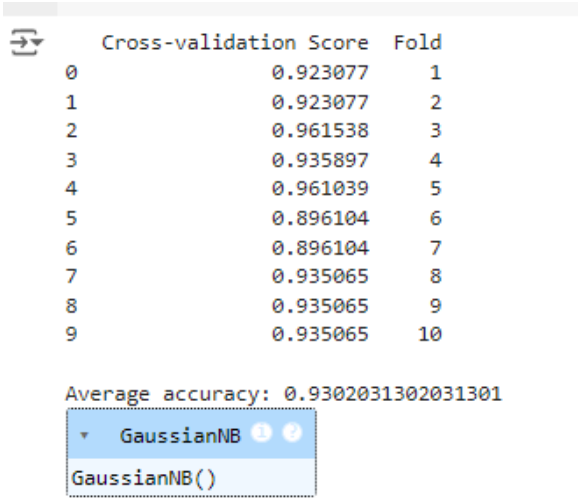
Gambar 11. Hasil Skenario 5

4.5. Evaluasi

Evaluasi model dilakukan untuk mengukur performa algoritma Naive Bayes, dengan pendekatan seperti cross-validation, metrik precision, recall, dan F1-Score, serta analisis pengaruh ukuran test set. Hasil evaluasi ini memberikan gambaran tentang kemampuan model dalam memprediksi data laku dan kurang laku.

4.5.1. Hasil Cross-Validation

Penelitian ini menggunakan 10-fold cross-validation, membagi dataset menjadi 10 bagian untuk pelatihan dan pengujian bergantian. Hasilnya, akurasi rata-rata model Naive Bayes adalah 0,93, menunjukkan kestabilan dan akurasi yang konsisten meskipun data uji bervariasi di setiap fold.



	Cross-validation Score	Fold
0	0.923077	1
1	0.923077	2
2	0.961538	3
3	0.935897	4
4	0.961039	5
5	0.896104	6
6	0.896104	7
7	0.935065	8
8	0.935065	9
9	0.935065	10

Average accuracy: 0.9302031302031301

GaussianNB

GaussianNB()

Gambar 12. Hasil K-Fold Cross-Validation

Hasil pemeriksaan menunjukkan bahwa akurasi model bervariasi pada setiap fold, dengan akurasi terendah 0,896 dan tertinggi 0,961. Variasi ini disebabkan oleh penggunaan data uji yang berbeda di setiap fold, namun rata-rata akurasi tetap tinggi, yaitu 0,93. Hal ini menunjukkan bahwa model memiliki performa yang baik dan tidak terlalu terpengaruh oleh pembagian data uji yang berbeda, serta menunjukkan kestabilan dalam memprediksi hasil di berbagai kondisi.

4.5.2. Evaluasi Metrik Precision, Recall, dan F1-Score

Evaluasi model juga dilakukan dengan metrik precision, recall, dan F1-Score untuk menilai kinerja dalam mengenali data laku dan kurang laku. Model dengan test size 0.5, yang memiliki akurasi tertinggi, dievaluasi menggunakan metrik-metrik tersebut dalam classification report.

Tabel 3. Evaluasi Metrik Precision, Recall, dan F1-Score

Kelas	***Precision***	***Recall***	**F1-*Score***
Kurang Laku	1.00	0.98	0.99
Laku	0.98	1.00	0.99

Hasil evaluasi menunjukkan precision 1.00 pada kelas kurang laku, artinya semua prediksi sebagai kurang laku benar. Recall 1.00 pada kelas laku menunjukkan model dapat menemukan semua instance laku. Precision 0.98 pada kelas laku mengindikasikan sedikit kesalahan dalam mengklasifikasikan produk yang tidak laku sebagai laku. Dengan F1-Score 0.99 pada kedua kelas, model Naive Bayes menunjukkan performa yang sangat baik dalam mengklasifikasikan produk ke dalam kategori laku dan kurang laku.

4.5.3. Pengaruh Ukuran Test Size pada Evaluasi

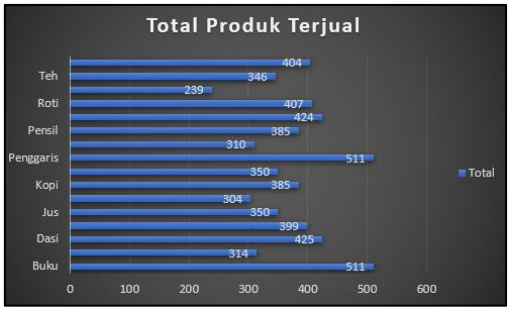
Selanjutnya, untuk memahami pengaruh ukuran test size terhadap hasil evaluasi, dilakukan pengujian dengan test size mulai dari 0.1 hingga 0.5. Ukuran test size mempengaruhi persentase data untuk pengujian, sementara sisanya digunakan untuk pelatihan model. Hasil pengujian menunjukkan perbedaan dalam akurasi berdasarkan variasi test size.

Gambar 13. Diagram akurasi dari setiap test size

Dapat dilihat bahwa akurasi model meningkat seiring dengan bertambahnya ukuran test size, dari 93.59% pada test size 0.1 hingga 98.97% pada test size 0.5. Namun, meskipun test size 0.5 memberikan akurasi tertinggi, cross-validation memberikan hasil yang lebih stabil dan akurat karena menguji model dengan variasi data uji di setiap fold.

4.6. Pembahasan

Penelitian ini menggunakan algoritma Naive Bayes untuk mengklasifikasikan data transaksi koperasi SMK PGRI 2 Karawang ke dalam kategori "laku" dan "kurang laku" melalui tahapan Knowledge Discovery in Database (KDD). Model diuji dengan berbagai test size (0.1–0.5), dengan akurasi tertinggi 98.97% pada test size 0.5. Evaluasi menggunakan 10-fold cross-validation menghasilkan akurasi rata-rata 93%. Hasil menunjukkan precision 1.00 untuk "kurang laku" dan 0.98 untuk "laku", dengan recall 1.00 untuk "laku" dan 0.98 untuk "kurang laku". F1-score 0.99 menunjukkan performa model yang sangat baik dalam klasifikasi.



Gambar 14. Grafik total penjualan produk

Berdasarkan grafik total penjualan, berikut adalah temuan utama:

a. **Buku dan Penggaris**: Terjual masing-masing 511 unit, menunjukkan permintaan tinggi untuk produk alat tulis sekolah.
b. **Dasi**: Terjual 425 unit, menunjukkan kebutuhan kuat untuk produk seragam sekolah.
c. **Pensil**: Terjual 424 unit, menunjukkan permintaan stabil untuk alat tulis.

d. **Teh**: Terjual 404 unit, menjadi produk konsumsi terlaris.
e. **Kopi**: Terjual 385 unit, meskipun masih di bawah teh, tetap menunjukkan minat konsumen.
f. **Jus**: Terjual 350 unit, menunjukkan permintaan stabil di kategori minuman.
g. **Roti**: Terjual 239 unit, dengan penjualan terendah, kemungkinan memerlukan strategi pemasaran lebih kuat.

Produk kebutuhan sekolah mendominasi penjualan, sementara roti memiliki penjualan lebih rendah yang perlu diperhatikan.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa penerapan algoritma Naïve Bayes dalam mengklasifikasikan data transaksi koperasi efektif untuk memprediksi apakah suatu produk akan laku atau tidak. Model ini berhasil mengidentifikasi pola penjualan berdasarkan fitur seperti jumlah produk terjual dan kategori produk dengan akurasi mencapai 98%. Evaluasi menunjukkan bahwa model memiliki performa yang baik, dengan metrik presisi dan recall yang menunjukkan keandalannya dalam mengklasifikasikan produk yang laku.

Namun, untuk meningkatkan akurasi, disarankan untuk mengembangkan model dengan menggunakan algoritma lain seperti Decision Tree atau Random Forest, yang lebih robust dalam menangani kompleksitas data. Selain itu, pengumpulan data yang lebih besar serta penambahan fitur relevan seperti promosi atau jenis pelanggan dapat meningkatkan kualitas model, memberikan prediksi yang lebih akurat untuk menentukan produk yang laku di masa depan.

DAFTAR PUSTAKA

[1] T. I. Verelladevanka Adryamarthanino, "Sejarah Perkembangan Koperasi Di Indonesia," 2022, [Online]. Available: https://www.kompas.com/stori/read/2022/12/21/110000179/sejarah-perkembangan-koperasi-di-indonesia

[2] R. I. Borman and M. Wati, "Penerapan Data Maining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes," *J. Ilm. Fak. Ilmu Komput.*, vol. 09, no. 01, pp. 25–34, 2020.

[3] A. Nurian, T. N. Padilah, and G. Garno, "Analisis Sentimen Terhadap Pelayanan Disdukcapil Karawang Menggunakan Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, 2024, doi: 10.23960/jitet.v12i2.4178.

[4] A. N. Aida, P. Arsi, R. P. Aji, and T. Tarwoto, "Analisis Sentimen Pengguna Aplikasi Instagram Pada Situs Google Play Menggunakan Metode Naïve Bayes," *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 704, 2024, doi: 10.30865/mib.v8i2.7388.

[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. 02, pp. 75–81, 2021, doi: 10.33884/jif.v9i02.3755.

[6] C. Reka, "Bulletin of Data Science Penerapan Data Mining Analisa Data Penjualan Obat Menggunakan Metode Random Forest," *Media Online)*, vol. 1, no. 3, p. 117, 2022, [Online]. Available: https://ejurnal.seminar-id.com/index.php/bulletinds

[7] I. G. T. Isa, "Aplikasi Asesmen Calon Debitur menggunakan Naive Bayes di Koperasi Mitra Sejahtera SMK Negeri 1 Kota Sukabumi," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 10, no. 1, pp. 31–39, 2021, doi: 10.32736/sisfokom.v10i1.1013.

[8] E. Widati and M. Herawati, "Pemahaman dan Kesadaran Pentingnya Koperasi Sekolah di SMK Nusa Bhakti Depok," *KANGMAS Karya Ilm. Pengabdi. Masy.*, vol. 1, no. 2, pp. 57–66, 2020, doi: 10.37010/kangmas.v1i2.40.

[9] A. Andika, S. Syarli, and C. R. Sari, "Data Mining Klasifikasi Kelulusan Mahasiswa Menggunakan Metode Naïve Bayes," *J. Peqguruang Conf. Ser.*, vol. 4, no. 1, p. 423, 2022, doi: 10.35329/jp.v4i1.2358.

[10] F. Solikhah, M. Febianah, A. L. Kamil, W. A. Arifin, and Shelly Janu Setyaning Tyas, "Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan," *Tematik*, vol. 8, no. 1, pp. 96–103, 2021, doi: 10.38204/tematik.v8i1.576.

[11] N. Syahfitri, E. Budianita, A. Nazir, and I. Afrianty, "Pengelompokan Produk Berdasarkan Data Persediaan Barang Menggunakan Metode Elbow dan K-Medoid," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 3, pp. 1668–1675, 2023, doi: 10.30865/klik.v4i3.1525.

[12] A. Triayudi and Sumiati, "Implementasi Klasifikasi Data Mining Untuk Penentuan Kelayakan Pemberian Kredit dengan Menggunakan Algoritma Naïve Bayes," *J. Sist. Komput. dan Inform. Hal 240−*, vol. 244, no. 1, pp. 240–244, 2022, doi: 10.30865/json.v4i1.4653.

[13] I. Orobor and N. Obi, "Machine Learning Pipeline for Multi-Class Text Classification," *J. Eng. Technol. Appl. Sci.*, vol. 7, no. 2, pp. 64–69, 2022, [Online]. Available: https://www.researchgate.net/publication/362091493_Machine_Learning_Pipeline_for_Multi-Class_Text_Classification