



Implementasi Data Mining untuk Analisis Data Penjualan dengan Menggunakan Algoritma *Naïve Bayes* (Studi Kasus : KPRI Kokarnaba Baturraden)

Lifa Sholiah¹, Ito Setiawan², Abdillah Teguh Permana³, Iqbal Yusuf Azhari⁴,
Wakhid Sayudha Rendra Graha Alrashid⁵

¹⁻⁵ Universitas Amikom Purwokerto, Indonesia

21SA2073@mhs.amikompurwokerto.ac.id¹, itosetiawan@amikompurwokerto.ac.id²,

4bdillahteguh@gmail.com³, 21sa2046@mhs.amikompurwokerto.ac.id⁴,

21sa2028@mhs.amikompurwokerto.ac.id⁵

Alamat: Jl. Letjend Pol. Soemarto No. 127, Watumas, Purwanegara, Kec. Purwokerto Utara,
Kabupaten Banyumas, Jawa Tengah

Korespondensi penulis: 21SA2073@mhs.amikompurwokerto.ac.id

Abstract. KPRI KOKARNABA Baturraden faces challenges in managing increasingly complex sales data, particularly in identifying the most in-demand products to maximize profit. This study aims to analyze sales patterns using the *Naïve Bayes* algorithm as a probability-based classification method. The collected sales data were analyzed to identify categories of best-selling and less popular products within the cooperative. The results indicate that the *Naïve Bayes* algorithm has an accuracy rate of 77.56% in predicting product categories. This research is expected to assist the cooperative in optimizing stock management and improving member satisfaction.

Keywords: *Naïve Bayes Algorithm, Sales Prediction, Stock Management.*

Abstrak. KPRI KOKARNABA Baturraden menghadapi tantangan dalam pengelolaan data penjualan yang terus meningkat, terutama dalam mengidentifikasi produk yang paling diminati untuk memaksimalkan profit. Penelitian ini bertujuan untuk menganalisis pola penjualan menggunakan algoritma *Naïve Bayes* sebagai metode klasifikasi berbasis probabilitas. Data penjualan yang terkumpul dianalisis untuk mengidentifikasi kategori produk terlaris dan kurang diminati di koperasi. Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* memiliki tingkat akurasi sebesar 77,56% dalam memprediksi kategori produk. Penelitian ini diharapkan dapat membantu koperasi dalam mengoptimalkan pengelolaan stok serta meningkatkan kepuasan anggota.

Kata kunci: Algoritma *Naïve Bayes*, Manajemen Stok, Prediksi Penjualan.

1. LATAR BELAKANG

Kemajuan teknologi informasi telah memberikan dampak signifikan terhadap berbagai sektor, termasuk sektor koperasi yang berperan penting dalam perekonomian local (Sinta & Zulfa Naftali, 2024). Koperasi Pegawai Republik Indonesia (KPRI) KOKARNABA Baturraden, yang berdiri pada 6 Juni 1977, berlokasi di Jalan Raya Baturraden, Desa Kemutug Lor, Kecamatan Baturraden, Kabupaten Banyumas, Jawa Tengah. Sebagai salah satu unit usaha di bawah Badan Penelitian dan Pengembangan Pertanian (Balitbangtan), koperasi ini memiliki peran penting dalam melayani kebutuhan pegawai Balai Besar Pembibitan Ternak Unggul dan Hijauan Pakan Ternak (BBPTUHPT), yang merupakan pelanggan utamanya. Dengan menyediakan berbagai produk kebutuhan sehari-hari yang lengkap dan harga yang

terjangkau, KPRI KOKARNABA berperan strategis dalam mendukung kesejahteraan pegawai sekaligus memperkuat ekosistem ekonomi lokal di sekitarnya.

Sebagai koperasi yang melibatkan banyak transaksi setiap harinya, pengelolaan data penjualan menjadi salah satu aspek penting untuk meningkatkan kinerja koperasi, baik dari segi efisiensi operasional maupun kepuasan anggota. dengan semakin bertambahnya jumlah data penjualan, KPRI KOKARNABA menghadapi tantangan dalam memahami dan menganalisis pola-pola yang ada di balik data tersebut. Untuk itu dari berbagai produk yang ada, akan penting untuk menentukan produk terlaris yang dijual agar dapat memaksimalkan profit dari penjualan dengan data mining (Nugraha et al., 2022).

Data mining adalah proses penggalian informasi berharga dari kumpulan data besar dengan tujuan mengidentifikasi pola, tren, atau hubungan yang tersembunyi (Saputra & Iskandar, 2023). Salah satu metode yang akan digunakan adalah metode naïve bayes sebuah metode klasifikasi yang berbasis pada Teorema Bayes (Muafa, 2022). Algoritma ini bekerja dengan menghitung probabilitas bersyarat untuk memprediksi kategori atau kelas suatu data, dengan asumsi bahwa setiap fitur dalam dataset bersifat independent (Khafifah et al., 2024). Keunggulan algoritma ini terletak pada kemampuannya untuk menangani data yang tidak terstruktur serta kemudahan dalam implementasi dan interpretasi hasil, sehingga sering digunakan dalam berbagai aplikasi, termasuk analisis data penjualan, deteksi spam, dan pengenalan pola perilaku pengguna (Santosa & Novianti, 2024).

Adapun penelitian sebelumnya yang dilakukan oleh (Prasetyo et al., 2020) dengan judul “Implementasi Data Mining Untuk Analisis Data Penjualan Dengan Menggunakan Algoritma Apriori (Studi Kasus Dapoerin's)” penelitian ini melakukan analisis data transaksi menggunakan data mining dengan metode algoritma apriori, menggunakan algoritma tersebut dapat diketahui produk kudapan yang paling banyak terjual, sehingga dapat menyiapkan persediaan bahan baku yang diprioritaskan dan menyusun strategi memasarkan produk dengan jenis kudapan yang lain dengan mengkaji keunggulan produk kudapan satu dengan dan lainnya yang sering banyak terjual.

Untuk mengatasi masalah tersebut, metode Naïve Bayes dapat diterapkan untuk menentukan produk terlaris berdasarkan data historis. Algoritma Naïve Bayes bekerja dengan menghitung probabilitas bersyarat dari berbagai faktor, seperti pola perilaku belanja, frekuensi transaksi, serta preferensi produk (Pebdika et al., 2023). Dengan prediksi yang lebih akurat, koperasi dapat melakukan optimalisasi stok barang dengan fokus pada produk yang paling diminati serta merancang strategi pencegahan churn, seperti memberikan insentif khusus atau program loyalitas bagi anggota yang berisiko.

Dengan demikian, penerapan metode Naïve Bayes diharapkan dapat memberikan solusi yang efektif bagi KPRI KOKARNABA dalam mengelola stok secara lebih efisien dan tepat sasaran. Melalui pemahaman pola penjualan yang lebih mendalam, koperasi dapat merumuskan strategi yang bukan hanya mampu meningkatkan profitabilitas, tetapi juga memperkuat keterlibatan anggota, menjaga loyalitas, serta mendukung keberlanjutan operasional koperasi di masa mendatang.

2. KAJIAN TEORITIS

No.	Nama Peneliti dan Tahun	Masalah	Metode	Hasil
1.	Prasetyo et al., (2020)	Penjualan produk kudapan pada DAPOERIN'S yang belum optimal.	Analisis data transaksi menggunakan data mining dengan metode algoritma apriori.	Dengan algoritma tersebut dapat diketahui produk kudapan yang paling banyak terjual,
2.	Pranata & Utomo, (2020)	Penumpukan data penjualan alat bengkel yang selama ini tidak diketahui kegunaannya.	Analisis data penjualan menggunakan data mining dengan metode algoritma FP-Growth	Dengan algoritma tersebut dapat diketahui persediaan produk yang harus ditingkatkan.
3.	Andarista & Jananto, (2022)	Masih terdapat kendaraan yang tidak layak jalan dan tidak lulus uji.	Analisis data menggunakan metode klasifikasi dengan algoritma C4.5	Dengan algoritma tersebut dapat diketahui pengaruh paling besar yaitu atribut kedalaman alur ban.

Algoritma Naive Bayes adalah metode klasifikasi berbasis probabilitas yang didasarkan pada Teorema Bayes. Algoritma ini mengasumsikan bahwa setiap fitur atau variabel bebas dan tidak saling mempengaruhi, sehingga disebut "naive" atau "sederhana." Dengan pendekatan ini bertujuan untuk memahami pola penjualan dan menentukan produk terlaris berdasarkan data historis serta untuk memprediksi kemungkinan anggota berhenti menggunakan layanan koperasi.

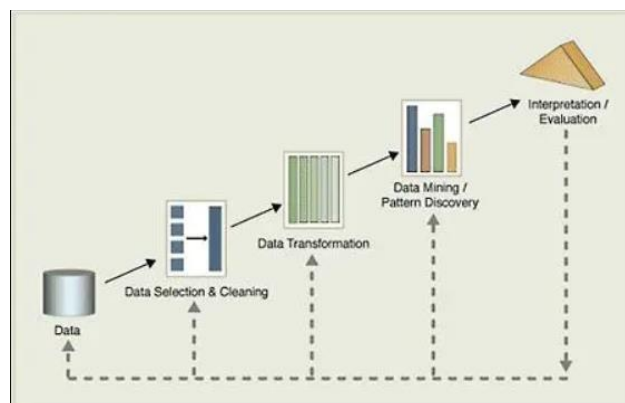
Penelitian ini berbeda dari penelitian sebelumnya karena menggunakan metode data mining dengan algoritma Naive Bayes khusus untuk menganalisis data penjualan di KPRI Kokarnaba Baturraden. Pendekatan ini diharapkan dapat memberikan wawasan baru yang lebih spesifik terhadap pola pembelian anggota koperasi, terutama dalam konteks kebutuhan

sehari-hari para pegawai BBPTUHPT, yang merupakan pelanggan utama. Melalui penelitian ini, hasil analisis diharapkan dapat membantu koperasi dalam memahami preferensi pelanggan serta mengoptimalkan pengelolaan persediaan produk.

3. METODE PENELITIAN

Bagian ini memuat rancangan penelitian meliputi disain penelitian, populasi/ sampel penelitian, teknik dan instrumen pengumpulan data, alat analisis data, dan model penelitian yang digunakan. Metode yang sudah umum tidak perlu dituliskan secara rinci, tetapi cukup merujuk ke referensi acuan (misalnya: rumus uji-F, uji-t, dll). Pengujian validitas dan reliabilitas instrumen penelitian tidak perlu dituliskan secara rinci, tetapi cukup dengan mengungkapkan hasil pengujian dan interpretasinya. Keterangan simbol pada model dituliskan dalam kalimat.

Metode penelitian yang digunakan adalah deskriptif analisis, yaitu sebuah metodologi yang bertujuan untuk menggambarkan atau mendeskripsikan objek berdasarkan data atau sampel yang telah terkumpul sebagaimana adanya, tanpa melakukan analisis mendalam atau menyusun kesimpulan yang bersifat umum. Langkah-langkah yang dilaksanakan berpedoman pada proses Knowledge Discovery in Data (KDD) sebagai berikut (Juanda Saputra & Izman Herdiansyah, 2022) :



Gambar.1

a) Data Selection

Data seleksi diambil dari data stok barang toko sembako, yang mencakup atribut dan stok barang. Data ini dipilih sesuai dengan persyaratan Knowledge Discovery in Database (KDD), sehingga tidak semua atribut yang diperoleh saat pengumpulan data digunakan nantinya (Zafira et al., 2024).

b) Pre-Processing

Sebelum proses data mining, ada proses pembersihan data, yang akan menjadi fokus KDD. Proses pembersihan data mencakup hal-hal seperti menghilangkan duplikat data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan data, seperti kesalahan cetak(Sultan Iskandar et al., 2020).

c) Data Transformation

Proses ini mengubah atau mengombinasikan data agar lebih sesuai untuk kegiatan penambangan dengan merangkum (agregasi) data tersebut(Anggriani et al., 2021).

d) Data Mining

Proses Data Mining adalah proses menemukan pola atau informasi menarik dalam data yang telah dipilih dengan menerapkan teknik, metode, atau algoritma tertentu sesuai dengan tujuan keseluruhan dari proses KDD(Anggriani et al., 2021).

e) Evaluation

Proses data mining menghasilkan pola informasi yang perlu disajikan dalam format yang mudah dipahami oleh pihak terkait. Ini merupakan tahap akhir dari KDD, bertujuan untuk memastikan bahwa pola atau informasi yang ditemukan tidak bertentangan dengan fakta atau hipotesis yang sudah ada sebelumnya(Sultan Iskandar et al., 2020).

Teorema Bayes menyediakan metode untuk menghitung probabilitas suatu hipotesis berdasarkan informasi yang ada. Persamaan Naïve Bayes ditunjukkan pada Persamaan 3 (Arfan & Paraga, 2024):

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

Keterangan

X = Sampel data yang memiliki Class (label) yang tidak diketahui

C = Hipotesis bahwa X adalah data Class (label)

P(C) = Probabilitas Hipotesis C

P(X) = Peluang dari data sampel yang diamati (probabilitas C)

P(X|C) = Probabilitas berdasarkan kondisi pada hipotesis

4. HASIL DAN PEMBAHASAN

Pengumpulan Data Pada penelitian ini menggunakan pengumpulan data penjualan KPRI KOKARNABA. Data yang diperoleh menunjukkan bahwa atribut yang dapat mempengaruhi penjualan adalah quantity, harga dan jumlah. Pada atribut quantity

merupakan banyak barang yang ada, atribut harga merupakan atribut yang menunjukkan harga barang tersebut terjual, sedangkan jumlah merupakan banyaknya jumlah barang yang terjual pada KPRI KOKARNABA. Dengan demikian data penjualan yang diperoleh tersebut akan dilakukan pengolahan dengan teknik klasifikasi data mining yang bertujuan untuk memprediksi penjualan dengan kategori barang yang laris dan tidak laris terjual. Dengan prediksi tersebut KPRI KOKARNABA dapat memperkirakan stok barang yang harus disediakan pada setiap jenis barang, juga untuk meminimalisir kerugian pada waktu yang akan datang.

No.	Kode Barang	Nama Barang	Satuan	Qty	Harga	Jumlah	
0	1 (90)MD273731084097(9	NaN	NaN	1	0	0	
1	2	0	NaN	NaN	1	0	
2	3	08968604170	INDOMIE GORENG JUMBO	PCS	27	5500	148500
3	4	08968604366	INDOMIE GORENG KEBAB RENDANG	PCS	2	3500	7000
4	5	762220170422	OREO SUMMER LTD 123.5G	PCS	5	8500	42500
...	
1754	1763	899305364109	PASEO BABY WIPES 50S	PCS	3	7500	22500
1755	1764	899305364109	PASEO BABY WIPES 50S	PCS	9	8000	72000
1756	1765	899305364110	PASEO TISU BASAH 25S	PCS	4	6000	24000
1757	1766	899305364110	PASEO TISU BASAH 25S	PCS	8	6500	52000
1758	1767	899701958164	GREEN SOOF 250	PCS	16	8000	128000

1759 rows x 7 columns

Gambar.2

Dataset yang ada berjumlah 1759 data dimana akan dilakukan data selection dan pre processing.

Data Pre Processing

```
[ ] data_bersih = df_cleaned.dropna()

# Menampilkan data setelah menghapus data NaN
print("\nData setelah menghapus data NaN:")
data_bersih

# Anda dapat menyimpan DataFrame yang telah dibersihkan ke dalam file Excel baru jika diperlukan
# df_cleaned.to_excel('data_bersih.xlsx', index=False)
```

Data setelah menghapus data NaN:

No.	Kode Barang	Nama Barang	Satuan	Qty	Harga	Jumlah
2	3 8.968604e+09	INDOMIE GORENG JUMBO	PCS	27.0	5500.0	148500.0
3	4 8.968604e+09	INDOMIE GORENG KEBAB RENDANG	PCS	2.0	3500.0	7000.0
4	5 7.622202e+11	OREO SUMMER LTD 123,5G	PCS	5.0	8500.0	42500.0
5	6 7.622210e+11	KRAFT 70G	PCS	3.0	13000.0	39000.0
6	7 8.850389e+11	MOGU MOGU STR 320ML	PCS	9.0	10000.0	90000.0
...
1754	1763 8.993054e+11	PASEO BABY WIPES 50S	PCS	3.0	7500.0	22500.0
1755	1764 8.993054e+11	PASEO BABY WIPES 50S	PCS	9.0	8000.0	72000.0
1756	1765 8.993054e+11	PASEO TISU BASAH 25S	PCS	4.0	6000.0	24000.0
1757	1766 8.993054e+11	PASEO TISU BASAH 25S	PCS	8.0	6500.0	52000.0
1758	1767 8.997020e+11	GREEN SOOF 250	PCS	16.0	8000.0	128000.0

1549 rows x 7 columns

Gambar.3

Gambar di atas menunjukkan proses pembersihan (data cleansing) data penjualan. Pada kode tersebut, dilakukan penghapusan baris yang mengandung nilai NaN dengan menggunakan fungsi `dropna()`. Setelah proses ini, data yang tersisa ditampilkan, yang terdiri dari beberapa kolom seperti *No.*, *Kode Barang*, *Nama Barang*, *Satuan*, *Qty* (Quantity), *Harga*, dan *Jumlah*.

Data ini menunjukkan informasi penjualan dari berbagai produk dengan detail jumlah satuan terjual, harga per unit, serta total harga. Proses pembersihan ini merupakan bagian awal data yang bersih dan bebas dari nilai kosong (NaN), yang akan menjadi input bagi algoritma Naïve Bayes untuk analisis lebih lanjut.

```
# prompt: HAPUS DATA OUTLIER

# Loop melalui kolom-kolom numerik dalam DataFrame
for column in data_bersih.select_dtypes(include=np.number):
    # Hitung IQR
    Q1 = data_bersih[column].quantile(0.25)
    Q3 = data_bersih[column].quantile(0.75)
    IQR = Q3 - Q1

    # Tentukan batas atas dan batas bawah untuk outlier
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR

    # Hapus baris dengan nilai outlier
    data_bersih = data_bersih[(data_bersih[column] >= lower_bound) & (data_bersih[column] <= upper_bound)]

# Menampilkan data setelah menghapus outlier
print("\nData setelah menghapus outlier:")
data_bersih

# Anda dapat menyimpan DataFrame yang telah dibersihkan ke dalam file Excel baru jika diperlukan
# data_bersih.to_excel('data_bersih_tanpa_outlier.xlsx', index=False)
```

Data setelah menghapus outlier:

No.	Kode Barang	Nama Barang	Satuan	Qty	Harga	Jumlah
12	13 899033381111	CHOCOPIE	PCS	27	2500	67500
13	14 899033381121	CHOCO PIE ISI 2	DUS	1	7000	7000
16	17 899100235301	DELBIS GANDUM COKLAT 39G	PCS	1	2000	2000
17	18 899199811831	KOPI GAJAH SP MIX	BALL	8	1500	12000
18	19 899200378722	PERMEN ZIPLONG EUCALYPTUS MINT 100G	PCS	2	6000	12000
...
1753	1762 899275953545	NICE TISSU ROLL 238 S	PCS	2	5500	11000
1754	1763 899305364109	PASEO BABY WIPES 50S	PCS	3	7500	22500
1755	1764 899305364109	PASEO BABY WIPES 50S	PCS	9	8000	72000
1756	1765 899305364110	PASEO TISU BASAH 25S	PCS	4	6000	24000
1757	1766 899305364110	PASEO TISU BASAH 25S	PCS	8	6500	52000

914 rows x 7 columns

Gambar.4

Gambar di atas menunjukkan proses penghapusan *outlier* dalam data penjualan. Dalam kode tersebut, dilakukan langkah-langkah untuk mengidentifikasi dan menghapus *outlier* berdasarkan nilai kuartil (*Interquartile Range* atau IQR) pada kolom-kolom numerik.

Prosesnya meliputi:

1. **Perhitungan IQR** untuk setiap kolom numerik, yaitu mencari kuartil pertama (Q1) dan kuartil ketiga (Q3) serta menghitung rentang IQR (Q3 - Q1).
2. **Penentuan batas bawah dan atas** untuk mendeteksi *outlier*, dengan menggunakan rumus $Q1 - 1.5 * IQR$ untuk batas bawah dan $Q3 + 1.5 * IQR$ untuk batas atas.
3. **Menghapus baris yang mengandung *outlier*** dengan menyaring data yang berada di luar batas bawah dan atas.

Setelah *outlier* dihapus, data yang sudah bersih ditampilkan dan menunjukkan 914 baris yang tersisa.

```
# Reset index setelah menghapus baris
data_bersih = data_bersih.reset_index(drop=True)

# Menambahkan kolom 'Index' dengan nilai urutan dari 0 hingga jumlah baris
data_bersih['Index'] = range(len(data_bersih))

# Mengurutkan data berdasarkan kolom 'Index'
data_bersih = data_bersih.sort_values('Index')

# Menampilkan data setelah pengurutan dan penambahan kolom 'Index'
print("\nData setelah pengurutan dan penambahan kolom 'Index':")
data_bersih
```

Data setelah pengurutan dan penambahan kolom 'Index':

No.	Kode Barang	Nama Barang	Satuan	Qty	Harga	Jumlah	Index
0	13 899033381111	CHOCOPIE	PCS	27	2500	67500	0
1	14 899033381121	CHOCO PIE ISI 2	DUS	1	7000	7000	1
2	17 899100235301	DELBISS GANDUM COKLAT 39G	PCS	1	2000	2000	2
3	18 899199811831	KOPI GAJAH SP MIX	BALL	8	1500	12000	3
4	19 899200378722	PERMEN ZIPLONG EUCALYPTUS MINT 100G	PCS	2	6000	12000	4
...
909	1762 899275953545	NICE TISSU ROLL 238 S	PCS	2	5500	11000	909
910	1763 899305364109	PASEO BABY WIPES 50S	PCS	3	7500	22500	910
911	1764 899305364109	PASEO BABY WIPES 50S	PCS	9	8000	72000	911
912	1765 899305364110	PASEO TISU BASAH 25S	PCS	4	6000	24000	912
913	1766 899305364110	PASEO TISU BASAH 25S	PCS	8	6500	52000	913

914 rows x 8 columns

```
[ ] data_bersih.to_excel('data_bersih_banget.xlsx', index=False)
```

Gambar.5

Gambar di atas menunjukkan proses pengaturan ulang indeks pada data penjualan setelah penghapusan baris dan *outlier*. Langkah-langkah yang dilakukan meliputi:

1. **Reset indeks** dengan `reset_index(drop=True)` untuk menghilangkan indeks lama dan membuat indeks baru yang dimulai dari 0.
2. **Menambahkan kolom 'Index'** yang berisi nilai urutan dari 0 hingga jumlah baris dalam data.
3. **Mengurutkan data berdasarkan kolom 'Index'** agar data tersusun rapi sesuai urutan.

Setelah proses ini, data yang bersih ditampilkan dengan indeks baru dan kolom tambahan 'Index'.

```

# prompt: label encoding kuantitas, Price, jml
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()

# Assuming 'kuantitas', 'Price', and 'jml' are the column names in your DataFrame
data['jml'] = le.fit_transform(data['jml'])

data.head()

```

Kode Barang	Nama Barang	Satuan	Qty	Harga	Jumlah	kuantitas	Price	jml	
0	89903381111	CHOCOPIE	PCS	27	2500	67500	Banyak	Murah	1
1	89903381121	CHOCO PIE ISI 2	DUS	1	7000	7000	Sedek	Mahal	0
2	89910023301	DELBSI GANDUM COKLAT 3KG	PCS	1	2000	2000	Sedek	Murah	0
3	899199611831	KOPRI GAJAH SF MIX	BALL	8	1500	12000	Sedek	Murah	0
4	899200378722	PERMEN ZIPLONG EUCALYPTUS MINT 100G	PCS	2	6000	12000	Sedek	Mahal	0

```

data['Price'] = le.fit_transform(data['Price'])

data.head()

```

Kode Barang	Nama Barang	Satuan	Qty	Harga	Jumlah	kuantitas	Price	jml
0	89903381111	CHOCOPIE	PCS	27	2500	67500	1	1
1	89903381121	CHOCO PIE ISI 2	DUS	1	7000	7000	0	0
2	89910023301	DELBSI GANDUM COKLAT 3KG	PCS	1	2000	2000	1	0
3	899199611831	KOPRI GAJAH SF MIX	BALL	8	1500	12000	1	0
4	899200378722	PERMEN ZIPLONG EUCALYPTUS MINT 100G	PCS	2	6000	12000	1	0

```

data['kuantitas'] = le.fit_transform(data['kuantitas'])

data.head()

```

Kode Barang	Nama Barang	Satuan	Qty	Harga	Jumlah	kuantitas	Price	jml
0	89903381111	CHOCOPIE	PCS	27	2500	67500	0	1
1	89903381121	CHOCO PIE ISI 2	DUS	1	7000	7000	1	0
2	89910023301	DELBSI GANDUM COKLAT 3KG	PCS	1	2000	2000	1	1
3	899199611831	KOPRI GAJAH SF MIX	BALL	8	1500	12000	1	1
4	899200378722	PERMEN ZIPLONG EUCALYPTUS MINT 100G	PCS	2	6000	12000	1	0

```

data.info()

```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 514 entries, 0 to 513
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   Kode Barang  514 non-null    int64
 1   Nama Barang  514 non-null    object
 2   Satuan       514 non-null    object
 3   Qty          514 non-null    int64
 4   Harga        514 non-null    int64
 5   Jumlah      514 non-null    int64
 6   kuantitas    514 non-null    int64
 7   Price        514 non-null    int64
 8   jml          514 non-null    int64
dtypes: int64(7), object(2)
memory usage: 64.4+ KB

```

```

data.describe()

```

	Kode Barang	Qty	Harga	Jumlah	kuantitas	Price	jml
count	514	914.000000	914.000000	914.000000	914.000000	914.000000	914.000000
mean	8.960198e+11	6.781182	6202.297593	30425.711160	0.787593	0.436543	0.164962
std	2.82419e+08	6.114841	4276.386682	24007.496679	0.402914	0.496228	0.388430
min	8.990344e+11	1.000000	500.000000	1000.000000	0.000000	0.000000	0.000000
25%	8.992770e+11	2.000000	2500.000000	12000.000000	1.000000	0.000000	0.000000
50%	8.994075e+11	5.000000	5500.000000	24250.000000	1.000000	0.000000	0.000000
75%	8.997214e+11	9.000000	8500.000000	42000.000000	1.000000	1.000000	0.000000
max	9.000000e+11	29.000000	19000.000000	108000.000000	1.000000	1.000000	1.000000

```

data.isnull().sum()

```

	Kode Barang	Nama Barang	Satuan	Qty	Harga	Jumlah	kuantitas	Price	jml
count	514	514	514	514	514	514	514	514	514
dtype	int64	object	object	int64	int64	int64	int64	int64	int64

```

data.duplicated().sum()

```

Gambar.6

Gambar di atas menunjukkan beberapa tahapan dalam *data preprocessing*.

1. **Label Encoding:** Kolom *Jml*, *Price*, dan *kuantitas* di-encode menggunakan LabelEncoder untuk mengonversi data kategorikal menjadi nilai numerik. Misalnya, nilai "Laris" dan "Murah" diubah menjadi 1, sedangkan "Tidak Laris" dan "Mahal" menjadi 0, yang penting untuk analisis algoritmik.

2. **Data Inspection:**

- `data.info()`: Menampilkan informasi tentang struktur data, jumlah entri, tipe data untuk setiap kolom, dan jumlah kolom yang non-null. Terlihat bahwa dataset memiliki 914 baris tanpa nilai kosong (*null*).
- `data.describe()`: Memberikan statistik deskriptif pada kolom numerik, seperti *mean*, *std*, *min*, dan *max*, yang membantu memahami distribusi data.

3. **Missing Values and Duplicates Check:**

- `data.isnull().sum()`: Memeriksa nilai kosong di setiap kolom dan menunjukkan bahwa tidak ada nilai kosong dalam dataset.
- `data.duplicated().sum()`: Mengecek apakah ada duplikat dalam data, yang akan menjadi penting untuk memastikan integritas data sebelum analisis lebih lanjut.

Tahapan ini bertujuan untuk membersihkan dan menyiapkan data.

Implementasi Naive Bayes

Implementasi menggunakan algoritma Naive Bayes dengan pendekatan CRM terhadap dataset penjualan KPRI KOKARNABA. Data yang diperoleh sebanyak 1759 dataset, kemudian dipisah dengan cara holdout sehingga menghasilkan 914 dataset dengan data training dan data testing dimana pembagian data dengan perbandingan 8:2. Prediksi penjualan dengan algoritma Naive Bayes dilakukan dengan menggunakan Google Collab sehingga diperoleh hasil performance sebagai berikut:

```
# prompt: buat model naive bayes dengan standar scaler

from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score, classification_report

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Scale the features using StandardScaler
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Create and train the Naive Bayes model
model = GaussianNB()
model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = model.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)

# Print classification report
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.86	0.86	0.86	148
1	0.41	0.40	0.41	35
accuracy			0.78	183
macro avg	0.64	0.63	0.63	183
weighted avg	0.77	0.78	0.77	183

Gambar.7

Gambar di atas menunjukkan implementasi kode Python menggunakan Google Collab untuk membangun model Naïve Bayes dengan preproses data menggunakan *Standard Scaler*.

Berikut penjelasan langkah-langkah dalam kode tersebut:

1. **Import Library:** Kode dimulai dengan mengimpor pustaka dari sklearn untuk melakukan pembagian data, preprocessing, pelatihan model Naïve Bayes, dan evaluasi.
2. **Pembagian Data:** Data dibagi menjadi data latih dan data uji menggunakan `train_test_split` dengan proporsi 80% untuk data latih dan 20% untuk data uji.
3. **Normalisasi Data:** Data fitur dinormalisasi menggunakan `StandardScaler` untuk memastikan bahwa setiap fitur memiliki skala yang sama, yang membantu meningkatkan performa model.
4. **Pembuatan dan Pelatihan Model:** Model Naïve Bayes (`GaussianNB`) dibuat dan dilatih menggunakan data latih.
5. **Prediksi:** Model melakukan prediksi pada data uji.
6. **Evaluasi Model:** Akurasi model dihitung, dan hasil klasifikasi ditampilkan dalam bentuk laporan yang mencakup metrik seperti *precision*, *recall*, dan *f1-score*.

Hasil evaluasi menunjukkan akurasi sebesar 77.96%, dengan nilai *precision*, *recall*, dan *f1-score* yang berbeda untuk setiap kelas.

Analisa Hasil

1. Akurasi Model

Model Naïve Bayes ini memiliki tingkat akurasi sebesar 77.56%. Artinya, sekitar 78% prediksi yang dilakukan oleh model ini sesuai dengan data asli. Ini merupakan tingkat akurasi yang cukup baik dalam konteks analisis penjualan, meskipun ada beberapa area yang dapat diperbaiki, terutama dalam hal mendeteksi barang yang kurang laris.

2. Kinerja Model Berdasarkan Kategori Barang

Kategori 0 (Barang Terlaris): Pada kategori ini, model menunjukkan performa yang cukup baik:

Precision sebesar 0.86, menunjukkan bahwa 86% dari barang yang diprediksi terlaris memang benar-benar terlaris.

Recall sebesar 0.86, yang berarti model mampu mengenali 86% dari total barang yang memang terlaris.

F1-score juga sebesar 0.86, menunjukkan keseimbangan yang baik antara precision dan recall pada kategori barang terlaris.

Kategori 1 (Barang Kurang Laris): Di sisi lain, pada kategori ini model menunjukkan performa yang lebih rendah:

Precision sebesar 0.41, yang berarti hanya 41% dari barang yang diprediksi sebagai kurang laris benar-benar masuk dalam kategori ini.

Recall sebesar 0.40, yang berarti model hanya mampu mengenali 40% dari total barang yang sebenarnya kurang laris.

F1-score sebesar 0.41, menunjukkan bahwa model ini kurang optimal dalam mendeteksi barang kurang laris.

Rendahnya precision dan recall pada kategori barang kurang laris menunjukkan bahwa model mengalami kesulitan dalam membedakan barang yang benar-benar kurang laris dari yang terlaris. Salah satu penyebabnya mungkin karena ketidakseimbangan data, di mana jumlah barang terlaris lebih banyak dibandingkan barang yang kurang laris.

3. Evaluasi Rata-rata (Macro dan Weighted Average)

Macro Average: Precision, recall, dan f1-score rata-rata untuk kedua kategori adalah sekitar 0.63. Ini menunjukkan bahwa model bekerja cukup baik secara keseluruhan tetapi masih ada ruang untuk perbaikan, khususnya pada kategori barang yang kurang laris.

Weighted Average: Precision, recall, dan f1-score tertimbang masing-masing adalah 0.77, 0.78, dan 0.77. Nilai ini menunjukkan kinerja model secara keseluruhan, tetapi didominasi oleh performa pada kategori barang terlaris (Kategori 0).

Model Naïve Bayes ini mampu memprediksi kategori barang terlaris dengan cukup baik, namun masih kurang dalam mengenali barang yang kurang laris. Hal ini menunjukkan bahwa model perlu disempurnakan, mungkin dengan teknik penyeimbangan data atau dengan mencoba algoritma lain yang lebih sesuai untuk mengatasi data yang tidak seimbang. Dengan perbaikan ini, model diharapkan dapat memberikan hasil yang lebih akurat dan andal, sehingga dapat membantu KPRI KOKARNABA Baturraden dalam mengoptimalkan strategi penjualan dan pengelolaan stok.

5. KESIMPULAN DAN SARAN

Kesimpulan dari penelitian ini menunjukkan bahwa penggunaan algoritma Naïve Bayes dalam analisis data penjualan di KPRI KOKARNABA Baturraden secara signifikan mampu memprediksi kategori produk terlaris dan kurang laris dengan tingkat akurasi sebesar 77,56%. Hasil penelitian ini menjawab tujuan utama, yaitu mengidentifikasi pola penjualan produk dan membantu dalam pengelolaan stok yang lebih efisien. Algoritma Naïve Bayes terbukti efektif untuk analisis ini, meskipun terdapat kendala dalam akurasi prediksi produk yang kurang laris, yang diduga disebabkan oleh ketidakseimbangan data antara kategori produk terlaris dan kurang laris.

Sebagai rekomendasi, disarankan agar koperasi mempertimbangkan metode penyeimbangan data untuk meningkatkan akurasi prediksi di kategori produk kurang laris. Selain itu, penelitian ini terbatas pada analisis data penjualan tanpa mempertimbangkan faktor eksternal yang mungkin memengaruhi pola pembelian anggota. Untuk penelitian mendatang, diharapkan dapat dilakukan pengembangan model dengan mempertimbangkan faktor-faktor tersebut, guna memperoleh hasil yang lebih komprehensif dan mendukung perencanaan strategi koperasi yang lebih efektif.

6. DAFTAR REFERENSI

- Andarista, R. R., & Jananto, A. (2022). Penerapan Data Mining Algoritma C4.5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor. *Jurnal Tekno Kompak*, 16(2), 29. <https://doi.org/10.33365/jtk.v16i2.1525>
- Anggriani, M., Umar, R., & Fadlil, A. (2021). Analisis Perbandingan Sistem Bangunan Pembelian Bahan Konsumen Dengan Data Mining (Studi Kasus Pt Sinar Valcosindo Teknik). ... *Sistem Informasi Dan ...*, 19–27. <https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/3%0Ahttps://sismatik.nusaputra.ac.id/index.php/sismatik/article/download/3/4>
- Arfan, U., & Paraga, N. (2024). Perbandingan Algoritma K-Means, Naïve Bayes dan Decision Tree Dalam Memprediksi Penjualan Bahan Bakar Minyak. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1379–1389. <https://doi.org/10.57152/malcom.v4i4.1566>
- Juanda Saputra, M., & Izman Herdiansyah, M. (2022). Penerapan Naive Bayes Dalam Memprediksi Penjualan Dan Persediaan Kain Jumputan Pada Toko Batiq Colet Tuan Kentang Palembang. *Jurnal Mantik*, 6(2), 2502–2507.
- Pranata, B. S., & Utomo, D. P. (2020). Penerapan Data Mining Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel Motor (Study Kasus Bengkel Sinar Service). *Bulletin of Information Technology (BIT)*, 1(2), 83–91.
- Prasetyo, A., Sastra, R., & Musyaffa, N. (2020). Implementasi Data Mining Untuk Analisis Data Penjualan Dengan Menggunakan Algoritma Apriori (Studi Kasus Dapoerin'S). *Jurnal Khatulistiwa Informatika*, 8(2). <https://doi.org/10.31294/jki.v8i2.8994>
- Sultan Iskandar, M., Syahra, Y., Si, S., & Kom, M. (2020). Implementasi Data Mining Menentukan Pola Penjualan Produk Pt. Sinergia Beaute Indonesia Dengan Algoritma Apriori. *Cyber Tech*, pp.
- Zafira, F., Irawan, B., & Bahtiar, A. (2024). Penerapan Data Mining Untuk Estimasi Stok Barang Dengan Metode K-Means Clustering. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 156–161. <https://doi.org/10.36040/jati.v8i1.8319>
- Andarista, R. R., & Jananto, A. (2022). Penerapan Data Mining Algoritma C4.5 Untuk Klasifikasi Hasil Pengujian Kendaraan Bermotor. *Jurnal Tekno Kompak*, 16(2), 29. <https://doi.org/10.33365/jtk.v16i2.1525>
- Anggriani, M., Umar, R., & Fadlil, A. (2021). Analisis Perbandingan Sistem Bangunan Pembelian Bahan Konsumen Dengan Data Mining (Studi Kasus Pt Sinar Valcosindo Teknik). ... *Sistem Informasi Dan ...*, 19–27. <https://sismatik.nusaputra.ac.id/index.php/sismatik/article/view/3%0Ahttps://sismatik.nusaputra.ac.id/index.php/sismatik/article/download/3/4>
- Arfan, U., & Paraga, N. (2024). Perbandingan Algoritma K-Means, Naïve Bayes dan Decision Tree Dalam Memprediksi Penjualan Bahan Bakar Minyak. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1379–1389. <https://doi.org/10.57152/malcom.v4i4.1566>
- Juanda Saputra, M., & Izman Herdiansyah, M. (2022). Penerapan Naive Bayes Dalam Memprediksi Penjualan Dan Persediaan Kain Jumputan Pada Toko Batiq Colet Tuan Kentang Palembang. *Jurnal Mantik*, 6(2), 2502–2507.

- Pranata, B. S., & Utomo, D. P. (2020). Penerapan Data Mining Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel Motor (Study Kasus Bengkel Sinar Service). *Bulletin of Information Technology (BIT)*, 1(2), 83–91.
- Prasetyo, A., Sastra, R., & Musyaffa, N. (2020). Implementasi Data Mining Untuk Analisis Data Penjualan Dengan Menggunakan Algoritma Apriori (Studi Kasus Dapoerin'S). *Jurnal Khatulistiwa Informatika*, 8(2). <https://doi.org/10.31294/jki.v8i2.8994>
- Sultan Iskandar, M., Syahra, Y., Si, S., & Kom, M. (2020). Implementasi Data Mining Menentukan Pola Penjualan Produk Pt. Sinergia Beaute Indonesia Dengan Algoritma Apriori. *Cyber Tech*, pp.
- Zafira, F., Irawan, B., & Bahtiar, A. (2024). Penerapan Data Mining Untuk Estimasi Stok Barang Dengan Metode K-Means Clustering. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 156–161. <https://doi.org/10.36040/jati.v8i1.8319>