

Penerapan Data Mining Untuk Klasifikasi Produk Terlaris Menggunakan Algoritma Naive Bayes Pada Bengkel Motor

Alvin Julianto¹, Sri Andayani²

^{1,2}Universitas Katolik Musi Charitas

^{1,2}Jl. Bangau No.60, Kota Palembang, Indonesia

Email: alvinjulianto080601@gmail.com¹, andayani_s@ukmc.ac.id²

Abstrak— Bengkel Aldo Motor merupakan salah satu bengkel motor yang ada di Kota Palembang, Sumatera Selatan. Bengkel ini melayani jasa untuk perbaikan motor serta menjual suku cadang motor seperti oli, lampu, busi dan lain-lain. Bengkel ini memiliki permasalahan yaitu produk yang laris sering kali kehabisan stok sedangkan untuk produk yang kurang laris, stok produknya mengalami penumpukan. Untuk memecahkan masalah tersebut dilakukan analisis data mining klasifikasi dengan menggunakan algoritma naive bayes menggunakan aplikasi rapid miner. Data yang digunakan adalah data penjualan dari bulan Mei 2023 – Oktober 2023. Berdasarkan hasil dari pengolahan data didapatkan hasil accuracy sebesar 86,77%, precision sebesar 85,39% dan nilai recall sebesar 76%. Hasil ini menunjukkan bahwa algoritma naive bayes dapat digunakan untuk menentukan produk terlaris pada bengkel aldo motor.

Kata Kunci: produk terlaris, rapidminer, naive bayes, klasifikasi

Abstract— Aldo Motor Workshop is one of motorcycle workshops in Palembang City, South of Sumatera. This workshop provides services for motorcycle repairs and sells motorcycle spare parts such as oil, lights, spark plugs and others. This workshop has a problem, namely that product that are in demand are often out of stock while for products that are less in demand, the stock of product has accumulated. To solve this problem, a classification data mining analysis is carried out using the naive bayes algorithm using the rapid miner application. The data used is sales data from May 2023 – October 2023. Based on the results of data processing, the accuracy results is 86.77%, precision is 85.395 and recall value is 76%. These results show that the naive bayes algorithm can be used to determine the best selling products at the Aldo Motor Workshop.

Keyword: best selling product, rapidminer, naive bayes, classification

I. PENDAHULUAN

Bengkel Aldo Motor adalah salah satu bengkel di Kota Palembang yang menyediakan jasa perbaikan motor dan menjual suku cadang kendaraan bermotor seperti oli, lampu, spion, rantai, ban luar, ban dalam dan lainnya. Banyaknya jumlah variasi produk yang dijual membuat pengelola Bengkel Aldo Motor kesulitan untuk mengelola stok di bengkel.

Produk yang sering dibeli oleh pelanggan sering kali kehabisan stok sedangkan produk yang kurang diminati oleh pelanggan mengalami penumpukan di gudang. Hal ini menyebabkan pengelola mengalami kesulitan untuk pembelian stok produk agar stok terjaga dan tidak mengalami penumpukan stok produk. Data mining dapat membantu bisnis dalam mengidentifikasi pola pembelian pelanggan mereka [1].

Data mining adalah proses mengekstraksi dan menambang pengetahuan yang dibutuhkan dari sejumlah besar data. Dalam data mining terdapat berbagai fungsi dimana salah satunya adalah klasifikasi. Klasifikasi adalah menggeneralisasi struktur yang diketahui untuk diaplikasikan pada data-data baru. Jenis klasifikasi seperti pohon keputusan, K-nearest neighbor, neural network, naive bayes dan support vector machines.

Metode Naive Bayes pada klasifikasi pada data mining merupakan salah satu metode yang paling sering digunakan. Salah satunya adalah klasifikasi produk terlaris pada penjualan voucher kuota. Jenis voucher yang dijual adalah telkomsel, voucher XL, voucher Indosat dan voucher 3. Produk-produk tersebut dikategorikan berkualitas rendah atau berkualitas tinggi. Hasil yang didapat adalah voucher 3 adalah voucher yang paling laris. Metode ini menghasilkan accuracy 97,50%, precision 100% dan skor recall 93,48% [2].

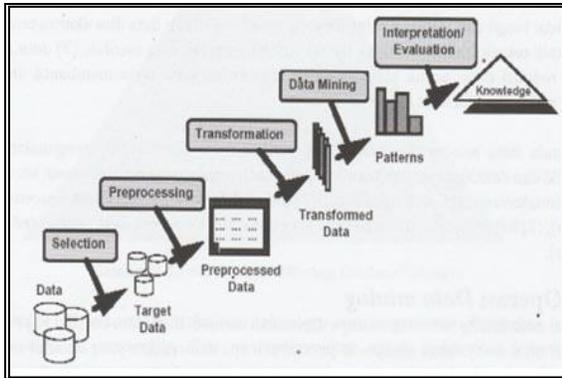
Implementasi data mining juga dapat dilakukan pada produk-produk yang tersedia di apotik misalnya vitamin, hal ini karena vitamin memiliki umur simpan yang sedikit berbeda dengan vitamin yang tersedia di apotik. Untuk mengatasi maka obat-obatan di vitamin dapat diklasifikasi berdasarkan variabel-variabel yang diperoleh. Nilai akurasi yang didapat adalah 88% [3].

Berdasarkan dari latar belakang di atas maka tujuan penelitian ini adalah melakukan implementasi algoritma naive bayes untuk mengklasifikasi produk terlaris pada Bengkel Aldo Motor.

II. LANDASAN TEORI

A. Data Mining

Data mining adalah proses yang menggunakan pembelajaran mesin untuk menganalisis dan mengekstraksi data secara otomatis [4].



Gambar 1. Tahapan Data Mining [5]

Tahapan Data Mining:

1. *Selection*: Data yang diperoleh disaring terlebih dahulu. Melalui seleksi data, proses pengolahan dapat ditingkatkan untuk mencapai tujuan penelitian dengan lebih efektif.
2. *Preprocessing*[6]: serangkaian proses persiapan data sebelum data *mining* dilakukan. Tahapan *preprocessing* biasanya meliputi pembersihan (*cleaning*), pengurangan (*reduction*), dan integrasi data.
3. *Transformation*: Data perlu mengalami transformasi sebelum diolah menggunakan data *mining*. Tujuannya adalah untuk menyesuaikan data yang akan diproses dengan algoritma dan perangkat lunak yang dipakai untuk mengolah data.
4. *Data Mining*: Data diproses berdasarkan algoritma sesuai dengan teknik data *mining*. Dalam penelitian ini algoritma *naive bayes* digunakan, ini adalah salah satu algoritma dalam metode klasifikasi, dan outputnya dapat berupa prediksi atau klasifikasi berdasarkan data yang telah diproses.
5. *Interpretation/Evaluation*: Tahapan di mana hasil yang diperoleh dari teknik data *mining* diinterpretasikan.

B. Klasifikasi

Klasifikasi adalah metode untuk mengembangkan model atau fungsi yang menerapkan model matematika pada data atau konsep untuk membuat inferensi statistik untuk data yang tidak diketahui [7]. Variabel kategori digunakan dalam klasifikasi target. Sebagai contoh, klasifikasi keuntungan dapat dibagi menjadi tiga kategori: kecil, menengah, dan besar [8].

C. Algoritma Naïve Bayes

Algoritma Naïve Bayes tidak memiliki aturan, metode ini menggunakan suatu cabang matematik yaitu konsep probabilitas. Konsep ini digunakan untuk menemukan probabilitas tertinggi dari proses klasifikasi yang dilakukan dengan melakukan pengamatan terhadap kemunculan setiap klasifikasi dalam data pelatihan.

Rumusan teorema bayes seperti pada persamaan berikut [9]:

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

Penjelasan dari persamaan 1 sebagai berikut:

- X : Data yang memiliki kelas yang tidak diketahui.
- H : Hipotesis untuk data X yang termasuk dalam kelas tertentu.
- P(H|X): Probabilitas nilai hipotesis H berdasarkan kondisi X.
- P(H) : Nilai probabilitas hipotesis H.
- P(X|H): Probabilitas nilai hipotesis X berdasarkan kondisi H.
- P(X) : Nilai probabilitas hipotesis X

a. Perhitungan Prior

Perhitungan yang dilakukan dengan menghitung perbandingan jumlah anggota kelas dengan seluruh sampel data.

$$P = \frac{X}{A} \dots\dots\dots(2)$$

Keterangan :

- P : Angka *prior*
- X : Total data untuk setiap kelas
- A : Kumpulan data untuk semua kelas

b. Perhitungan Likelihood

Proses menghitung nilai probabilitas/kemungkinan untuk setiap kelas probabilitas kemunculan karakteristik dari karakteristik yang dipilih.

$$L = \frac{F}{B} \dots\dots\dots (3)$$

Keterangan :

- L : Nilai probabilitas
- F : Total data atribut untuk setiap kelas
- B : Total seluruh data tiap kelas

c. Perhitungan Posterior

Hasil dari perhitungan *likelihood* berupa probabilitas atribut untuk suatu kelas, yang digunakan untuk menentukan probabilitas bahwa masing-masing atribut dimasukkan ke dalam kelas. Perhitungan kembali dihitung dengan mengalikan kemungkinan karakteristik input dengan kuadrat, dalam proses ini diperoleh probabilitas akhir dari hasil tugas.

$$P(c) \pi P(a|c) \dots\dots\dots (4)$$

Keterangan :

- P(C) : Angka *prior* tiap kelas
- P(a|c) : Angka probabilitas

Hasil dari proses klasifikasi dibuat dengan melakukan perbandingan nilai *posterior* dari *class* yang ada. Angka *posterior* tertinggi adalah nilai yang dipilih dari klasifikasi.

D. Confusion Matrix

Confusion Matrix merupakan alat visualisasi umum dalam pembelajaran yang diawasi. Setiap kolom dalam matriks mewakili turunan dari kelas predikat, dan setiap baris mewakili turunan aktual dari kelas tersebut [10].

Tabel 1 Confusion Matrix

		Predicted Sample	
		True	False
Actual Sample	True	TP	FP
	False	FN	TN

Keterangan :

TP : *True positives*, banyaknya data dengan kelas positif tergolong positif.

TN : *True negative*, banyaknya data dengan kelas negatif tergolong negatif.

FP : *False positives*, total data dengan kelas positif tergolong negatif.

FN : *False negative*, banyaknya data dengan kelas negatif yang tergolong positif

Efisiensi matriks konfusi adalah ukuran akurasi, presisi dan nilai memori dari suatu model algoritma [11]. Ada tiga nilai untuk mengukur kemampuan sistem klasifikasi yakni akurasi, presisi dan recall [12]. Nilai akurasi adalah persentase akurasi antara nilai perkiraan dan nilai yang sebenarnya. Nilai presisi adalah presisi dari kelas yang diprediksi. Nilai *recall* adalah persentase dari nilai keberhasilan algoritma yang digunakan [7].

Nilai akurasi merupakan rasio dari jumlah total perkiraan yang benar, nilai akurasi yang tinggi dapat memperkirakan kebenaran lebih banyak, dapat dilakukan perhitungan menggunakan rumus berikut:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \dots\dots\dots (5)$$

Recall dihitung untuk menilai kemampuan model dalam memprediksi kelas tertentu. Ini persis rasio antara jumlah data di kelas tertentu dibagi dengan jumlah semua kelas.

$$Recall = \frac{TP}{TP + FN} \dots\dots\dots (6)$$

Precision digunakan untuk menilai seberapa baik model dapat memperkirakan class. Akurasi ditentukan dengan menghitung rasio nilai data untuk kelas tertentu dibagi dengan jumlah total prediksi untuk kelas tersebut.

$$Precision = \frac{TP}{TP + FP} \dots\dots\dots (7)$$

Nilai akurasi dapat dinilai berdasarkan beberapa kategori sesuai dengan Tabel 2 [10].

Tabel 2 Kategori Nilai AUC

Nilai AUC	Kategori Klasifikasi
0.90 – 1.00	<i>Excellent Classification</i>
0.80 – 0.90	<i>Good Classification</i>
0.70 – 0.80	<i>Fair Classification</i>
0.60 – 0.70	<i>Poor Classification</i>
0.50 – 0.60	<i>Failure</i>

Presisi dan recall diinterpretasikan dengan kategori tiga kelas. Metode yang digunakan skala interval dengan menghitung perbedaan antara nilai presisi dan recall tertinggi (1) dan nilai presisi dan recall terendah (0) sesuai dengan Tabel 3.

Tabel 3 Kategori Precision dan Recall

Nilai Precision dan Recall	Kategori Klasifikasi
0.67 – 1.00	Tinggi
0.34 – 0.66	Sedang
0.00 – 0.33	Rendah

E. Penelitian Terdahulu

Algoritma Naïve Bayes digunakan untuk menentukan produk laris dan tidak laris pada konter HP Bayu Cell untuk penjualan handphone. Hal ini untuk mengatasi penumpukan produk yang tidak terjual. Algoritma Naïve Bayes memanfaatkan peluang suatu kejadian berdasarkan kejadian lain yang telah terjadi. Hasil dari penelitian ini adalah handphone dengan merek REDMI NOTE 9 termasuk dalam kategori laris dengan nilai peluang sebesar 0.005211 [13].

Penggunaan algoritma Naïve Bayes untuk klasifikasi produk dilakukan pada penjualan voucher kuota di Edi Cell. Penelitian ini didasari oleh kemajuan teknologi informasi dan komunikasi yang terus berlangsung dengan cepat dan harga yang semakin terjangkau sehingga voucher kuota menjadi kebutuhan yang mendesak bagi masyarakat. Produk voucher yang tersedia adalah voucher Telkomsel, XL, Indosat dan 3. Produk-produk tersebut dikategorikan berkualitas rendah dan berkualitas tinggi. Tujuannya adalah mempelajari bagaimana penggunaan algoritma Naïve Bayes dalam data mining untuk menentukan klasifikasi produk berbasis kredit dan keakuratan data yang diperoleh. Hasil dari penelitian ini adalah bahwa voucher kuota yang paling laris adalah voucher 3. Metode Naïve Bayes pada penelitian ini menghasilkan skor accuracy 97.50%, precision 100% dan recall 93.48% [2].

Penelitian lain yang menerapkan algoritma Naïve Bayes adalah penelitian dengan tujuan untuk peningkatan strategi pemasaran. Hal ini untuk mengatasi masalah yang dihadapi toko Universal Homeware di mana toko hanya menyimpan data transaksi penjualan tanpa mengetahui manfaatnya. Toko tersebut tidak mengetahui bagaimana mengolah data untuk strategi bisnis masa depan. Hasil dari penelitian ini adalah 36 titik data menghasilkan akurasi 80% [14].

Algortima Naïve Bayes juga digunakan untuk memperkirakan penjualan di toko VJCakes Pematang Siantar. Penjualan produk kue diperkirakan akan terus mendongkrak pendapatan. Salah satu caranya adalah dengan menawarkan barang sesuai dengan permintaan konsumen untuk mencegah kerugian. Hingga saat ini, banyak pelaku bisnis yang sering mengalami kerugian akibat tidak adanya sistem penjualan. Hasil dari penelitian ini adalah algoritma Naïve Bayes dapat diimplementasikan untuk memprediksi penjualan dengan tingkat akurasi sebesar 83.44% [15].

Penerapan data mining dengan algoritma Naïve Bayes juga dilakukan untuk memprediksi pembelian cat. Algoritma Naïve Bayes diimplementasikan dalam bentuk aplikasi yang diharapkan dapat memprediksi dan mendeteksi lebih awal minat beli calon pelanggan berdasarkan transaksi masa lalu sehingga perusahaan dapat lebih memahami cat yang diminati pelanggan. Hasil penerapan algoritma Naïve Bayes dari 60 data yang diuji berhasil mengklasifikasi 48 data dengan persentase keakuratan sebesar 80% [6].

III. METODE

Penelitian ini mempunyai beberapa tahapan sehingga pelaksanaan menjadi lebih terstruktur seperti:

- a. Melakukan identifikasi masalah
Mengidentifikasi masalah yang ada pada Bengkel Aldo Motor adalah belum diklasifikasikan produk-produk sehingga sering kali bengkel mengalami kehabisan stok sedangkan untuk produk yang tidak laris, produk mengalami penumpukan.
- b. Wawancara
Dilakukan pada pemilik Bengkel Aldo Motor sehingga didapatkan lebih jelas permasalahan yang ada di bengkel serta melakukan pengumpulan data yang diperlukan.
- c. Studi Literatur
Melakukan studi literatur dari buku, jurnal ilmiah yang mempunyai algoritma yang sama atau objek yang sama.
- d. Pengumpulan Data di Tahapan KDD
Data yang telah diperoleh kemudian diolah dengan menggunakan tahapan Knowledge Discovery in Database yaitu:
 1. *Selection* (data dipilih yaitu data penjualan dari Bulan Mei 2023 sampai dengan bulan Oktober 2023)
 2. *Preprocessing*, pada tahap ini ada 3 proses yaitu:
 Cleaning: data dibersihkan untuk memastikan keakuratan data dalam penelitian
 Reduction: Data yang sudah dibersihkan kemudian masuk ke dalam reduction yaitu data akan dikurangi jika memiliki data yang kurang relevan atau tidak diperlukan.
 Integration: Data yang sudah dikurangi maka akan ada proses untuk menggabungkan atribut data yang akan menghasilkan kelas baru yang dapat membantu memperlancar proses penelitian.
 3. *Transformation*: pada tahap ini seluruh data pada tahapan sebelumnya akan ditransformasikan dalam bentuk kategorial.

- e. Pengolahan data menggunakan rapidminer dengan menerapkan algoritma Naïve Bayes.
- f. Pembahasan hasil dari pengolahan data menggunakan rapidminer serta melakukan analisis terhadap hasil yang didapatkan.
- g. Kesimpulan dan Saran. Menarik kesimpulan atas hasil yang didapat dan memberikan saran perbaikan untuk penelitian selanjutnya mengenai kelemahan yang ditemukan dalam penelitian ini.

IV. HASIL DAN PEMBAHASAN

Data penelitian yang didapatkan dari Bengkel Aldo Motor adalah data penjualan bulan Mei – Oktober 2023. Informasi data seperti pada Tabel 4.

Tabel 4 Informasi Data Penjualan

Data Penjualan	
Jumlah Data	1857
Atribut	6
Tahun Data	2023

Berdasarkan Tabel 4, data penjualan memiliki jumlah 1875 record data dan jumlah atribut sebanyak enam atribut. Atribut-atribut yang ada adalah nomor, tanggal penjualan, nama produk, jenis produk, harga, jumlah penjualan.

Tabel 5 Atribut Data Sebelum Diproses/Diolah

No	Nama Atribut	Jenis Atribut
1	Nomor	Numerik
2	Tanggal Penjualan	Karakter
3	Nama Produk	Karakter
4	Jenis Produk	Karakter
5	Harga	Numerik
6	Total Penjualan	Numerik

Tabel 5 menunjukkan atribut pada kumpulan data yang memiliki tipe fungsi yang berbeda-beda yaitu nilai atribut berupa tipe klasifikasi, atribut numerik yang menunjukkan nilai berbentuk angka dan atribut karakter berbentuk karakter.

- a. Selection
Data penjualan bengkel Aldo Motor dari Bulan Mei 2023 - Oktober 2023 adalah data yang akan dipakai dalam penelitian ini. Data ini didapatkan dari hasil merekap pembukuan beserta nota penjualan dari bengkel Aldo Motor. Data ini terdiri dari nomor, tanggal penjualan, nama produk, jenis produk, harga, dan total penjualan. Data-data tersebut kemudian digabungkan sehingga menjadi satu dataset. Semua data beserta atribut yang ada akan diolah menggunakan tahapan KDD (*Knowledge Discovery in Database*).
- b. *Preprocessing Data*
Pada tahapan ini data yang sudah didapatkan akan melalui pra pemrosesan yaitu membersihkan data, mengurangi data serta menggabungkan data. Dalam tahapan ini terbagi

ini harga murah dan mahal ditentukan dari hasil wawancara dengan pemilik bengkel Aldo motor. Hasil transformasi bisa dilihat pada Tabel 9.

Tabel 9 Atribut Harga Yang Telah Ditransformasikan

Bulan	Nama Produk	Jenis Produk	Harga	Total Penjualan	Status Penjualan
Mei	Yamalube Matic	Oli	Murah	35	Laris
Mei	Yamalube Silver	Oli	Murah	23	Laris
Mei	Yamalube Gold	Oli	Murah	2	Tidak Laris
Mei	Yamalube Sport	Oli	Mahal	9	Tidak Laris
Mei	Yamalube Super Matic	Oli	Mahal	3	Tidak Laris
Mei	Yamalube XP	Oli	Murah	9	Tidak Laris
Mei	Gear Yamalube 1	Oli	Murah	9	Tidak Laris
Mei	Gear Yamalube 2	Oli	Murah	8	Tidak Laris
Mei	Gear Yamalube 3	Oli	Murah	2	Tidak Laris
Mei	MPX 1 (0,8l)	Oli	Murah	11	Laris
Mei	MPX 1 (1l)	Oli	Mahal	5	Tidak Laris
Mei	MPX 2	Oli	Murah	77	Laris
Mei	MPX Gear	Oli	Murah	58	Laris
Mei	Enduro 4T	Oli	Murah	9	Tidak Laris
Mei	Enduro Racing	Oli	Mahal	1	Tidak Laris
Mei	Enduro Matic	Oli	Murah	13	Tidak Laris
Mei	Mesran X	Oli	Murah	6	Tidak Laris
Mei	Shell Matic	Oli	Murah	35	Laris
Mei	Shell X	Oli	Murah	20	Laris
Mei	Shell UTR (0,8l)	Oli	Murah	1	Tidak Laris
Mei	Shell UTR (1l)	Oli	Murah	6	Tidak Laris
Mei	Shell UJ (0,8l)	Oli	Murah	9	Tidak Laris
Mei	Shell UJ (1l)	Oli	Murah	2	Tidak Laris
Mei	Shell Matic 2 (0,8l)	Oli	Mahal	1	Tidak Laris
Mei	Shell Matic 2 (1l)	Oli	Mahal	3	Tidak Laris
Mei	Federal X	Oli	Murah	20	Laris
Mei	Federal Racing	Oli	Murah	1	Tidak Laris
.....
Oktober	Air Radiator	Air Radiator	Murah	3	Tidak Laris

Atribut harga diganti menjadi atribut harga tetapi dengan interval yang lebih kecil yaitu murah dan mahal. Kategori harga dapat dilihat pada Tabel 10.

Tabel 10 Kategori Atribut Harga

Range harga	Kategori
Diatas Rp 60000	Mahal
Dibawah Rp 60000	Murah

Kategori harga dibagi menjadi dua yaitu murah dan mahal. Range yang termasuk kategori mahal adalah diatas Rp 60000, dan range yang termasuk kategori murah adalah dibawah Rp 60000 (Sumber : Pemilik Bengkel pada lampiran 7).

b. Proses transformasi pada atribut total penjualan yang ditransformasikan menjadi atribut total penjualan tetapi dengan kategori banyak, sedang dan sedikit berdasarkan hasil dari wawancara. Hasil transformasi sesuai dengan Tabel 11.

Tabel 11. Atribut Total Penjualan yang ditransformasikan

Bulan	Nama Produk	Jenis Produk	Harga	Total Penjualan	Status Penjualan
Mei	Yamalube Matic	Oli	Murah	Banyak	Laris
Mei	Yamalube Silver	Oli	Murah	Banyak	Laris
Mei	Yamalube Gold	Oli	Murah	Sedikit	Tidak Laris
Mei	Yamalube Sport	Oli	Mahal	Sedikit	Tidak Laris
Mei	Yamalube Super Matic	Oli	Mahal	Sedikit	Tidak Laris
Mei	Yamalube XP	Oli	Murah	Sedikit	Tidak Laris
Mei	Gear Yamalube 1	Oli	Murah	Sedikit	Tidak Laris
Mei	Gear Yamalube 2	Oli	Murah	Sedikit	Tidak Laris
Mei	Gear Yamalube 3	Oli	Murah	Sedikit	Tidak Laris
Mei	MPX 1 (0,8l)	Oli	Murah	Sedikit	Tidak Laris
Mei	MPX 1 (1l)	Oli	Mahal	Sedikit	Tidak Laris
Mei	MPX 2	Oli	Murah	Banyak	Laris
Mei	MPX Gear	Oli	Murah	Banyak	Laris
Mei	Enduro 4T	Oli	Murah	Sedikit	Tidak Laris
Mei	Enduro Racing	Oli	Mahal	Sedikit	Tidak Laris
Mei	Enduro Matic	Oli	Murah	Sedikit	Tidak Laris
Mei	Mesran X	Oli	Murah	Sedikit	Tidak Laris
Mei	Shell Matic	Oli	Murah	Banyak	Laris
Mei	Shell X	Oli	Murah	Banyak	Laris
Mei	Shell UTR (0,8l)	Oli	Murah	Sedikit	Tidak Laris
Mei	Shell UTR (1l)	Oli	Murah	Sedikit	Tidak Laris
Mei	Shell UJ (0,8l)	Oli	Murah	Sedikit	Tidak Laris
Mei	Shell UJ (1l)	Oli	Murah	Sedikit	Tidak Laris
Mei	Shell Matic 2 (0,8l)	Oli	Mahal	Sedikit	Tidak Laris
Mei	Shell Matic 2 (1l)	Oli	Mahal	Sedikit	Tidak Laris
Mei	Federal X	Oli	Murah	Banyak	Laris
Mei	Federal Racing	Oli	Murah	Sedikit	Tidak Laris
.....
Oktober	Air Radiator	Air Radiator	Murah	Sedikit	Tidak Laris

Atribut total penjualan diganti menjadi atribut total penjualan tetapi dengan interval yang lebih kecil yaitu banyak, sedang dan sedikit seperti pada Tabel 12.

Tabel 12. Kategori Atribut Total Penjualan

Range Total Penjualan	Kategori
≥ 20	Banyak
10-19	Sedang
1-9	Sedikit

Data Mining

a. Data Training

Data training yang digunakan terlihat pada Tabel 13.

Tabel 13. Sampel Data Training

No	Bulan	Nama Produk	Jenis Produk	Harga	Total Penjualan	Status Penjualan
1	Mei	Yamalube Matic	Oli	Murah	Banyak	Laris
2	Mei	Yamalube Silver	Oli	Murah	Banyak	Laris
3	Mei	Yamalube Gold	Oli	Murah	Sedikit	Tidak Laris
4	Mei	Ban Dalam	Ban	Murah	Sedang	Laris
5	Mei	Ban Luar	Ban	Mahal	Sedikit	Tidak Laris
6	Mei	Yamalube XP	Oli	Murah	Sedikit	Tidak Laris
7	Mei	Busi Daytona	Busi	Murah	Banyak	Laris
8	Mei	Busi Denso	Busi	Murah	Sedang	Tidak Laris
9	Mei	Gear Yamalube 3	Oli	Murah	Sedikit	Tidak Laris
10	Mei	Bola Lampu Denshin	Lampu	Murah	Sedikit	Tidak Laris
11	Mei	Bola Lampu Osram	Lampu	Murah	Sedikit	Tidak Laris
12	Mei	MPX 2	Oli	Murah	Banyak	Laris
13	Mei	MPX Gear	Oli	Murah	Banyak	Laris
14	Mei	Dispet	Dispet	Murah	Sedang	Laris
15	Mei	Klahar	Klahar	Murah	Sedang	Tidak Laris
16	Mei	Enduro Matic	Oli	Murah	Sedang	Tidak Laris
17	Mei	Mesran X	Oli	Murah	Sedikit	Tidak Laris
18	Mei	Baterai	Baterai	Mahal	Sedikit	Laris
19	Mei	Baterai NKP	Baterai	Mahal	Sedikit	Laris
20	Mei	Shell UTR (0,8l)	Oli	Murah	Sedikit	Tidak Laris

b. Data Testing

Model simulasi data testing diperoleh dari sumber data berjumlah 20 record.

Tabel 14. Sampel Data Testing

No	Bulan	Nama Produk	Jenis Produk	Harga	Total Penjualan	Status Penjualan
1	Mei	Yamalube Matic	Oli	Murah	Banyak	Laris
2	Mei	Yamalube Silver	Oli	Murah	Banyak	Laris
3	Mei	Yamalube Gold	Oli	Murah	Sedikit	Tidak Laris
4	Mei	Ban Dalam	Ban	Murah	Sedang	Laris
5	Mei	Ban Luar	Ban	Mahal	Sedikit	Tidak Laris
6	Mei	Yamalube XP	Oli	Murah	Sedikit	Tidak Laris
7	Mei	Busi Daytona	Busi	Murah	Banyak	Laris
8	Mei	Busi Denso	Busi	Murah	Sedang	Tidak Laris
9	Mei	Gear Yamalube 3	Oli	Murah	Sedikit	Tidak Laris
10	Mei	Bola Lampu Denshin	Lampu	Murah	Sedikit	Tidak Laris
11	Mei	Bola Lampu Osram	Lampu	Murah	Sedikit	Tidak Laris
12	Mei	MPX 2	Oli	Murah	Banyak	Laris
13	Mei	MPX Gear	Oli	Murah	Banyak	Laris
14	Mei	Dispet	Dispet	Murah	Sedang	Laris
15	Mei	Klahar	Klahar	Murah	Sedang	Tidak Laris
16	Mei	Enduro Matic	Oli	Murah	Sedang	Tidak Laris
17	Mei	Mesran X	Oli	Murah	Sedikit	Tidak Laris
18	Mei	Baterai	Baterai	Mahal	Sedikit	Laris
19	Mei	Baterai NKP	Baterai	Mahal	Sedikit	Laris
20	Mei	Shell UTR (0,8l)	Oli	Murah	Sedikit	Tidak Laris

Algoritma Naïve Bayes digunakan sebagai model untuk proses simulasi pengujian dengan melakukan perhitungan terhadap nilai likelihood dari setiap kelas.

a. Perhitungan Nilai Prior

Dilakukan perhitungan nilai pada setiap kelas data yang ada. Cara perhitungan nilai prior sesuai dengan persamaan 2. Di mana untuk nilai prior kelas laris didapatkan nilai 0.45 dan untuk kelas tidak laris didapatkan nilai 0.55.

b. Perhitungan nilai likelihood

Menghitung nilai probabilitas sesuai dengan persamaan 3. Sebagai contoh untuk bulan Mei dan laris didapatkan nilai 1 dan bulan Mei tidak laris didapatkan nilai 1. Untuk jenis produk Oli laris didapatkan nilai 0.44 dan produk Oli tidak laris didapatkan nilai 0.54. Hasil *likelihood* seperti pada Tabel 15.

Tabel 15 Perhitungan Likelihood

Atribut	Laris	Tidak Laris	Probabilitas		
			Laris	Tidak Laris	
Total Kasus	20	9	11	0,45	0,55
Bulan	9	11	1	1	
Jenis Produk					
Oli	4	6	0,44	0,54	
Busi	1	1	0,11	0,09	
Lampu	0	2	0	0,18	
Ban	1	1	0,11	0,09	
Dispet	1	0	0,11	0	
Klabar	0	1	0	0,11	
Baterai	2	0	0,22	0	
Harga					
Murah	7	10	0,77	0,90	
Mahal	2	1	0,22	0,09	
Atribut	Laris	Tidak Laris	Probabilitas		
			Laris	Tidak Laris	
Total Penjualan					
Sedikit	2	8	0,22	0,72	
Sedang	2	3	0,22	0,27	
Banyak	5	0	0,55	0	

c. Perhitungan nilai probabilitas posterior

Jumlah Posterior yang dibutuhkan adalah jumlah posterior pada setiap kelas tingkat kelulusan sesuai dengan persamaan 4. Sebagai contoh untuk probabilitas posterior pada data testing pertama untuk status laris didapatkan nilai 0.083 dan untuk status tidak laris didapatkan nilai 0. Hasil perhitungan nilai posterior dapat dilihat pada Tabel 16. Hasil prediksi status penjualan diperoleh dengan membandingkan nilai posterior. Jika nilai posterior kelas laris lebih tinggi dari kelas tidak laris, maka hasilnya laris begitu juga sebaliknya.

Tabel 16. Perhitungan Posterior

No	Sampel	Posterior Laris	Posterior Tidak Laris	Kelas	Prediksi
1	Sampel 1	0,083	0	Laris	Laris
2	Sampel 2	0,083	0	Laris	Laris
3	Sampel 3	0,033	0,192	Tidak Laris	Tidak Laris
4	Sampel 4	0,008	0,012	Laris	Tidak Laris
5	Sampel 5	0,005	0,003	Tidak Laris	Laris
6	Sampel 6	0,033	0,192	Tidak Laris	Tidak Laris
7	Sampel 7	0,002	0	Laris	Laris
8	Sampel 8	0,008	0,012	Tidak Laris	Tidak Laris
9	Sampel 9	0,033	0,192	Tidak Laris	Tidak Laris
10	Sampel 10	0	0,064	Tidak Laris	Tidak Laris
11	Sampel 11	0	0,064	Tidak Laris	Tidak Laris
12	Sampel 12	0,083	0	Laris	Laris
13	Sampel 13	0,083	0	Laris	Laris
14	Sampel 14	0,008	0	Laris	Laris
15	Sampel 15	0,033	0,072	Tidak Laris	Tidak Laris
16	Sampel 16	0,033	0,192	Tidak Laris	Tidak Laris
17	Sampel 17	0,033	0,192	Tidak Laris	Tidak Laris
18	Sampel 18	0,005	0	Laris	Laris
19	Sampel 19	0,005	0	Laris	Laris
20	Sampel 20	0,033	0,192	Tidak Laris	Tidak Laris

Pengujian Menggunakan Rapid Miner

Untuk mempermudah dalam proses pengujian algoritma digunakan aplikasi rapid miner. Pengujian ini menggunakan data penjualan sebagai dataset testing. Saat menggunakan dataset asli akan menghasilkan hasil yang lebih akurat daripada menggunakan sampel data pada pengujian manual.

accuracy: 75.00% +/- 26.35% (micro average: 75.00%)			
	true Laris	true Tidak Laris	class precision
pred. Laris	7	3	70.00%
pred. Tidak Laris	2	8	80.00%
class recall	77.78%	72.73%	

Gambar 1 Hasil Pengujian Algoritma Naïve Bayes Menggunakan Data Simulasi

Nilai yang didapat pada Gambar 1 kemudian dimasukkan ke dalam tabel confusion matrix untuk menghitung nilai performanya sesuai dengan Tabel 17.

Tabel 17 Confusion Matrix Data Simulasi

Prediction Sample			
Actual Sample		True	False
	True	7	3
	False	2	8

Perhitungan nilai akurasi sesuai dengan persamaan 5 didapatkan hasil sebesar 75%, nilai presisi sesuai dengan persamaan 6 didapatkan nilai sebesar 77.77% dan nilai recall sesuai dengan persamaan 7 didapatkan nilai sebesar 70%.

Sedangkan untuk pengujian menggunakan semua data penjualan dapat dilihat pada Gambar 2.

accuracy: 86.77% • +/- 5.04% (micro average: 86.74%)			
	true Laris	true Tidak Laris	class precision
pred. Laris	76	24	76.00%
pred. Tidak Laris	13	166	92.74%
class recall	85.39%	87.37%	

Gambar 2 Hasil Pengujian Keseluruhan Data Menggunakan Rapid Miner

Keseluruhan data kemudian dimasukkan ke dalam tabel confusion matrix untuk mendapat nilai akurasi, presisi dan recall seperti terlihat pada Tabel 18.

Tabel 18 Confusion Matrix Pengujian Keseluruhan Data

Prediction Sample			
Actual Sample		True	False
	True	76	24
	False	13	166

Nilai akurasi didapatkan sebesar 86.77%, nilai presisi sebesar 85.39% dan nilai recall sebesar 76%.

Tabel 19 Hasil Pengolahan Dengan Rapid Miner

...	Status Pe...	prediction(Statu...	confide...	confide...	Bulan	Nama Barang	Jenis Barang	Harga	Total Penjualan
1	Tidak Laris	Laris	0.695	0.305	Mai	Busi Denso	Busi	Murah	Sedang
2	Laris	Laris	1.000	0.000	Juni	MPX 2	Oli	Murah	Banyak
3	Tidak Laris	Tidak Laris	0.001	0.999	Juni	Enduro 4T	Oli	Murah	Sedikit
4	Tidak Laris	Tidak Laris	0.387	0.613	Juni	Busi Denso	Busi	Murah	Sedikit
5	Tidak Laris	Tidak Laris	0.000	1.000	Juli	Yamalube Gold	Oli	Murah	Sedikit
6	Tidak Laris	Tidak Laris	0.000	1.000	Juli	Enduro 4T	Oli	Murah	Sedikit
7	Laris	Laris	1.000	0.000	Juli	Shell Matic	Oli	Murah	Banyak
8	Tidak Laris	Tidak Laris	0.000	1.000	Juli	Federal Racing	Oli	Murah	Sedikit
9	Tidak Laris	Tidak Laris	0.000	1.000	Agustus	Yamalube Super Matic	Oli	Mahal	Sedikit
10	Laris	Laris	1.000	0.000	Agustus	MPX 2	Oli	Murah	Banyak
11	Tidak Laris	Tidak Laris	0.349	0.651	Agustus	Shell UTR (0.8)	Oli	Murah	Sedang
12	Tidak Laris	Laris	0.553	0.447	Agustus	Busi Denso	Busi	Murah	Sedang
13	Tidak Laris	Tidak Laris	0.000	1.000	Septemb...	Yamalube Sport	Oli	Mahal	Sedikit
14	Tidak Laris	Tidak Laris	0.000	1.000	Septemb...	Yamalube XP	Oli	Murah	Sedikit
15	Tidak Laris	Tidak Laris	0.000	1.000	Septemb...	Shell Matic 2 (0.8)	Oli	Mahal	Sedikit

Interpretation/Evaluation

Nilai akurasi sebesar 86,77% termasuk dalam klasifikasi yang baik (good classification), nilai presisi sebesar 85.39% termasuk ke dalam nilai tinggi sedangkan nilai recall sebesar 76% termasuk ke dalam nilai tinggi

V. KESIMPULAN

Penentuan produk laris dan tidak laris dengan menggunakan metode klasifikasi dan algoritma naïve bayes pada aplikasi rapid miner diperoleh bahwa algoritma naïve bayes dapat dipakai untuk melakukan penelitian terkait produk lairs dan tidak laris pada bengkel motor Aldo Motor.

Berdasarkan hasil penelitian yang didapatkan, maka saran sebagai berikut:

1. Untuk penelitian selanjutnya data yang digunakan bisa

dilakukan penambahan pada berbagai data penjualan bengkel yang ada di Kota Palembang untuk dapat melihat produk yang diminat.

2. Melakukan perbandingan dengan algoritma klasifikasi lainnya seperti ID3, C.45, Neural Network, Support Vector Machine (SVM) dan Apriori.

DAFTAR PUSTAKA

[1] S. Rahmatullah, M. Mukrim, and M. N. Pramitha, "Data mining untuk menentukan produk terlaris menggunakan metode naive bayes," *J. Inf. Dan Komput.*, vol. 7, pp. 57–64, 2019, doi: 10.35959/jik.v7i2.150.

[2] 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, doi: 10.31004/innovative.v3i2.1293.

[3] H. D. Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *J. Inform.*, vol. 7, no. 1, pp. 1–7, 2020, doi: 10.31311/ji.v7i1.6203.

[4] Fajar Astuti Hermawati, "Data Mining," 1st ed., Yogyakarta: ANDI, 2013.

[5] and P. S. Usama Fayyad, Gregory Piatetsky-Shapiro, "From Data Mining to Knowledge Discovery in Databases," *AI Magazine*, Rhode Island, pp. 37–54, 1996.

[6] F. Harahap, N. E. Saragih, E. T. Siregar, and H. Sariangisah, "Penerapan Data Mining Dengan Algoritma Naive Bayes Classifier Dalam Memprediksi Pembelian Cat," *J. Ilm. Inform.*, vol. 9, no. 01, pp. 19–23, 2021, doi: 10.33884/jif.v9i01.3702.

[7] Han Jiawei and Kamber Micheline, "Data Mining: Concepts and Techniques," Second., San Francisco: Morgan Kaufmann, 2006.

[8] Kusriani and Luthfi Emha Taufiq, "Algoritma Data Mining," 1st ed., Yogyakarta: CV Andi Offset, 2009.

[9] Bustami, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi," *Techsi*, vol. 5, no. 2, 2013, doi: https://doi.org/10.29103/techsi.v5i2.154.

[10] F. Gorunescu, "Data Mining: Concepts, Models and Techniques," Verlag Berlin Heidelberg: Springer, 2011.

[11] J. Han, M. Kamber, and J. Pei, *Data Mining: Concept and Techniques*, Third Edit. Waltham: Morgan Kauffman Publishers, 2012.

[12] jagat S. Challa, P. Goyal, S. Nikhil, A. Mangla, and S. S, "DD-Rtree: A dynamic distributed data structure for efficient data distribution among cluster nodes for spatial data mining algorithms," in *2016 IEEE International Conference on Big Data (Big Data)*, IEEE, 2016, pp. 27–36.

[13] 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>
- [14] K. R. P. Lubis and S. Sitohang, “Penerapan Data Mining Dengan Metode Naive Bayes Classifier Pada Penjualan Barang Untuk Optimasi Strategi Pemasaran,” *Comput. Sci. Ind. Eng.*, vol. 9, no. 7, p. 30, 2023, doi: 10.33884/comasiejournal.v9i7.7910.
- [15] Juwita, M. Safii, and B. Efendi Damanik, “Algoritma Naive Bayes Untuk Memprediksi Penjualan Pada Toko VJCakes Pematang Siantar,” *J. Mach. Learn. Artif. Intell.*, vol. 1, no. 4, pp. 337–346, 2022, doi: 10.55123/jomlai.v1i4.1674.