

**PREDIKSI PENJUALAN TOKO SNACK MENGGUNAKAN
MODEL *FORECASTING NAÏVE BAYES* : STUDI KASUS
PADA [TOKO SNACK IDA]**

LAPORAN TUGAS AKHIR

Laporan ini Disusun untuk memenuhi salah satu syarat memperoleh
Gelar Sarjana (S1) pada Program Studi Teknik Informatika Fakultas
Teknologi Industri Universitas Islam Sultan Agung Semarang



DISUSUN OLEH :
MUMAMMAD HAIKAL ADZIM
NIM 32602000046

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM SULTAN AGUNG
SEMARANG**

2025

FINAL PROJECT
***SNACK SHOP SALES PREDICTION USING NAÏVE BAYES
FORECASTING MODEL : CASE STUDY ON [IDA SNACK
SHOP]***

*Proposed to complete the requirement to obtain a bachelor's degree
(S1) at Informatics Engineering Departement of Industrial Technology
Faculty Sultan Agung Islamic University*



ARRANGED BY :
MUHAMMAD HAIKALADZIM
NIM 32602000046

***MAJORING OF INFORMATICS ENGINEERING
INDUSTRIAL TECHNOLOGY FACULTY
SULTAN AGUNG ISLAMIC UNIVERSITY***

2025

**LEMBAR PENGESAHAN
TUGAS AKHIR**

**PREDIKSI PENJUALAN TOKO SNACK MENGGUNAKAN MODEL
FORECASTING NAÏVE BAYES : STUDI KASUS PADA [TOKO SNACK
IDA]**

MUHAMMAD HAIKAL ADZIM

NIM 32602000046

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir

Program Studi Teknik Informatika

Universitas Islam Sultan Agung

Pada tanggal : 24 Maret 2025

TIM PENGUJI UJIAN SARJANA :

Andi Riansyah, S.T., M.Kom

NIDN. 0609108802

(Ketua Penguji)

24-03-2025

Mustafa, S.T., M.M., M.Kom

NIDN. 0623117703

(Anggota Penguji)

24-03-2025

Ghufron, S.T., M.Kom

NIDN. 062079005

(Pembimbing 1)

24-03-2025

Semarang, 25 Maret 2025

Mengetahui,

Kaprodi Teknik Informatika
Universitas Islam Sultan Agung

Moch Taufik, ST., MT

NIDN. 0622037502

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini

Nama : Muhammad Haikal Adzim

NIM : 32602000046

Judul Tugas Akhir :Prediksi Penjualan Toko Snack Menggunakan Model
Forecasting Naive Bayes : Studi Kasus pada [Toko Snack Ida]

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 25 Maret 2025

Yang Menyatakan,



Muhammad Haikal Adzim

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Muhammad Haikal Adzim
NIM : 32602000046
Program Studi : Teknik Informatika
Fakultas : Teknologi industri
Alamat Asal : Pemalang

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan agung.

Semarang, 25 Maret 2025

Yang menyatakan,



Muhammad Haikal Adzim

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT, yang telah memberikan rahmat, taufik serta hidayah-Nya, sehingga laporan Tugas Akhir ini dapat terselesaikan.

Tanpa lupa penulis mengucapkan terima kasih kepada :

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto S.H., M.Hum yang mengizinkan penulis menimba ilmu di kampus ini.
2. Dekan Fakultas Teknologi Industri Ibu Dr. Hj. Novi Marlyana, ST., MT., IPU.
3. Dosen pembimbing 1 Bapak Ghufroon, S.T., M.Kom yang telah membimbing selama tugas akhir berlangsung.
4. Orang tua penulis yang telah mengizinkan untuk menyelesaikan laporan ini.
5. Dan kepada semua pihak yang tidak dapat saya sebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan laporan ini masih terdapat banyak kekurangan, untuk itu penulis mengharap kritik dan saran dari pembaca untuk sempurnanya laporan ini. Semoga dengan ditulisnya laporan ini dapat menjadi sumber ilmu bagi setiap pembaca.

Semarang, 27 Februari 2025



Muhammad Haikal Adzim

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	i
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR.....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iii
KATA PENGANTAR.....	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR	viii
DAFTAR TABEL	xi
ABSTRAK	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Pembatasan Masalah	3
1.4 Tujuan.....	3
1.5 Manfaat	3
1.6 Sistematika Penelitian	3
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	5
2.1 Tinjauan Pustaka	5
2.2 Dasar Teori	6
2.2.1 Toko Snack Ida.....	6
2.2.2 Prediksi dan Peramalan (<i>Forecasting</i>)	6
2.2.3 <i>Naïve Bayes</i>	7
2.2.4 <i>Python</i>	8
2.2.5 <i>Website</i>	8
BAB III METODE PENELITIAN	9

3.1	Metodologi Penelitian	9
3.1.1	Studi Literatur	10
3.1.2	Pengumpulan Data	12
3.1.3	Analisis kebutuhan	13
3.1.4	Perancangan Sistem	15
2.1.5	Implementasi Algoritma <i>Naïve Bayes</i>	18
3.2	Perancangan <i>User Interface</i>	32
3.2.1	<i>Login</i>	32
3.2.2	<i>Menu</i>	33
3.2.3	<i>Dashboard</i>	33
3.2.4	Laporan Penjualan	34
3.2.5	Prediksi Penjualan	35
3.2.6	<i>Logout</i>	42
BAB IV HASIL DAN ANALISIS PENELITIAN		43
4.1	Analisis Penelitian	43
4.1.1	Mempersiapkan data	43
4.1.2	Pembagian data <i>training</i> dan data <i>testing</i>	44
4.1.3	Evaluasi model	52
4.2	Hasil Sistem	55
4.2.1	Halaman <i>Login</i>	55
4.2.2	Halaman <i>Menu</i>	55
4.2.3	Halaman Utama (<i>Dasboard</i>)	56
4.2.4	Halaman Laporan Penjualan	56
4.2.5	Halaman Prediksi Penjualan	59
4.2.6	Halaman Logout	67

BAB V KESIMPULAN DAN SARAN	68
5.1 Kesimpulan	68
5.2 Saran.....	69
DAFTAR PUSTAKA.....	70

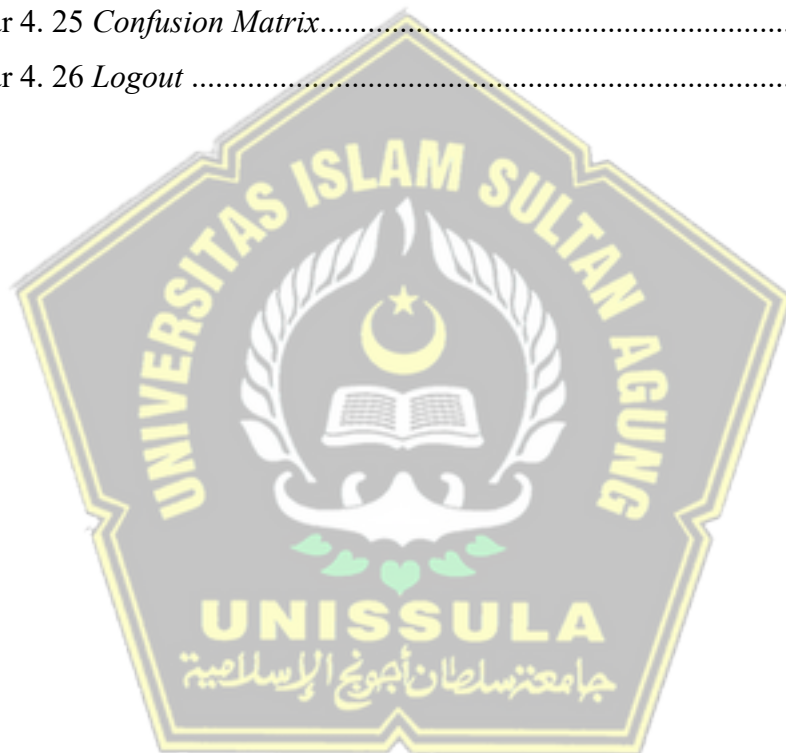


DAFTAR GAMBAR

Gambar 3. 1 <i>System Model</i>	9
Gambar 3. 2 <i>Flowchart</i> Penelitian	10
Gambar 3. 3 <i>Flowchart</i> Perancangan Sistem	16
Gambar 3. 4 <i>Use Case Diagram</i>	17
Gambar 3. 5 Kode <i>Drop Data</i>	19
Gambar 3. 6 Kode <i>Cleaning Data</i>	21
Gambar 3. 7 Kode Klasifikasi <i>Naive Bayes</i>	22
Gambar 3. 8 Kode Memuat Variabel <i>x</i>	23
Gambar 3. 9 Kode Memuat Variabel <i>Dependen</i>	23
Gambar 3. 10 Kode <i>Library Sklearn</i>	24
Gambar 3. 11 Kode <i>Split Data</i>	24
Gambar 3. 12 Kode <i>x Train</i>	25
Gambar 3. 13 Kode <i>y Train</i>	25
Gambar 3. 14 Kode <i>x Test</i>	26
Gambar 3. 15 Kode Prediksi	27
Gambar 3. 16 Hasil Prediksi	27
Gambar 3. 17 Kode Probabilitas	27
Gambar 3. 18 Hasil Probabilitas	28
Gambar 3. 19 Kode Evaluasi Akurasi	29
Gambar 3. 20 Kode Evaluasi	29
Gambar 3. 21 Kode Laporan Klasifikasi	30
Gambar 3. 22 Hasil Klasifikasi Laporan	30
Gambar 3. 23 Kode <i>Confusion Matrix</i>	31
Gambar 3. 24 <i>Confusion Matrix</i>	31
Gambar 3. 25 <i>Login</i>	32
Gambar 3. 26 <i>Menu</i>	33
Gambar 3. 27 <i>Dashboard</i>	33

Gambar 3. 28 Laporan Penjualan.....	34
Gambar 3. 29 <i>Input</i> Data Penjualan	35
Gambar 3. 30 Grafik Penjualan.....	35
Gambar 3. 31 Atribut Data	36
Gambar 3. 32 Menghapus Kolom	36
Gambar 3. 33 <i>Cleaning</i> Data.....	37
Gambar 3. 34 Tahap Klasifikasi <i>Naïve Bayes</i>	38
Gambar 3. 35 Variabel <i>y</i>	38
Gambar 3. 36 Data Training Variabel <i>x</i>	39
Gambar 3. 37 Data Training Variabel <i>y</i>	39
Gambar 3. 38 Data Testing	40
Gambar 3. 39 Hasil Prediksi Penjualan.....	40
Gambar 3. 40 Probabilitas Prediksi.....	41
Gambar 3. 41 Evaluasi Model.....	41
Gambar 3. 42 <i>Logout</i>	42
Gambar 4. 1 <i>Login</i>	55
Gambar 4. 2 <i>Menu</i>	55
Gambar 4. 3 <i>Dashboard</i>	56
Gambar 4. 4 Laporan Penjualan.....	56
Gambar 4. 5 Upload Data.....	57
Gambar 4. 6 Pilih Tahun	57
Gambar 4. 7 Buat Laporan dan <i>Download</i>	57
Gambar 4. 8 Setelah di <i>Download</i>	58
Gambar 4. 9 Laporan Penjualan.....	58
Gambar 4. 10 Halaman Prediksi Penjualan	59
Gambar 4. 11 Upload Data.....	59
Gambar 4. 12 Data yang diunggah.....	60
Gambar 4. 13 Grafik Penjualan.....	60
Gambar 4. 14 Atribut Data	61
Gambar 4. 15 Hapus Kolom	61
Gambar 4. 16 <i>Cleaning</i> Data.....	62

Gambar 4. 17 Tahap Klasifikasi <i>Naive Bayes</i>	62
Gambar 4. 18 Variabel <i>y</i>	63
Gambar 4. 19 Pembagian Data Training dan Testing.....	63
Gambar 4. 20 Data Training Variabel <i>y</i>	64
Gambar 4. 21 Testing Data.....	64
Gambar 4. 22 Hasil Prediksi	65
Gambar 4. 23 Probabilitas Prediksi.....	66
Gambar 4. 24 Hasil Evaluasi Model	66
Gambar 4. 25 <i>Confusion Matrix</i>	67
Gambar 4. 26 <i>Logout</i>	67



DAFTAR TABEL

Tabel 3. 1 Analisis Sistem	13
Tabel 3. 2 Menghapus Kolom	20
Tabel 3. 3 <i>Cleaning</i> Data.....	22
Tabel 3. 4 Variabel x.....	23
Tabel 3. 5 Variabel y.....	24
Tabel 3. 6 x Train	25
Tabel 3. 7 y Train	26
Tabel 3. 8 Testing	26
Tabel 4. 1 Memuat Data	43
Tabel 4. 2 Training Data.....	44
Tabel 4. 3 Testing Data.....	45
Tabel 4. 4 Menghitung Probabilitas Kelas	45
Tabel 4. 5 Menghitung Probabilitas Kategori Rasa	46
Tabel 4. 6 Menghitung Probabilitas Kategori Jenis	46
Tabel 4. 7 Menghitung Probabilitas Kategori Kemasan	46
Tabel 4. 8 Menghitung Probabilitas Bulan.....	47
Tabel 4. 9 Menghitung Probabilitas Tahun	47
Tabel 4. 10 Evaluasi Model.....	52

ABSTRAK

Toko Snack Ida di Pasar Comal, Kabupaten Pemalang, menghadapi kendala dalam mengelola stok barang. Untuk mengatasinya, metode *Naïve Bayes* diterapkan guna mengklasifikasikan pola penjualan berdasarkan variabel seperti nama produk, rasa, jenis, kemasan, bulan, dan tahun. Dengan data *historis* sebanyak 1.620 *entri* (1.296 data latih, 324 data uji), model ini menghasilkan prediksi kategori Laris (123) dan Kurang Laris (201), dengan akurasi 0,59, presisi 0,59, *recall* 0,59, dan *F1-score* 0,58. Meskipun akurasi masih dapat ditingkatkan, model ini telah diimplementasikan dalam sistem berbasis *web* menggunakan *Streamlit* untuk memudahkan prediksi penjualan secara interaktif.

Kata kunci : prediksi penjualan, *naïve bayes*, toko snack

ABSTRACT

Snack Store Ida in Pasar Comal, Pemalang Regency, faces challenges in managing stock levels. To address this, the Naïve Bayes method is applied to classify sales patterns based on variables such as product name, flavor, type, packaging, month, and year. Using historical data of 1,620 entries (1,296 training data, 324 test data), the model predicts High Sales (123) and Low Sales (201) categories, achieving an accuracy of 0.59, precision of 0.59, recall of 0.59, and an F1-score of 0.58. Although the accuracy can be improved, the model has been implemented in a web-based system using Streamlit to facilitate interactive sales predictions.

Keywords: sales prediction, naïve Bayes, snack shop

BAB I PENDAHULUAN

1.1 Latar Belakang

Industri makanan ringan atau snack memiliki peran yang *signifikan* dalam pasar *ritel*, dengan pertumbuhan yang terus meningkat seiring dengan perubahan gaya hidup dan *preferensi* konsumen. Di tengah persaingan yang semakin ketat, manajemen toko *snack* memerlukan strategi yang efektif untuk mengoptimalkan penjualan, mengelola stok dengan baik, dan meningkatkan kepuasan pelanggan. *Forecasting* digunakan untuk memprediksi jumlah barang yang akan dibutuhkan dimasa mendatang berdasarkan pengalaman penjualan barang dimasa lampau (Wardhani dkk., 2022)

Forecasting atau peramalan stok barang memiliki potensi untuk mempengaruhi tingkat kelancaran penjualan barang tersebut, karena membantu dalam merencanakan produksi, pengadaan, dan strategi pemasaran yang lebih tepat sesuai dengan permintaan pasar yang diperkirakan. *Naïve Bayes* adalah metode klasifikasi probabilistik yang sederhana, di mana probabilitas diperkirakan dengan menghitung *frekuensi* dan kombinasi nilai dari data yang ada dalam dataset yang diberikan (Saputra dan Herdiansyah, 2022). Dengan menggunakan metode *Naïve Bayes* untuk meramalkan penjualan stok *snack*, pengelola dapat memperoleh pemahaman yang lebih baik mengenai seberapa laku stok tersebut dipasaran.

Toko Snack Ida dibangun pada tahun 2002. Toko snack ida menghadapi permasalahan dalam pengelolaan barang yang kurang stabil, sehingga sering kali mengalami kesulitan dalam menentukan jumlah stok yang optimal. Untuk mengatasi hal tersebut, penelitian ini berencana mengembangkan sebuah aplikasi prediksi yang dapat menganalisis data penjualan dan memberikan estimasi mengenai tingkat keberhasilan penjualan suatu produk, apakah termasuk kategori laris atau kurang laris. dengan adanya aplikasi ini, diharapkan pengelolaan stok dapat lebih efektif sehingga mengurangi resiko kelebihan atau kekurangan barang serta meningkatkan efesiensi operasional toko. Algoritma *naïve bayes* adalah klasifikasi data yang bekerja secara efektif dengan memaksimalkan pengawasan dalam estimasi probabilitas secara akurat, berdasarkan asumsi bahwa setiap atribut

bersifat *independen* secara kondisional terhadap atribut lainnya jika nilai *output* telah diketahui (Juwita dkk., 2022). Proses pemodelan bertujuan untuk membangun pola dan menguji model *naïve bayes* dengan membagi dataset menjadi dua yaitu data *training* dan data *testing* dengan proporsi 80:20, dimana 80% digunakan untuk pelatihan dan 20% untuk pengujian model (Pramana dkk., 2023). Algoritma *naïve bayes* digunakan untuk mengelompokkan ulasan produk ke dalam tiga kategori sentimen, yaitu positif, negatif dan netral. Evaluasi kinerja model menunjukkan bahwa tingkat akurasi yang mencapai adalah 75,90%. Selain itu, nilai presisi sebesar 66,45% dan recall sebesar 97,42% juga menjadi aspek penting yang diperhatikan (Hajaroh dkk., 2024).

Tujuan dari pengembangan aplikasi prediksi ini adalah untuk mengembangkan model prediksi penjualan snack menggunakan metode *naïve bayes* serta menerapkan pada toko snack idn guna menganalisis *tren* penjualan berdasarkan data *historis* dari tahun 2017 hingga mei 2024, meningkatkan keuntungan dengan menyesuaikan jumlah produk yang tersedia sesuai dengan permintaan konsumen. Pengumpulan data penjualan makanan ringan dari januari hingga desember dalam rentang tahun 2018 – 2021 menjadi sumber informasi yang berharga untuk analisis. Dengan menerapkan algoritma *naïve bayes* yang merupakan salah satu metode umum dalam analisis *prediktif*, toko ini dapat memperkirakan harga penjualan makanan ringan dengan lebih akurat (Gulo dkk., 2023)

Manfaat dari penelitian ini membantu pemilik toko dalam mengoptimalkan strategi penjualan dengan menganalisis data penjualan untuk mengidentifikasi produk yang berpotensi laris dan kurang laris. Dengan demikian, pemilik toko dapat mengelola stok secara lebih efisien, mengurangi resiko kelebihan atau kekurangan barang serta meningkatkan keuntungan melalui perencanaan yang lebih tepat berdasarkan pola permintaan konsumen.

Sistem ini dirancang untuk membantu pemilik toko dalam mengelola penjualan serta memprediksi tren penjualan pada tahun mendatang yaitu 2025. Selain itu, sistem ini dapat digunakan oleh pengguna untuk mengelola toko dengan lebih efektif serta memantau laporan penjualan secara sistematis.

1.2 Perumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut :

1. Bagaimana menerapkan metode *Naïve Bayes* untuk memprediksi penjualan di toko snack ida?
2. Apa saja faktor – faktor yang mempengaruhi penjualan di toko snack ida yang perlu dipertimbangkan dalam model prediksi?

1.3 Pembatasan Masalah

Adapun batasan masalah dari penulisan ini yaitu sebagai berikut:

1. Evaluasi terhadap akurasi prediksi yang dihasilkan oleh model *Naïve Bayes*.
2. Penggunaan metode *Naïve Bayes* sebagai metode untuk prediksi penjualan.

1.4 Tujuan

Tujuan penelitian ini adalah :

1. Mengimplementasikan metode *Naïve Bayes* untuk memprediksi penjualan di Toko Snack Ida.
2. Menganalisis faktor – faktor yang mempengaruhi penjualan ditoko tersebut dan mengintegrasikan ke dalam model prediksi.

1.5 Manfaat

Penelitian ini membantu pemilik toko dalam mengoptimalkan strategi penjualan dan mengelola stok dengan efisien mengidentifikasi produk yang berpotensi laris dan kurang laris.

1.6 Sistematika Penelitian

Penulisan dalam membuat laporan tugas akhir mengikuti sistematika penulisan sebagai berikut :

BAB I : PENDAHULUAN

Pada bab ini penulis mengutarakan latar belakang pemilihan judul, rumusan masalah, tujuan penelitian, metodologi penelitian, serta sistematika penulisan.

BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI

Bab ini memuat penelitian – penelitian sebelumnya dan dasar teori yang berguna untuk membantu sebuah penulis untuk memahami bagaimana teori yang berhubungan dengan algoritma *naïve bayes* untuk penelitian ini.

BAB III : METODE PENELITIAN

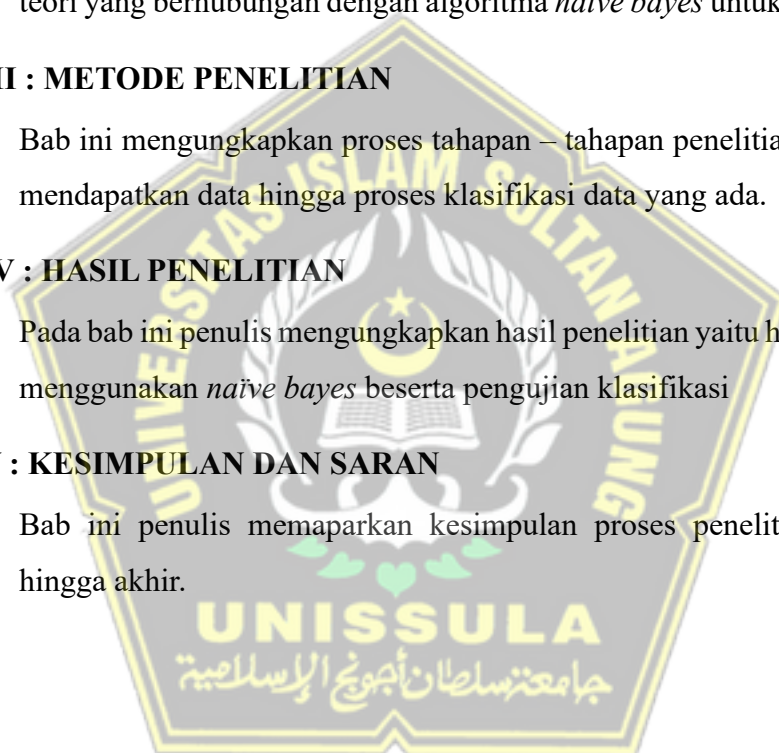
Bab ini mengungkapkan proses tahapan – tahapan penelitian dimulai dari mendapatkan data hingga proses klasifikasi data yang ada.

BAB IV : HASIL PENELITIAN

Pada bab ini penulis mengungkapkan hasil penelitian yaitu hasil klasifikasi menggunakan *naïve bayes* beserta pengujian klasifikasi

BAB V : KESIMPULAN DAN SARAN

Bab ini penulis memaparkan kesimpulan proses penelitian dari awal hingga akhir.



BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Forecasting (peramalan) secara umum adalah suatu perkiraan, namun dengan menggunakan metode – metode tersebut, peramalan menjadi lebih dari sekedar perkiraan. Selain itu, *forecasting* merupakan proses perkiraan berbagai kebutuhan dimasa depan, termasuk kebutuhan dalam hal kuantitas, kualitas, waktu dan lokasi yang diperlukan untuk memenuhi permintaan barang atau jasa. (Wardhani dkk., 2022).

Pembahasan implementasi algoritma lainnya berkaitan yaitu penelitian oleh Azhar Syaputra dkk (2020). Pada penelitiannya mengenai “Implementasi Metode *Double Exponential Smoothing* Pada Aplikasi Prediksi Harga Emas Dunia” dengan hasil dan penggunaan *smoothing* atau pemulusan peramalan pertama dan kedua untuk implementasi algoritma pada prediksi pergerakan harga emas tersebut (Al Husaini dkk., 2022).

Penelitian yang dilakukan oleh Anugrah Angga Ronaldi dan rekan – rekannya pada tahun 2020 menganalisi data penjualan CV Mitra Sejati dari tahun 2017, 2018, 2019 dan 2020. Hasil perhitungan prediksi menggunakan algoritma *Naïve Bayes* menunjukkan tingkat akurasi prediksi sebesar 94,59%. Untuk kelas presisi, kategori “Ya” mencapai 100.000% dan kategori “Tidak” mencapai 94,44%, sedangkan kelas *recall*, kategori “Ya” mencapai 33 (Supendar dkk., 2023).

Dari penelitian Husaini dkk., (2024) hasil klasifikasi data uji menunjukkan akurasi sebesar 90,62% dengan rincian 3 klasifikasi salah dan 29 data yang benar. Saputra dkk., (2022) akurasi penelitian ini mencapai 89.47% dengan rincian hasil yang meningkat dan menurun. Penelitian ini menunjukkan bahwa ada kecenderungan hasil prediksi yang sejalan dengan data nyata dilapangan. Penelitian yang akan dilakukan direncanakan untuk mendapatkan hasil perkiraan akurasi sekitar 92,00%, hasil klasifikasi yang akan dicapai lebih unggul dibandingkan dengan penelitian husaini dkk., (2024) yang memiliki akurasi 90,62% dan saputra dkk., (2022) dengan akurasi 89,47%, mengidentifikasi bahwa metode *naïve bayes*

yang digunakan dalam penelitian ini lebih efektif dalam menghasilkan prediksi yang sesuai dengan data nyata pada penjualan toko snack Ida.

2.2 Dasar Teori

2.2.1 Toko Snack Ida

Toko snack ida adalah salah satu toko yang berada di daerah pasar comal, kabupaten pemalang. Toko ini selalu menambahkan produk – produk terbaru dengan berbagai rasa, jenis makanan dan minuman serta kemasan setiap bulan hingga setiap tahun. Dalam metode ini, penulis terlibat langsung dalam aktivitas sehari – hari untuk mengamati objek penelitian, sambil mencatat berbagai kejadian yang terjadi di pt. tunas dwipa matra way kanan (Anggraini dkk., 2024).

Mengumpulkan data untuk melatih dan menguji model *naïve bayes* sangat penting. Semakin banyak data yang tersedia, semakin banyak baik model dalam menyelesaikan permasalahan (Harahap dkk., 2023). Pengumpulan data perlu dilakukan melalui observasi dan wawancara dengan pemilik toko agar data dapat dicatat dan dianalisis dalam sistem. Oleh karena itu, diperlukan sistem prediksi untuk mengetahui apakah suatu produk memiliki tingkat penjualan laris atau kurang laris dalam pengelolaan stok barang. Sistem ini akan membantu pemilik toko dalam memantau ketersediaan produk, sehingga dapat melakukan pembelian sesuai dengan *preferensi* konsumen.

2.2.2 Prediksi dan Peramalan (*Forecasting*)

Prediksi adalah proses memperkirakan secara sistematis tentang apa yang kemungkinan besar akan terjadi dimasa depan berdasarkan informasi yang ada dimasa lalu dan sekarang, dengan tujuan meminimalkan kesalahan (perbedaan antara kejadian yang sebenarnya terjadi dan perkiraan). Sedangkan peramalan (*forecasting*) adalah prosedur untuk menghasilkan informasi faktual mengenai situasi sosial di masa depan berdasarkan informasi yang sudah tersedia mengenai masalah kebijakan (Putra dkk., 2022).

2.2.3 Naïve Bayes

Naïve Bayes adalah metode klasifikasi terkenal yang menggunakan probabilitas sederhana. Metode ini juga dapat didefinisikan sebagai metode mesin learning yang mengandalkan perhitungan probabilitas dan statistik. Ditemukan oleh ilmuwan inggris, *thomas bayes*, metode ini digunakan untuk memperkirakan probabilitas kejadian dimasa depan berdasarkan data masa lalu (Rozi dkk., 2023).

Ada beberapa langkah dalam menentukan hasil dari algoritma *naïve bayes*, yaitu sebagai berikut :

1. Membaca Data Training
2. Menghitung probabilitas kelas
3. Menghitung probabilitas atribut
4. Menghitung probabilitas akhir setiap kelas
5. Kalikan semua variabel kelas

Untuk memahami algoritma *naïve bayes*, berikut adalah rumus umum *teorema bayes* yang menjadi dasar dari metode *naïve bayes* :

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

Dimana :

X : Data dengan *class* yang belum diketahui

H : *Hipotesis* data X merupakan suatu *class* spesifik

$P(H|X)$: probabilitas *hipotesis* H berdasarkan kondisi X (*posteriori probabilitas*)

$P(H)$: Probabilitas *hipotesis* H (*prior probabilitas*)

$P(X|H)$: Probabilitas X berdasarkan kondisi *hipotesis* H

$P(X)$: Probabilitas dari X

2.2.4 Python

Python adalah bahasa pemrograman yang mampu menjalankan berbagai instruksi secara langsung dan *interpretatif*, tanpa memerlukan proses kompilasi yang terpisah. *Python* juga mendukung paradigma pemrograman berorientasi objek

(*object oriented programming*) dan memiliki semantik dinamis yang meningkatkan keterbacaan sintaksnya (Zulkifli dkk., 2022).

Bahasa ini pertama kali muncul pada tahun 1991 dan dirancang oleh *guido van rossum*. *Python* mendukung hampir semua sistem operasi dan bahkan hampir semua distribusi *linux* sudah menyertakan *python* secara *default* (Aris, 2021).

Numpy adalah salah satu pustaka dalam bahasa pemrograman *python* yang digunakan untuk melakukan berbagai perhitungan matematis, seperti aljabar, statistika dan perhitungan kompleks lainnya (Putra dkk., 2024).

2.2.5 Website

Website adalah sekumpulan halaman *web* yang saling terkait dan diakses melalui internet menggunakan *browser web*. Halaman – halaman ini biasanya berbagi *domain* yang sama dan dapat diakses melalui alamat URL (*Unifrom Resource Locator*).

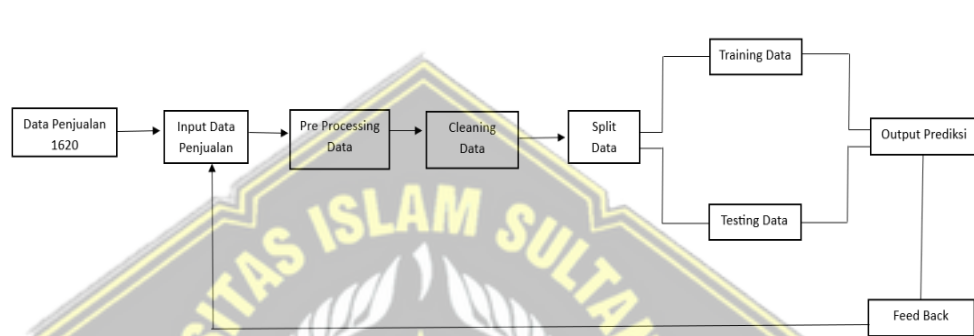
Streamlite merupakan *web application framework* yang dirancang untuk mempermudah *developer* dalam mengembangkan dan membangun aplikasi web, khususnya dibidang *machine learning* (Hadi & Benedict, 2024).

BAB III

METODE PENELITIAN

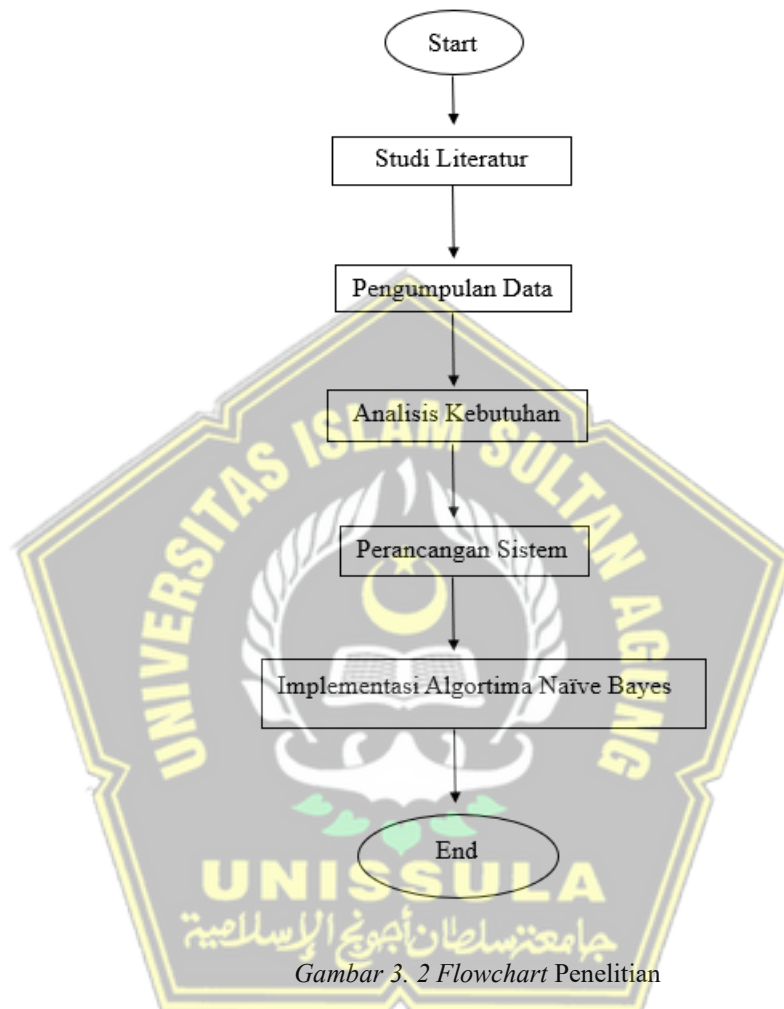
3.1 Metodologi Penelitian

Dalam penelitian ini, penjualan ditoko snack ida diprediksi menggunakan metode *naïve bayes*. Metode ini digunakan untuk mengidentifikasi dan memprediksi penjualan. Penelitian ini terdiri dari beberapa tahapan yang dilakukan secara sistematis. Berikut tahapan yang dilakukan pada penelitian ini.



Gambar 3. 1 System Model

Pada gambar 3. 1 dalam sistem model prediksi penjualan, data penjualan dalam format *excel* diunggah ke dalam sistem. Proses awal yang dilakukan adalah penghapusan kolom yang tidak diperlukan. Selanjutnya, dilakukan proses pembersihan data (*cleaning data*), di mana data dalam bentuk teks diubah menjadi format *numerik*. Setelah itu, data dibagi menjadi dua bagian dalam tahap pemisahan data (*split data*), yaitu 80% untuk data pelatihan (*training data*) dan 20% untuk data pengujian (*testing data*). Model kemudian dilatih menggunakan data pelatihan dan diuji dengan data pengujian. Setelah proses pengujian selesai, sistem menghasilkan *output* berupa hasil prediksi. Jika proses prediksi telah berhasil sistem akan kembali ke tahap awal untuk menerima input data baru.



3.1.1 Studi Literatur

Pada era digital dan persaingan bisnis yang semakin ketat prediksi penjualan menjadi salah satu aspek penting dalam strategi pemasaran dan pengelolaan stok barang, terutama bagi toko snack. Dengan menganalisis data penjualan pemilik toko dapat mengoptimalkan ketersediaan produk untuk mengatasi kehabisan barang atau stok serta meningkatkan kepuasan pelanggan. Salah satu metode yang digunakan dalam prediksi penjualan adalah *naïve bayes* sebuah algoritma klasifikasi berbasis probabilitas yang mampu mengolah data dengan cepat dan

efisien. penelitian ini menggunakan metode *naïve bayes* sebagai solusi untuk mengelola stok barang sehingga pemilik toko dapat mengetahui produk yang banyak diminati (laris) atau kurang diminati (kurang laris). Setelah itu, pemilik toko dapat menyesuaikan pembelian produk sesuai dengan kebutuhan pelanggan.

Sistem dalam penelitian ini memanfaatkan *google colab* untuk menganalisis dan menampilkan hasil prediksi.

Google colab atau sering disebut *google colab* adalah *platform* gratis dari *google* yang mendukung pembelajaran mesin. Layanan ini memungkinkan pengguna untuk menulis dan mengeksekusi kode *python*, serta mengunggah, menyimpan, dan berbagi kode maupun data (Al Fatah., 2024). Dalam tahap ini, dilakukan proses pembersihan data sebelum diolah. Pembersihan mencakup penghapusan beberapa kolom yang tidak diperlukan serta perubahan nama *header* agar lebih sesuai. (Harmaja dkk., 2022).

Pada langkah selanjutnya *transformasi* data, dilakukan beberapa operasi untuk meningkatkan kesiapan data dalam analisis, salah satunya adalah mengonversi format bulan dari teks ke angka agar analisis selanjutnya lebih mudah dilakukan (Purnama dkk., 2024). Klasifikasi merupakan proses pengelompokkan dan prediksi data baru berdasarkan atribut atau variabel dengan acuan dari data yang sudah tersedia. Dalam klasifikasi, terdapat dua jenis data yaitu data latih (*training data*) dan data uji (*testing data*) (Mahar dkk., 2023). Evaluasi *confusion matrix* di *google colab* menunjukkan hasil yang baik dalam mengenali sentimen penggunaan aplikasi grab. Dengan akurasi 87%, model ini dapat mengklasifikasikan sentimen dengan cukup akurat, termasuk dalam hal presisi, *recall* dan *f1-score* (Suryana dkk., 2024). Pengembangan ini dilakukan menggunakan *framework streamlite* yang dirancang khusus untuk keperluan analisis sentimen (Maulidan dkk., 2023). Maka, *framework streamlite* digunakan sebagai *platform* utama dalam pengembangan aplikasi ini guna mendukung analisis data, visualisasi hasil prediksi serta penyajian informasi yang interaktif dan mudah diakses oleh pengguna.

Penelitian ini bertujuan untuk mengembangkan sistem prediksi penjualan dengan model *naïve bayes* berbasis *web*. Kajian literatur menunjukkan bahwa metode *naïve bayes* sangat efektif untuk memprediksi penjualan toko snack.

Diharapkan bahwa penggunaan *naïve bayes* dapat memprediksi penjualan toko snack secara akurat dan bermanfaat bagi pemilik toko untuk mengetahui produk mana yang diprediksi laku pada tahun mendatang, sehingga pengelolaan produk dapat dilakukan secara efektif. Selain itu, penelitian ini juga berpotensi menjadi referensi untuk pengembangan sistem prediksi di masa depan.

3.1.2 Pengumpulan Data

Dalam penelitian ini, pengumpulan data melibatkan pengambilan informasi penjualan snack dari Toko Snack Ida, seperti laporan penjualan dan data transaksi, untuk mendapatkan variabel-variabel yang mempengaruhi penjualan.

Adapun pengumpulan data yang dilakukan pada penelitian ini adalah sebagai berikut :

1. Observasi

Melakukan observasi langsung di toko snack ida untuk mencatat penjualan, penilaian pelanggan, waktu pembelian dan faktor faktor lain yang mungkin berpengaruh. Pada penelitian ini mencatat data penjualan yang mencakup informasi mengenai produk, rasa, jenis, kemasan, isi, jumlah produk, harga produk, serta jumlah penjualan dari bulan hingga tahun. Dalam penelitian ini, data yang digunakan berasal dari periode januari 2017 hingga mei 2024 untuk berbagai jenis produk.

2. Wawancara

Melakukan wawancara dengan pemilik toko, karyawan dan pelanggan untuk mendapatkan informasi lebih mendalam mengenai pola penjualan, promosi yang dilakukan dan pereferensi pelanggan. Dalam penelitian ini, wawancara dilakukan untuk memperoleh informasi mengenai toko tersebut guna mendukung pengelolaan produk. Setelah wawancara dan pencatatan data penjualan, informasi yang diperoleh akan digunakan sebagai dasar dalam sistem prediksi untuk pengelolaan stok barang.

3.1.3 Analisis kebutuhan

Pada tahap penelitian analisis sistem, kebutuhan berbagai alat dan perangkat yang diperlukan untuk pengembangan sistem ini akan diidentifikasi. Berikut ini adalah alat dan perangkat yang akan digunakan tersedia dalam tabel 7.1 :

Tabel 3. 1 Analisis Sistem

No	Tool	Version
1	Laptop Asus Tuf Gaming FX505GT-FX50GT	Windows 11
2	Visual Studio Code	1.91.1
3	Excel	Office 2019
4	Web Browser	-
5	Drawio.io	1.0
6	Google colab	-
7	Figma	WEB
8	Microsoft word	Office 19

Untuk membangun aplikasi berbasis *web* menggunakan bahasa pemrograman *python*, diperlukan berbagai *library* yang mendukung pengembangan sistem tersebut.

Adapun *library* yang digunakan dalam pembuatan sistem adalah sebagai berikut :

1. *Streamlite*

Streamlite adalah sebuah *framework open – source* yang digunakan untuk membangun antarmuka pengguna interaktif untuk aplikasi data *science*. *Framework* ini dirancang untuk mempermudah pengembangan aplikasi *web* dengan menggunakan *python*.

2. *Pandas*

Pandas adalah *library* yang biasanya digunakan untuk mempermudah dalam mengolah dan menganalisis data. Dalam *library* ini *pandas* berfungsi untuk

mengolah data dalam bentuk *dataframe*, seperti membaca, menulis, dan memproses dataset. *Dataframe* adalah struktur data dua dimensi atau tabel yang terdiri dari baris dan kolom.

3. *Scikit – learn*

Scikit – learn adalah salah satu pustaka yang disediakan oleh *python* untuk memproses algoritma *meachine learning*. Dalam pustaka ini terdapat fungsi *train_test_split* yang digunakan untuk membagi dataset menjadi data latih dan data uji dalam proses *meachine learning*. Selain itu, terdapat *gaussian naïve bayes* (*GaussianNB*) yang digunakan sebagai model klasifikasi probabilistik. Pustaka ini menyediakan *clasification_report* dan *confusion_matrix* yang berfungsi untuk mengukur kinerja model berdasarkan metrik seperti akurasi, presisi, *recall* dan *f1 – score*. Selanjutnya, *labelencoder* digunakan untuk mengonversi data kategorikal menjadi angka sehingga dapat digunakan dalam model *meachine learning*.

4. *Matplotlib*

Matplotlib merupakan salah satu pustaka visualisasi data yang paling umum digunakan di *python*. Pustaka ini merupakan ekstensi matematika *numerik* dari pustaka *numpy*. *Matplotlib* ini digunakan untuk membuat grafik dan visualisasi data.

5. *Seaborn*

Seaborn adalah *library* visualisasi data berbasis *matplotlib* di *python* yang digunakan untuk membuat grafik yang lebih informatif dan menarik secara visual. *Seaborn* menyediakan antarmuka untuk menggambarkan grafik statistik seperti *heatmap confusion matrix*.

6. *Plotly express*

Plotly express merupakan visualisasi data berbasis *plotly* di *python* yang memungkinkan pembuatan grafik interaktif dengan kode yang sederhana dibandingkan dengan *plotly* standar. *Plotly express* ini digunakan untuk membuat grafik interaktif dalam aplikasi *web*.

7. *Numpy*

Numpy merupakan salah satu *library python* yang berfungsi untuk proses komputasi *numerik*. *Numpy* sendiri digunakan untuk mengolah data numerik seperti *array* dan perhitungan matematika.

8. *Time*

Time adalah *library bawaan python* yang digunakan untuk mengolah waktu seperti mendapatkan waktu saat ini, menghitung eksekusi program, menunda eksekusi dan mengonversi format waktu.

9. *Joblib*

Joblib adalah *library python* yang digunakan untuk penyimpanan dan pemrosesan paralel yang efisien. *Library* ini digunakan dalam *machine learning* untuk menyimpan model yang sudah dilatih agar dapat digunakan kembali tanpa perlu melatih ulang kembali.

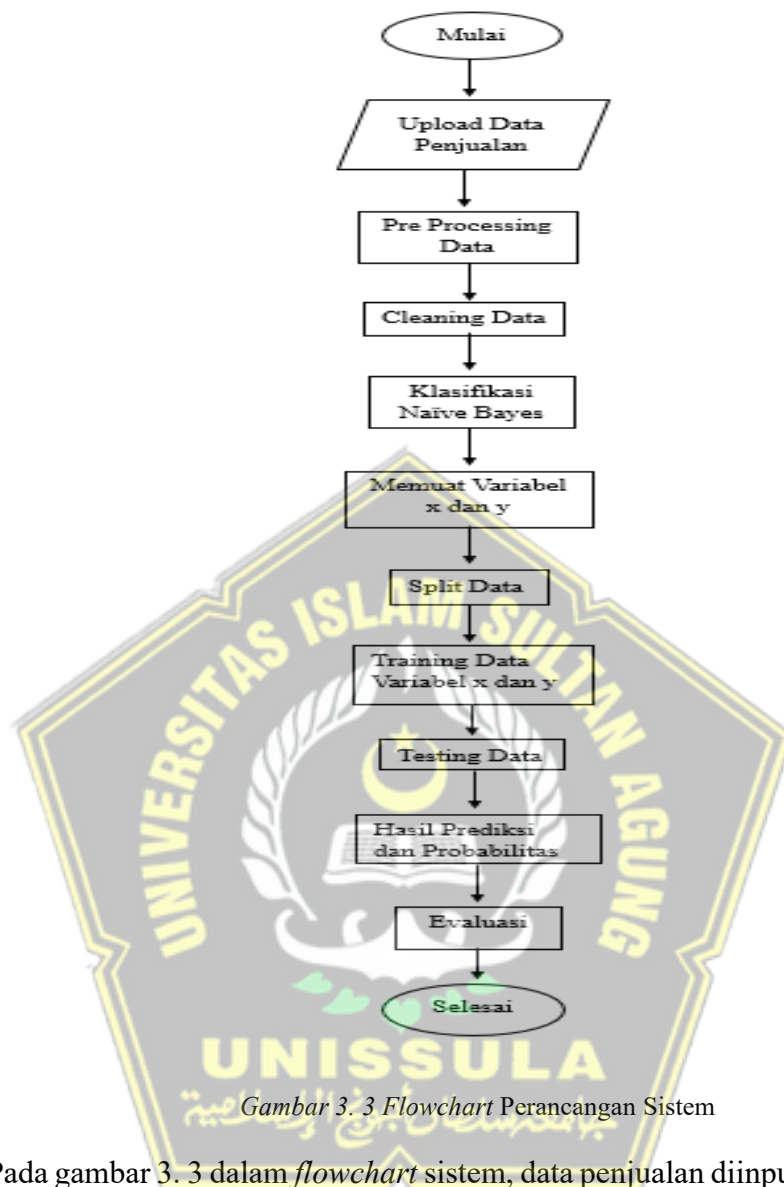
10. *Python – docx*

Python – docx adalah pustaka *python* yang digunakan untuk membuat, mengedit dan membaca dokumen *microsoft word* berformat (*docx*). Pustaka ini memungkinkan pengguna untuk mengatur perataan teks (*alignment*), mengubah ukuran dan format *font*, serta mengedit elemen XML dalam dokumen.

3.1.4 Perancangan Sistem

A. Analisis Alur Sistem

Dalam analisis alur sistem ini, terdapat *flowchart* yang menguraikan langkah - langkah urutan operasi dari sistem penelitian ini. Proses ini dimulai dari tahap penginputan data hingga menghasilkan keluaran berupa prediksi yang dihasilkan oleh sistem. *Flowchart* ini dapat dilihat pada gambar 7.3.



Gambar 3. 3 Flowchart Perancangan Sistem

Pada gambar 3. 3 dalam *flowchart* sistem, data penjualan diinputkan ke dalam sistem. Tahap pertama dalam pemrosesan adalah menghapus kolom yang tidak diperlukan. Setelah itu, data melakukan proses pembersihan yaitu mengubah data berbentuk teks menjadi format numerik. Selanjutnya data diklasifikasikan menggunakan algoritma *naïve bayes*. Pada tahap ini, dilakukan persiapan *variabel x* dan *variabel y*. setelah *variabel* disiapkan, data dibagi menjadi dua bagian yaitu data latih dan data uji. Data latih digunakan untuk melatih model pada *variabel x* dan *y*. setelah proses pelatihan selesai, model diuji menggunakan data uji untuk menghasilkan prediksi serta probabilitas dari data yang diuji. Tahap terakhir adalah

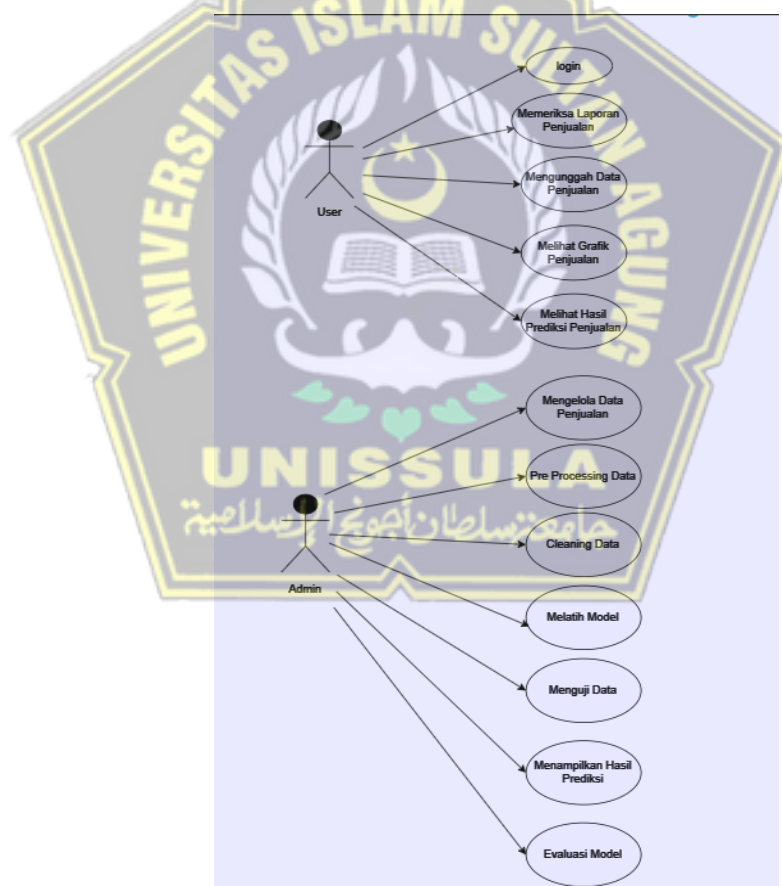
evaluasi hasil prediksi untuk menilai performa model. Setelah melakukan evaluasi proses prediksi telah selesai.

B. Perancangan Berorientasi Objek

Dalam perancangan berorientasi objek, analisis sistem dilakukan menggunakan aplikasi perancangan seperti drawio untuk menentukan use case diagram.

1. Use case diagram

Use case diagram adalah alat yang sangat berguna untuk menggambarkan interaksi antara pengguna (aktor) dan sistem. Dalam konteks proses login, dapat membuat use case diagram yang mencakup aktor dan use case utama.



Gambar 3. 4 Use Case Diagram

Gambar 3. 4 merupakan diagram *use case* yang menggambarkan interaksi antara dua aktor, yaitu *user* dan *admin* dengan sistem prediksi penjualan. *user*

memiliki beberapa fungsi, seperti *login*, memeriksa laporan penjualan, mengunggah data penjualan, melihat grafik penjualan dan melihat hasil prediksi penjualan. sementara itu, *admin* memiliki peran lebih luas termasuk mengelola data penjualan, melakukan *preprocessing* data, membersihkan data (*cleaning* data), melatih model, menguji data, menampilkan hasil prediksi dan mengevaluasi model. Diagram ini menunjukkan bagaimana masing – masing peran berinteraksi dengan sistem untuk menjalankan tugas mereka.

2.1.5 Implementasi Algoritma *Naïve Bayes*

Dalam penelitian ini, proses pengolahan data untuk membangun model prediksi penjualan dilakukan menggunakan *google colab* yang berfungsi sebagai platform untuk melatih dan menguji guna memperoleh hasil prediksi.

1. Data Penjualan

Penelitian ini menggunakan data penjualan snack ditoko ida dari tahun 2017 hingga mei 2024. Data tersebut diperoleh melalui wawancara dengan pemilik toko. Jumlah data yang digunakan dalam penelitian ini adalah 1.620 *entri* dengan 11 kolom. Data yang digunakan dalam analisis untuk memprediksi berasal dari produk aoka.

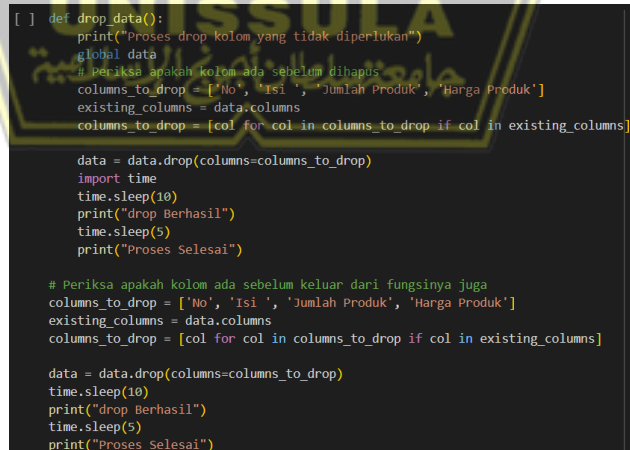
2. *Preprocessing* Data

Preprocessing data merupakan tahap awal dalam analisis data yang bertujuan untuk menyiapkan data agar dapat digunakan dalam model prediksi. Model *development* (pengembangan model) dimana proses ini membangun, melatih dan mengoptimalkan model *meachine learning* menggunakan data yang telah diproses. Tujuan model *development* membuat model yang dapat memprediksi penjualan snack dengan akurat, mengoptimalkan performa model sebelum digunakan aplikasi, memastikan model dapat bekerja dengan baik pada data. Proses ini mencakup pembersihan data dengan menghapus kolom yang tidak diperlukan, seperti no, isi, jumlah produk dan harga produk.

- a. Kolom no tidak digunakan karena hanya berfungsi sebagai nomor urut data dan tidak memiliki pengaruh terhadap prediksi penjualan.

- b. Kolom isi tidak digunakan karena tidak memberikan informasi langsung yang dapat digunakan dalam model klasifikasi. Model lebih berfokus pada fitur yang memiliki korelasi kuat dengan hasil prediksi.
- c. Kolom jumlah produk tidak digunakan karena sudah terdapat atribut lain yang memiliki kesamaan, seperti, jumlah penjualan. Penghapusan kolom ini bertujuan untuk menghindari redudansi data yang dapat menyebabkan bias dalam model.
- d. Kolom harga produk tidak digunakan karena dalam konteks prediksi, harga bukan merupakan faktor utama dalam menentukan kategori penjualan. Faktor ini lebih berkaitan dengan strategi bisnis yang berada di luar cakupan prediksi.

Selanjutnya data yang relevan, seperti nama produk, rasa, jenis, kemasan, bulan, tahun, dan jumlah penjualan, dipersiapkan untuk diproses dalam sistem. Faktor – faktor yang mempengaruhi penjualan seperti nama produk, rasa, jenis, kemasan, bulan, tahun, dan jumlah penjualan akan dianalisis untuk melihat pengaruhnya terhadap tingkat penjualan. Setelah itu, model *naive bayes* dilatih dengan mempertimbangkan faktor – faktor yang mempengaruhi penjualan, kemudian diuji menggunakan data testing untuk mengevaluasi kinerja model berdasarkan akurasi, *confusion matrix*, dan *clasification report*



```
[ ] def drop_data():
    print("Proses drop kolom yang tidak diperlukan")
    global data
    # Periksa apakah kolom ada sebelum dihapus
    columns_to_drop = ['No', 'isi ', 'Jumlah Produk', 'Harga Produk']
    existing_columns = data.columns
    columns_to_drop = [col for col in columns_to_drop if col in existing_columns]

    data = data.drop(columns=columns_to_drop)
    import time
    time.sleep(10)
    print("drop Berhasil")
    time.sleep(5)
    print("Proses Selesai")

    # Periksa apakah kolom ada sebelum keluar dari fungsinya juga
    columns_to_drop = ['No', 'isi ', 'Jumlah Produk', 'Harga Produk']
    existing_columns = data.columns
    columns_to_drop = [col for col in columns_to_drop if col in existing_columns]

    data = data.drop(columns=columns_to_drop)
    time.sleep(10)
    print("drop Berhasil")
    time.sleep(5)
    print("Proses Selesai")
```

Gambar 3. 5 Kode *Drop Data*

Pada gambar 3. 5 kode tersebut berfungsi untuk menghapus kolom yang tidak diperlukan dari sebuah *dataframe* bernama *data*, dengan terlebih dahulu memeriksa

keberadaan kolom yang akan dihapus agar tidak terjadi *error*. Selain itu, terdapat jeda waktu menggunakan *time.sleep()* untuk memberikan simulasi proses sebelum mencetak pesan keberhasilan dan penyelesaian proses.

Tabel 3. 2 Menghapus Kolom

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun	Jumlah Penjualan
Aoka	Cokelat	Roti	Sachet	Mei	2024	Laris
Aoka	Strawberry	Roti	Sachet	Mei	2024	Laris
Aoka	Pandan	Roti	Sachet	Mei	2024	Laris
Aoka	Duren	Roti	Sachet	Mei	2024	Kurang Laris
Aoka	Nanas	Roti	Sachet	Mei	2024	Kurang Laris
...	...	...	...	...	...	...
Aoka	Cokelat	Roti	Sachet	Januari	2017	Kurang Laris
Aoka	Strawberry	Roti	Sachet	Januari	2017	Kurang Laris
Aoka	Pandan	Roti	Sachet	Januari	2017	Kurang Laris
Aoka	Duren	Roti	Sachet	Januari	2017	Laris
Aoka	Nanas	Roti	Sachet	Januari	2017	Laris

Tabel 3.2 merupakan proses setelah menghapus kolom no, isi, jumlah produk, harga produk data yang akan diproses ke dalam sistem mencakup nama produk, rasa, jenis, kemasan, bulan dan jumlah penjualan.

3. *Cleaning Data*

Cleaning data adalah proses pembersihan, memperbaiki dan mengorganisir data agar siap digunakan untuk analisis atau pemodelan. Dalam proses ini, bertujuan untuk mengubah data kategorikal menjadi *numerik* agar dapat digunakan dalam model prediksi. Untuk data yang akan diubah ke dalam *numerik* yaitu nama produk, rasa, jenis, kemasan, bulan, tahun dan jumlah penjualan.

```

from os import name
def cleaning_data():
    global data
    nam = data['Nama Produk']
    nam = nam.drop_duplicates()
    nam = [nam for nam in nam]
    # Rentangnya harus (1, len(nam) + 1) untuk memastikan jumlah nilai yang benar
    i = [i for i in range(1, len(nam) + 1)]
    data['Nama Produk'] = data['Nama Produk'].replace(to_replace=nam, value=i)
    time.sleep(10)
    print("Cleaning Data Nama Produk Berhasil")

    ras = data['Rasa ']
    ras = ras.drop_duplicates()
    ras = [ras for ras in ras]
    # Rentangnya harus (1, len(ras) + 1) untuk memastikan jumlah nilai yang benar
    i = [i for i in range(1, len(ras) + 1)]
    data['Rasa '] = data['Rasa '].replace(to_replace=ras, value=i)
    time.sleep(10)
    print("Cleaning Data Rasa Berhasil")

    jen = data['Jenis']
    jen = jen.drop_duplicates()
    jen = [jen for jen in jen]
    # Rentangnya harus (1, len(jen) + 1) untuk memastikan jumlah nilai yang benar
    j = [j for j in range(1, len(jen) + 1)]
    data['Jenis'] = data['Jenis'].replace(to_replace=jen, value=j)
    time.sleep(10)
    print("Cleaning Data Jenis Berhasil")

    kem = data['Kemasan']
    kem = kem.drop_duplicates()
    kem = [kem for kem in kem]
    # Rentangnya harus (1, len(kem) + 1) untuk memastikan jumlah nilai yang benar
    k = [k for k in range(1, len(kem) + 1)]
    data['Kemasan'] = data['Kemasan'].replace(to_replace=kem, value=k)
    time.sleep(10)
    print("Cleaning Data Kemasan Berhasil")

    bul = data['Bulan ']
    bul = bul.drop_duplicates()
    bul = [bul for bul in bul]
    # Rentangnya harus (1, len(bul) + 1) untuk memastikan jumlah nilai yang benar
    l = [l for l in range(1, len(bul) + 1)]
    data['Bulan '] = data['Bulan '].replace(to_replace=bul, value=l)
    time.sleep(10)
    print("Cleaning Data Bulan Berhasil")

    # Validasi kolom 'Tahun'
    if 'Tahun' in data.columns:
        tah = data['Tahun'].drop_duplicates().reset_index(drop=True)
        tah_mapping = {v: m + 1 for m, v in enumerate(tah)}
        data['Tahun'] = data['Tahun'].map(tah_mapping)
        time.sleep(2)
        print("Cleaning Data Tahun Berhasil")

    # Validasi kolom 'Jumlah Penjualan'
    if 'Jumlah Penjualan' in data.columns:
        # Perbaikan pada pengantian nilai
        data['Jumlah Penjualan'] = data['Jumlah Penjualan'].replace({
            'Laris': 1,
            'Kurang Laris': 2
        })
        time.sleep(2)
        print("Cleaning Data Jumlah Penjualan Berhasil")

    time.sleep(1)
    print("Proses Selesai")

```

Gambar 3. 6 Kode *Cleaning Data*

Gambar 3. 6 kode ini melakukan proses *cleaning* data dengan mengganti nilai unik dalam beberapa kolom kategorikal dengan angka. Langkah – langkahnya adalah sebagai berikut :

- Menghapus duplikasi pada setiap kolom kategorikal seperti nama produk, rasa, jenis, kemasan, bulan dan tahun.
- Mengganti nilai unik dalam kolom dengan angka berurutan.
- Mengonversi jumlah penjualan dari kategori laris “1” dan kurang laris “2”.
- Menambah jeda waktu (*time.sleep*) setelah setiap langkah untuk memberi waktu pemrosesan.
- Menampilkan pesan keberhasilan untuk setiap tahap *cleaning*.

Tabel 3. 3 *Cleaning Data*

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun	Jumlah Penjualan
1	1	1	1	1	1	1
1	2	1	1	1	1	1
1	3	1	1	1	1	1
1	4	1	1	1	1	2
1	5	1	1	1	1	2
...	...	...	...	...	...	...
1	1	1	1	5	8	2
1	2	1	1	5	8	2
1	3	1	1	5	8	2
1	4	1	1	5	8	1
1	5	1	1	5	8	1

Tabel 3. 3 proses mengubah data dari teks menjadi *numerik* pada kolom nama produk, rasa, jenis, kemasan, bulan, tahun dan jumlah penjualan dengan melalui proses *cleaning*.

4. Klasifikasi *naïve bayes*

Tahap awal dalam klasifikasi menggunakan algoritma *naïve bayes* dimulai dengan mengimpor pustaka yang diperlukan, yaitu *gaussiannb* dari *sklearn.naive.bayes*.

```
#import algoritma naive bayes
from sklearn.naive_bayes import GaussianNB

#import model gaussian
modelnb = GaussianNB()
```

Gambar 3. 7 Kode Klasifikasi *Naive Bayes*

Gambar 3. 7 kode ini mengimpor kelas *gaussiannb* dari modul *sklearn.naive_bayes*, yang merupakan implementasi algoritma *naïve bayes* dengan distribusi *gaussian*. Selanjutnya, membuat sebuah *instance* dari model *gaussiannb* dengan nama *modelnb*, yang artinya dapat digunakan untuk pelatihan dan prediksi data.

5. Memuat variabel X dan Y

Siapkan variabel x yaitu nama produk, rasa, jenis, kemasan, bulan dan tahun serta variabel y yaitu jumlah penjualan. Kemudian muat data variabel *independen* ke dalam variabel x . selanjutnya, muat data variabel *dependen*, yaitu jumlah penjualan yang dikategorikan sebagai laris atau kurang laris, ke dalam variabel y .

```
#memuat data variabel independen ke dalam variabel x
x = data.drop(["Jumlah Penjualan"], axis=1) # Ubah 'sumbu' menjadi 'sumbu'
print(x)
```

Gambar 3. 8 Kode Memuat Variabel x

Gambar 3. 8 kode tersebut memuat data variabel *independen* ke dalam variabel x dengan menghapus kolom “jumlah penjualan” dari *dataframe* data pada sumbu kolom (*axis=1*) dan kemudian mencetak hasilnya.

Tabel 3. 4 Variabel x

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun
1	1	1	1	1	1
1	2	1	1	1	1
1	3	1	1	1	1
1	4	1	1	1	1
1	5	1	1	1	1

Tabel 3.4 merupakan memuat data variabel *independen* ke dalam variabel x dengan menghapus jumlah penjualan.

```
# memuat data variabel dependen, yaitu variabel jumlah penjualan dengan laris atau kurang laris ke dalam variabel x
y = data["Jumlah Penjualan"]
print(y)
```

Gambar 3. 9 Kode Memuat Variabel *Dependen*

Gambar 3. 9 kode tersebut memuat variabel *dependen* “jumlah penjualan” dari *dataframe* data ke dalam variabel y , yang berisi kategori seperti laris atau kurang laris, lalu mencetak hasilnya.

Tabel 3. 5 Variabel y

Jumlah Penjualan
1
1
1
2
2

Tabel 3.5 memuat data variabel *independen* yaitu variabel jumlah penjualan dengan laris atau kurang laris ke dalam variabel x .

6. *Split data*

Setelah memuat data variabel x dan y , langkah selanjutnya adalah mengimpor *library machine learning scikit-learn*, yang akan digunakan untuk proses pelatihan dan pengujian model. Setelah itu, data x dan y dibagi menjadi dua bagian, yaitu data *training* untuk melatih model dan data *testing* untuk menguji akurasi model dalam melakukan prediksi. Dalam pembagian data ini, proses dilakukan secara acak sebanyak 80% data digunakan untuk pelatihan, sementara 20% data digunakan untuk pengujian.

```
#import library machine learning sklearn
from sklearn.model_selection import train_test_split
```

Gambar 3. 10 Kode *Library Sklearn*

Gambar 3. 10 kode tersebut mengimpor modul *train_test_split* dari *scikit-learn (sklearn)*, yang digunakan untuk membagi dataset menjadi data pelatihan (*training*) dan data pengujian (*testing*) dalam proses *meachine learning*.

```
#meload data x dan y ke dalam data training dan juga data test
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 123)
```

Gambar 3. 11 Kode *Split Data*

Gambar 3. 11 kode tersebut membagi dataset menjadi data pelatihan (x_{train} , y_{train}) dan data pengujian (x_{test} , y_{test}) menggunakan fungsi *train_test_split()*, dimana 20% data digunakan untuk pengujian ($test_size=0.2$), sementara 80%

digunakan pelatihan. parameter *random_state*=123 memastikan bahwa pembagian data selalu konsisten setiap kali kode dijalankan.

7. *Training* Data variabel *x* dan *y*

Data yang digunakan untuk melatih model terdiri dari variabel *x* dan *y*, yang dipilih secara acak. Sebanyak 80% dari total data, yaitu $0,8 * 1.620 = 1.296$ data, digunakan sebagai data training untuk proses pembelajaran model.

```
x_train.head(5)
```

Gambar 3. 12 Kode *x Train*

Gambar 3. 12 kode *x_train* berfungsi untuk menampilkan lima baris pertama dari *dataframe* *x_train*, yang merupakan bagian dari data pelatihan setelah pembagian dataset menggunakan *train_test_split()*.

Tabel 3. 6 *x Train*

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun
1	5	1	1	2	5
1	3	1	1	3	5
1	3	1	1	10	2
1	5	1	1	7	6
1	4	1	1	4	7

Tabel 3. 6 menampilkan lima data dari *x_train* setelah melalui proses pembagian data secara acak untuk melatih data.

```
y_train.head(5)
```

Gambar 3. 13 Kode *y Train*

Gambar 3. 13 kode *y_train* digunakan untuk menampilkan lima baris pertama dari *dataframe* atau *series* *y_train*, yang berisi data target (*label*) dari data pelatihan setelah pembagian dataset menggunakan *train_test_split()*.

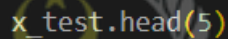
Tabel 3. 7 *y Train*

Jumlah Penjualan
1
2
1
2
1

Tabel 3. 7 menampilkan lima data dari *y_train* setelah melalui proses pembagian data secara acak untuk melatih data.

8. *Testing Data*

Data yang digunakan untuk pengujian dipilih secara acak sebanyak 20% dari total data. Sebanyak $0,2 * 1.620 = 324$ data digunakan sebagai data *testing* untuk mengevaluasi apakah model dapat memprediksi kategori laris atau kurang laris dengan akurat.



```
x_test.head(5)
```

Gambar 3. 14 Kode *x Test*

Gambar 3. 14 kode *x_test* digunakan untuk menampilkan lima baris pertama dari dataframe *x_test*, yang berisi data uji untuk variabel *independen* setelah proses pembagian dataset menggunakan *train_test_split()*.

Tabel 3. 8 *Testing*

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun
1	3	1	1	7	4
1	4	1	1	10	5
1	5	1	1	8	5
1	4	1	1	1	7
1	5	1	1	6	8

Tabel 3. 8 menampilkan lima baris data uji yang akan digunakan untuk menguji dan menghasilkan prediksi apakah suatu produk laris atau kurang laris.

9. *Output* prediksi dan probabilitas

Hasil prediksi pada *testing* berupa nilai probabilitas yang menunjukkan kemungkinan suatu data masuk ke dalam kategori laris atau kurang laris. pada proses pengujian ini, sebanyak 324 data digunakan untuk mengukur sejauh mana model dapat melakukan prediksi dengan akurat berdasarkan pola yang telah dipelajari dari data *training*.

```
#prediksi
y_pred = nbtrain.predict(x_test)
y_pred
```

Gambar 3. 15 Kode Prediksi

Gambar 3. 15 kode ini menggunakan model *naïve bayes* yang telah dilatih (*nbtrain*) untuk memprediksi nilai berdasarkan data uji(x_{test}) dan menyimpan hasilnya dalam variabel *y_pred*. setelah itu, *y_pred* akan menampilkan hasil prediksi tersebut.

[illegible]

Gambar 3. 16 Hasil Prediksi

Hasil prediksi ditampilkan dalam bentuk *array* yang berisi kode 1 (laris) dan kode 2 (kurang laris) dengan total sebanyak 324 data.

```
#probabilitas dari data test
probabilitas
```

Gambar 3. 17 Kode Probabilitas

Gambar 3. 17 kode ini bertujuan untuk menampilkan probabilitas hasil prediksi dari data uji. Probabilitas ini menunjukkan seberapa besar kemungkinan setiap sampel dalam data uji termasuk ke dalam masing – masing kelas.

a. Akurasi

$$akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

b. Presisi

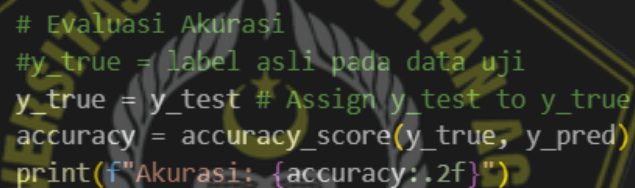
$$precision = \frac{TP}{TP + FP} \quad (3)$$

c. Recall

$$Recall = \frac{TN}{TN + FP} \quad (4)$$

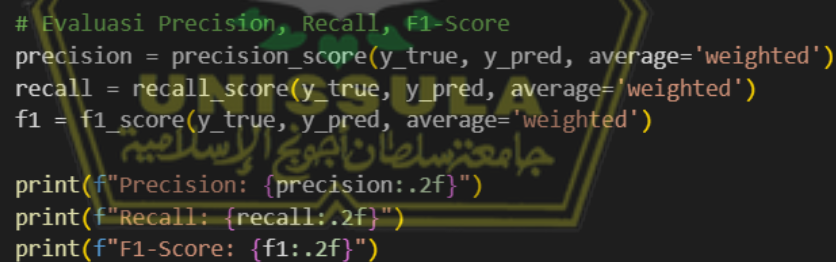
d. F1-score

$$f1 - score = 2 * \frac{precision * recall}{precision + recall} \quad (5)$$



```
# Evaluasi Akurasi
#y_true = label asli pada data uji
y_true = y_test # Assign y_test to y_true
accuracy = accuracy_score(y_true, y_pred)
print(f"Akurasi: {accuracy:.2f}")
```

Gambar 3. 19 Kode Evaluasi Akurasi



```
# Evaluasi Precision, Recall, F1-Score
precision = precision_score(y_true, y_pred, average='weighted')
recall = recall_score(y_true, y_pred, average='weighted')
f1 = f1_score(y_true, y_pred, average='weighted')

print(f"Precision: {precision:.2f}")
print(f"Recall: {recall:.2f}")
print(f"F1-Score: {f1:.2f}")
```

Gambar 3. 20 Kode Evaluasi

Gambar 3. 20 Berdasarkan hasil evaluasi, model menghasilkan akurasi sebesar 0.59 dengan *precision* 0.59, *recall* 0.59 dan *f1-score* 0.58, yang menunjukkan bahwa hasil prediksi sesuai dengan label asli yang terdapat pada data uji. Nilai *precision*, *recall*, *f1-score* dengan metode *average* yang mempertimbangkan distribusi kelas dalam data uji untuk memberikan evaluasi yang lebih seimbang terhadap performa model.

```
# Laporan Klasifikasi
print("\nLaporan Klasifikasi:\n")
print(classification_report(y_true, y_pred))
```

Gambar 3. 21 Kode Laporan Klasifikasi

Gambar 3. 21 kode tersebut menampilkan laporan klasifikasi yang berisi metrik evaluasi seperti presisi, *recall*, dan *f1-score* dengan menggunakan fungsi *classification_report(y_true, y_pred)* dari pustaka *sklearn*.

Laporan Klasifikasi:

	precision	recall	f1-score	support
1	0.62	0.47	0.53	163
2	0.57	0.71	0.63	161
accuracy			0.59	324
macro avg	0.59	0.59	0.58	324
weighted avg	0.59	0.59	0.58	324

Gambar 3. 22 Hasil Klasifikasi Laporan

Gambar 3. 22 hasil klasifikasi menunjukkan bahwa model memiliki akurasi sebesar 59% dengan nilai *precision*, *recall* dan *f1-score* untuk masing – masing kelas.

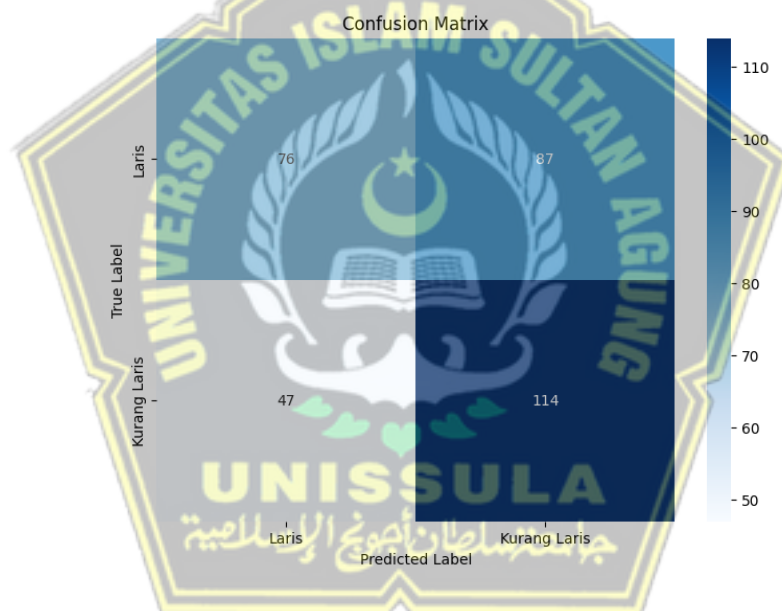
- Kelas 1 (laris) memiliki *precision* 62%, *recall* 47% dan *f1-score* 53% menunjukkan bahwa model lebih sering salah dalam mengidentifikasi kelas ini.
- Kelas 2 (kurang laris) memiliki *precision* 57%, *recall* 71%, dan *f1-score* 63%, yang berarti model lebih baik dalam mendeteksi kelas ini dibanding kelas 1.
- *Macro avg* dan *weighted avg* masing – masing bernilai sekitar 59%, yang menunjukkan performa rata – rata model terhadap seluruh kelas.

Secara keseluruhan model cenderung lebih ke akurat dalam mengidentifikasi produk yang kurang laris dibandingkan dengan laris.

```
# Visualisasi Confusion Matrix
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=['Laris', 'Kurang Laris'], yticklabels=['Laris', 'Kurang Laris'])
plt.xlabel('Predicted Label')
plt.ylabel('True Label')
plt.title('Confusion Matrix')
plt.show()
```

Gambar 3. 23 Kode *Confusion Matrix*

Gambar 3. 23 kode ini digunakan untuk membuat visualisasi *confusion matrix* menggunakan *matplotlib* dan *seaborn* dengan tampilan *heatmap* berwarna biru. Label pada sumbu *x* menunjukkan hasil prediksi model, sedangkan label pada sumbu *y* menunjukkan nilai sebenarnya, sehingga memudahkan dalam mengevaluasi performa klasifikasi.

Gambar 3. 24 *Confusion Matrix*

Gambar 3. 24 menunjukkan *confusion matrix*, yang menggambarkan performa model klasifikasi dalam membedakan kategori laris dan kurang laris.

- 76 produk yang sebenarnya laris diprediksi dengan benar sebagai laris (*true positive*)
- 87 produk yang sebenarnya laris salah diprediksi sebagai kurang laris (*false negative*)
- 47 produk yang sebenarnya kurang laris salah diprediksi sebagai laris (*false positive*)

- 144 produk yang sebenarnya kurang laris diprediksi benar sebagai kurang laris (*true negative*)

Hasil ini menunjukkan bahwa model lebih sering salah memprediksi produk laris sebagai kurang laris yang dapat berdampak pada keputusan bisnis terkait stok dan strategi penjualan.

3.2 Perancangan *User Interface*

Dalam pembuatan aplikasi prediksi penjualan diperlukan perancangan antarmuka pengguna (*user interface*) yang dirancang dengan baik untuk memudahkan pengguna dalam mengoperasikan aplikasi. Berikut ini adalah tampilan rancangan antarmuka pengguna yang telah disiapkan.

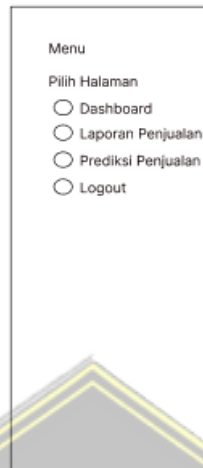
3.2.1 *Login*



Gambar 3. 25 *Login*

Pada gambar 3. 25 merupakan tampilan *login* yang dimana halaman ini hanya bisa dimasukkan oleh *admin* dan *user* karena jika ingin masuk ke halaman ini harus memiliki *username* dan *password*. Untuk masuk ke login aplikasi harus memasukkan *username* dan *password* terlebih dahulu lalu klik tombol login agar bisa masuk ke dalam aplikasi.

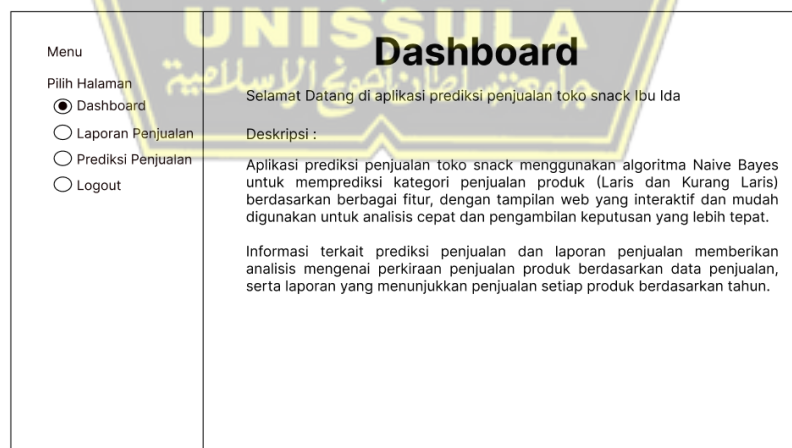
3.2.2 Menu



Gambar 3. 26 Menu

Pada gambar 3. 26 menunjukkan *menu* navigasi pada sebuah aplikasi yang memungkinkan pengguna untuk memilih halaman yang ingin diakses. Menu ini berisi beberapa opsi berupa tombol yaitu *dashboard*, *upload data*, *laporan penjualan*, *prediksi penjualan* dan *logout*.

3.2.3 Dashboard



Gambar 3. 27 Dashboard

Pada gambar 3. 27 halaman *dashboard*, ditampilkan deskripsi mengenai penjelasan aplikasi prediksi penjualan menggunakan algoritma *naïve bayes*.

Selanjutnya, halaman ini menyampaikan informasi terkait langkah – langkah prediksi penjualan dan pembuatan laporan penjualan.

3.2.4 Laporan Penjualan

Gambar 3. 28 Laporan Penjualan

Pada gambar 3. 28 sistem ini, pengguna dapat mengunggah file *excel* yang berisi data penjualan, kemudian memilih tahun tertentu, misalnya tahun 2024 untuk menampilkan data penjualan pada tahun tersebut. setelah data ditampilkan sistem akan menampilkan laporan penjualan berdasarkan informasi yang telah dipilih. Laporan tersebut dapat diunduh dalam format *file word* untuk melihat laporan dan dokumentasi lebih lanjut.

3.2.5 Prediksi Penjualan

1. Input data penjualan

Menu
Pilih Halaman
☐ Dashboard
☐ Laporan Penjualan
☒ Prediksi Penjualan
☐ Logout

Prediksi Penjualan Toko Snack

upload data untuk prediksi

Upload File
Browse Files

data yang diunggah

No	Nama Produk	Rasa	Isi	Kemasan	Uk	Jumlah Produk	Harga Produk	Bulan	Tahun	Jumlah Penjualan
1	Ades	Coklat	Batu	Isi	60 Dzs	5000	Rp. 110.000	Mar	2018	1471
2	Ades	Strawberry	Batu	Isi	60 Dzs	5000	Rp. 110.000	Mar	2018	1471
3	Ades	Pisang	Batu	Isi	60 Dzs	5000	Rp. 110.000	Mar	2018	1471
4	Ades	Durian	Batu	Isi	60 Dzs	5000	Rp. 110.000	Mar	2018	Kurang Laji
5	Ades	Nanas	Batu	Isi	60 Dzs	5000	Rp. 110.000	Mar	2018	Kurang Laji
6	Chikapan	Strawberry	Batu	Isi	60 Dzs	5000	Rp. 110.000	Mar	2018	1471
7	Chikapan	Coklat	Batu	Isi	60 Dzs	5000	Rp. 110.000	Mar	2018	1471
8	Chikapan	Nanas	Batu	Isi	60 Dzs	5000	Rp. 110.000	Mar	2018	Kurang Laji
9	Kakao	Coklat	Batu	Isi	60 Dzs	5000	Rp. 110.000	Mar	2018	1471
10	Kakao	Pisang	Batu	Isi	60 Dzs	5000	Rp. 110.000	Mar	2018	Kurang Laji
11	Kakao	Strawberry	Batu	Isi	60 Dzs	5000	Rp. 110.000	Mar	2018	1471
12	Skotlandia	Coklat	Batu	Isi	60 Dzs	5000	Rp. 110.000	Mar	2018	1471

Gambar 3. 29 Input Data Penjualan

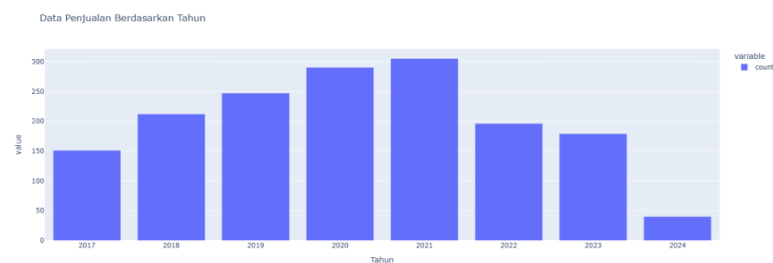
Gambar 3. 29 tahap perencanaan dalam melakukan prediksi penjualan adalah dengan mengunggah file data penjualan dalam format excel. Setelah file berhasil diunggah sistem akan menampilkan data yang telah diunggah untuk memastikan informasi yang dimasukkan sesuai dan siap digunakan dalam proses analisis lebih lanjut.

2. Grafik Data Penjualan

Grafik Data Penjualan:

pilih kolom untuk menampilkan grafik:

Tahun



Gambar 3. 30 Grafik Penjualan

Gambar 3. 30 grafik data penjualan dapat ditampilkan dengan memilih parameter berdasarkan nama produk, rasa, jenis, kemasan, bulan dan tahun. Pilihan ini memungkinkan pengguna untuk memvisualisasikan data penjualan secara spesifik sesuai dengan kategori yang diinginkan.

3. Menampilkan atribut dari data

☐ Tampilkan Atribut Data

Atribut Data:

No
Nama Produk
Rasa
Jenis
Kemasan
Isi
Jumlah Produk
Harga Produk
Bulan
Tahun
Jumlah Penjualan

Gambar 3. 31 Atribut Data

Gambar 3. 31 tampilkan atribut data yang terdiri dari : nomor, nama produk, rasa, jenis, kemasan, isi, jumlah produk, harga produk, bulan, tahun dan jumlah penjualan. Atribut – atribut ini memberikan informasi lengkap untuk mendukung analisis dan prediksi penjualan.

4. Menghapus kolom tidak diperlukan

hapus kolom tidak diperlukan :

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun	Jumlah Penjualan
Aoka	Cokelat	Roti	Sachet	Mei	2024	Laris
Aoka	Strawberry	Roti	Sachet	Mei	2024	Laris
Aoka	Pandan	Roti	Sachet	Mei	2024	Laris
Aoka	Duren	Roti	Sachet	Mei	2024	Kurang Laris
Aoka	Nanas	Roti	Sachet	Mei	2024	Kurang Laris
Oishipan	Strawberry	Roti	Sachet	Mei	2024	Laris
Oishipan	Cokelat	Roti	Sachet	Mei	2024	Laris
Oishipan	Nanas	Roti	Sachet	Mei	2024	Kurang Laris
Nabati	Cokelat	Wafer	Sachet	Mei	2024	Laris
Nabati	Keju	Wafer	Sachet	Mei	2024	Kurang Laris

Gambar 3. 32 Menghapus Kolom

Gambar 3. 32 lakukan penghapusan kolom yang tidak diperlukan seperti nomor, isi, jumlah produk dan harga produk untuk menyederhanakan data dan memfokuskan analisis pada atribut yang relevan.

5. Cleaning data

Cleaning Data

Setelah melakukan cleaning data :

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun	Jumlah Penjualan
1	1	1	1	1	1	1
1	2	1	1	1	1	1
1	3	1	1	1	1	1
1	4	1	1	1	1	2
1	5	1	1	1	1	2
2	2	1	1	1	1	1
2	1	1	1	1	1	1
2	5	1	1	1	1	2
3	1	2	1	1	1	1
3	6	2	1	1	1	2

Gambar 3. 33 Cleaning Data

Gambar 3. 33 Cleaning data dilakukan setelah proses penghapusan kolom untuk memastikan data yang tersisa bersih dan konsisten. Setelah itu mengubah teks menjadi kode atau numerik untuk mempermudah proses analisis data dan memastikan keseimbangan dengan algoritma naïve bayes.

6. Tahapan Algoritma naïve bayes

Tahapan algoritma naïve bayes meliputi penghitungan probabilitas setiap fitur terhadap kelas menggunakan teorema bayes, mengasumsikan independen antar fitur kemudian menentukan kelas dengan probabilitas tertinggi berdasarkan data latih.

1. Mempersiapkan data pada variabel x dan y

Pada tahap awal klasifikasi menggunakan algoritma naïve bayes, data fitur dimuat ke dalam variabel x, yang berisi atribut – atribut seperti nama produk, rasa, jenis, kemasan, bulan dan tahun sedangkan data label dimasukkan ke dalam variabel y yang mempresentasikan kategori laris dan kurang laris.

Tahapan Klasifikasi Algoritma Naive Bayes

Jumlah data pada variabel X: 1620 (6 kolom)

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun
1	1	1	1	1	1
1	2	1	1	1	1
1	3	1	1	1	1
1	4	1	1	1	1
1	5	1	1	1	1
2	2	1	1	1	1
2	1	1	1	1	1
2	5	1	1	1	1
3	1	2	1	1	1
3	6	2	1	1	1

Gambar 3. 34 Tahap Klasifikasi *Naïve Bayes*

Jumlah data pada variabel y: 1620 (1 kolom)

Jumlah Penjualan
1
1
1
2
2
1
1
2
1
2

Gambar 3. 35 Variabel y

2. Pembagian data training dan testing

Data yang telah dimuat kemudian dibagi menjadi dua bagian, yaitu data training untuk melatih model dan data testing untuk mengevaluasi performa model naïve bayes dalam memprediksi kategori.

- **Data training variabel x dan y**

Data training yang tersimpan dalam variabel `x_train` dan `y_train` merupakan subset dari dataset asli yang digunakan untuk melatih model. Variabel `x_train` berisi atribut – atribut atau fitur, sedangkan variabel `y_train` berisi label atau kategori target yang sesuai dengan data tersebut.

Pembagian Data Training dan Testing

Jumlah data untuk training: 1296 (80%)

Jumlah data untuk testing: 324 (20%)

Data Training Variabel x (1296 data, 6 kolom):

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun
16	11	6	1	11	5
6	1	3	1	12	5
13	14	8	1	11	2
3	6	2	1	2	5
14	2	8	1	11	7
7	2	4	2	12	5
3	6	2	1	9	4
7	1	4	2	2	5
3	6	2	1	4	8
4	1	2	1	11	7

Gambar 3. 36 Data Training Variabel x

Data Training Variabel y : 1296

Jumlah Penjualan
2
2
2
1
2
1
1
2
1
2

Gambar 3. 37 Data Training Variabel y

- Data testing**

Setelah proses pelatihan, model diuji menggunakan data testing yang tersimpan dalam variabel `x_test` dan `y_test`. Variabel `x_test` digunakan untuk melatih model sedangkan `y_test` menyimpan digunakan untuk menguji dalam memprediksi kategori berdasarkan target yang sesuai dengan data.

Data Testing (324 data, 6 kolom):

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun
6	1	3	1	8	4
11	4	7	1	7	5
16	11	6	1	5	4
6	9	3	1	8	7
12	16	7	1	1	7
6	1	3	1	11	6
1	4	1	1	3	5
1	4	1	1	2	5
14	2	8	1	12	5
11	1	7	1	3	3

Gambar 3. 38 Data *Testing*

3. Hasil prediksi penjualan

Hasil prediksi menunjukkan kategori target yang diperkirakan oleh model dengan menampilkan data yang diprediksi setelah pengujian.

Hasil Prediksi Penjualan Tahun 2025

Data Penjualan (324 data, 8 kolom):

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun	Prediksi Jumlah Penjualan	Kategori Prediksi
6	1	3	1	8	4	2	Kurang Laris
11	4	7	1	7	5	2	Kurang Laris
16	11	6	1	5	4	1	Laris
6	9	3	1	8	7	2	Kurang Laris
12	16	7	1	1	7	1	Laris
6	1	3	1	11	6	2	Kurang Laris
1	4	1	1	3	5	2	Kurang Laris
1	4	1	1	2	5	2	Kurang Laris
14	2	8	1	12	5	2	Kurang Laris
11	1	7	1	3	3	2	Kurang Laris

Gambar 3. 39 Hasil Prediksi Penjualan

4. Probabilitas prediksi

Setelah melakukan prediksi selanjutnya menampilkan probabilitas yang digunakan untuk membandingkan probabilitas “laris” dan “kurang laris”. Dalam hal ini, hasil prediksi ditentukan dengan probabilitas tertinggi.

Probabilitas Prediksi
Data Testing (324 data, 2 kolom):

Probabilitas Laris	Probabilitas Kurang Laris
0.3518016903203484	0.6481983096796511
0.3709624878626937	0.6290375121373071
0.5135828254070368	0.48641717459296374
0.34601440995703076	0.6539855900429696
0.543069176771808	0.45693082322819223
0.29309795406806444	0.7069020459319358
0.33520392007181704	0.6647960799281838
0.33673431834803896	0.6632656816519613
0.3798485798360926	0.6201514201639068
0.4298527299167448	0.5701472700832548

Gambar 3. 40 Probabilitas Prediksi

5. Evaluasi model

Evaluasi model dilakukan setelah proses prediksi dengan membandingkan hasil prediksi dengan data aktual menggunakan metrik seperti akurasi, presisi, recall, dan f1-score untuk mengukur seberapa baik model mampu memprediksi kategori target secara tepat.

Hasil Evaluasi Model

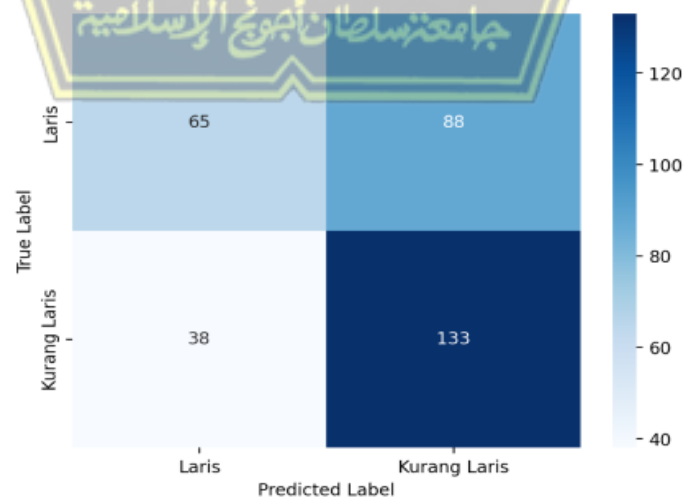
Akurasi: 0.61

Precision: 0.62

Recall: 0.61

F1-Score: 0.60

Confusion Matrix:



Gambar 3. 41 Evaluasi Model

3.2.6 Logout



Gambar 3. 42 Logout

Pada gambar 3. 42 menunjukkan halaman logout dari sebuah aplikasi prediksi penjualan. Dibagian kiri terdapat menu navigasi yang memudahkan pengguna untuk berpindah antarhalaman dalam aplikasi. Halaman ini dirancang untuk digunakan setelah pengguna selesai melakukan prediksi penjualan memberi opsi keluar dari aplikasi dengan aman.

BAB IV

HASIL DAN ANALISIS PENELITIAN

4.1 Analisis Penelitian

4.1.1 Mempersiapkan data

Langkah pertama sebelum melakukan tahap klasifikasi naïve bayes adalah mempersiapkan data untuk dianalisis. Pada tahap ini, data yang digunakan berupa nilai numerik atau kode setelah melalui proses pembersihan data. Berikut ini adalah tabel yang digunakan.

Tabel 4. 1 Memuat Data

Nama Produk	Rasa	Jenis	kemasan	Bulan	Tahun	Jumlah Penjualan
1	1	1	1	1	1	1
1	2	1	1	1	1	1
1	3	1	1	1	1	1
1	4	1	1	1	1	2
1	5	1	1	1	1	2
1	1	1	1	1	1	1
1	2	1	1	1	1	1
1	3	1	1	1	1	1
1	4	1	1	1	1	2
1	5	1	1	1	1	2

Tabel 4.1 berisi data yang digunakan sebagai dasar dalam proses analisis klasifikasi menggunakan metode Naive Bayes, yang bertujuan untuk mengidentifikasi pola dan menentukan kategori berdasarkan probabilitas masing-masing fitur.

4.1.2 Pembagian data *training* dan data *testing*

Data dibagi menjadi dua bagian, yaitu data pelatihan (*training*) dan data pengujian (*testing*). Sebanyak 80% dari data yang digunakan sebagai data latih, sementara 20% digunakan sebagai data pengujian. Pembagian data tersebut dilakukan sebagai berikut :

- *Training* data (80%) = $0.8 * 1620 = 1296$
- *Testing* data (20%) = $0.2 * 1620 = 324$

Pembagian ini dilakukan secara seimbang untuk setiap kategori, yaitu “laris” dan “kurang laris”. pemilihan data pelatihan dilakukan secara acak dari total 1.620 data, sehingga diperoleh 1.296 data untuk dilatih. Sementara itu, data pengujian dipilih dari kategori jumlah terbanyak dalam 324 data, dimana yang paling sering muncul digunakan sebagai pengujian.

a. Membaca *Training* data

Tabel 4. 2 *Training* Data

Nama produk	Rasa	Jenis	Kemasan	Bulan	Tahun	Jumlah Penjualan
1	5	1	1	2	5	1
1	3	1	1	3	5	2
1	3	1	1	10	2	1
1	5	1	1	7	6	2
1	4	1	1	4	7	1
1	1	1	1	4	5	2
1	5	1	1	9	4	1
1	5	1	1	7	6	2
1	5	1	1	5	8	1
1	2	1	1	5	7	2

Tabel 4.2 menampilkan data pelatihan yang terdiri dari 1.296 sampel yang dipilih secara acak dengan tujuan melatih data training.

b. *Testing data*Tabel 4. 3 *Testing Data*

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun	Jumlah Penjualan
1	3	1	1	7	4	?
1	4	1	1	10	5	?
1	5	1	1	8	5	?
1	4	1	1	1	7	?
1	5	1	1	6	8	?

Tabel 4.3 menampilkan data uji yang terdiri dari 324 sampel yang dipilih secara acak untuk menentukan hasil prediksi apakah suatu produk laris atau kurang laris.

1. Menghitung Probabilitas kelas

Tabel 4. 4 Menghitung Probabilitas Kelas

Jumlah penjualan	Jumlah kemunculan	Probabilitas prior
Laris	633	$633/1.296 = 0,488425926$
Kurang laris	663	$663/1.296 = 0,511574074$

Tabel 4.4 menghitung probabilitas kelas berdasarkan jumlah kemunculan masing-masing kategori dalam data pelatihan, di mana kategori “Laris” muncul sebanyak 633 kali dengan probabilitas 0,4884, sedangkan kategori “Kurang Laris” muncul sebanyak 663 kali dengan probabilitas 0,5115 dari total 1.296 data pelatihan.

2. Menghitung probabilitas atribut

a. Probabilitas kategori rasa

Tabel 4. 5 Menghitung Probabilitas Kategori Rasa

Rasa	Laris	Kurang laris	Probabilitas laris	Probabilitas kurang laris
3	118	138	$118/633 = 0,137614679$	$138/663 = 0,269470405$
4	145	119	$145/633 = 0,068807339$	$119/663 = 0,035825545$
4	138	114	$138/633 = 0,060747664$	$114/663 = 0,033639144$
5	145	119	$145/633 = 0,017133956$	$119/663 = 0,032110092$
5	138	114	$138/633 = 0,020249221$	$114/663 = 0,027522931$

b. Probabilitas kategori jenis

Tabel 4. 6 Menghitung Probabilitas Kategori Jenis

Jenis	Laris	Kurang laris	Probabilitas laris	Probabilitas kurang laris
1	633	663	$633/633 = 1$	$663/663 = 1$
1	633	663	$633/633 = 1$	$663/663 = 1$
1	633	663	$633/633 = 1$	$663/663 = 1$
1	633	663	$633/633 = 1$	$663/663 = 1$
1	633	663	$633/633 = 1$	$663/663 = 1$

c. Probabilitas kategori kemasan

Tabel 4. 7 Menghitung Probabilitas Kategori Kemasan

Jenis	Laris	Kurang laris	Probabilitas laris	Probabilitas kurang laris

1	633	663	$633/633 = 1$	$663/663 = 1$
1	633	663	$633/633 = 1$	$663/663 = 1$
1	633	663	$633/633 = 1$	$663/663 = 1$
1	633	663	$633/633 = 1$	$663/663 = 1$
1	633	663	$633/633 = 1$	$663/663 = 1$

d. Probabilitas bulan

Tabel 4. 8 Menghitung Probabilitas Bulan

Bulan	Laris	Kurang laris	Probabilitas laris	Probabilitas kurang laris
7	43	54	$43/633 = 0,06793049$	$54/663 = 0,081447964$
10	50	58	$50/633 = 0,078988942$	$58/663 = 0,087481146$
1	53	55	$53/633 = 0,08372878$	$55/663 = 0,082956259$
8	55	53	$55/633 = 0,086887836$	$53/663 = 0,079939668$
6	46	63	$46/633 = 0,072669826$	$63/663 = 0,095022624$

e. Probabilitas Tahun

Tabel 4. 9 Menghitung Probabilitas Tahun

Tahun	Laris	Kurang laris	Probabilitas laris	Probabilitas kurang laris
4	100	95	$100/633 = 0,157977883$	$95/663 = 0,143288084$
5	78	119	$78/633 = 0,123222749$	$119/663 = 0,1794487179$
7	78	119	$78/633 = 0,123222749$	$119/663 = 0,1794487179$
5	83	107	$83/633 = 0,131121643$	$107/663 = 0,161387632$
8	86	106	$86/633 = 0,13586079$	$106/663 = 0,159879336$

Berikut adalah perhitungan pada setiap produk :

1. Menghitung probabilitas produk pada kode 1

a) Menghitung probabilitas prior $P(H)$

$$P(\text{Laris}) = 633/1.296 = 0,488425926$$

$$P(\text{Kurang Laris}) = 663/1.296 = 0,511574074$$

b) Menghitung probabilitas likelihood $P(X|H)$

$$P(\text{Rasa}) \text{ Kode 3} \mid \text{"Laris"} \text{ Kode 1} = 118/633 = 0,137614679$$

$$P(\text{Rasa}) \text{ Kode 3} \mid \text{"Kurang Laris"} \text{ Kode 2} = 138/663 = 0,269470405$$

$$P(\text{Jenis}) \text{ Kode 1} \mid \text{"Laris"} \text{ Kode 1} = 633/633 = 1$$

$$P(\text{Jenis}) \text{ Kode 1} \mid \text{"Kurang Laris"} \text{ Kode 2} = 663/663 = 1$$

$$P(\text{Kemasan}) \text{ Kode 1} \mid \text{"Laris"} \text{ Kode 1} = 633/633 = 1$$

$$P(\text{Kemasan}) \text{ Kode 1} \mid \text{"Kurang Laris"} \text{ Kode 2} = 663/663 = 1$$

$$P(\text{Bulan}) \text{ Kode 7} \mid \text{"Laris"} \text{ Kode 1} = 43/633 = 0,06793049$$

$$P(\text{Bulan}) \text{ Kode 7} \mid \text{"Kurang Laris"} \text{ Kode 2} = 54/663 = 0,081447964$$

$$P(\text{Tahun}) \text{ Kode 4} \mid \text{"Laris"} \text{ Kode 1} = 100/633 = 0,157977883$$

$$P(\text{Tahun}) \text{ Kode 4} \mid \text{"Kurang Laris"} \text{ Kode 2} = 95/663 = 0,143288084$$

c) Menghitung probabilitas posterior $P(H|X)$

Probabilitas laris

$$= P(\text{Rasa}) * P(\text{Jenis}) * P(\text{Kemasan}) * P(\text{Bulan}) * P(\text{Tahun})$$

$$= 0,137614679 * 1 * 1 * 0,06793049 * 0,157977883 = 0,002000504$$

Probabilitas kurang laris

$$= P(\text{Rasa}) * P(\text{Jenis}) * P(\text{Kemasan}) * P(\text{Bulan}) * P(\text{Tahun})$$

$$= 0,269470405 * 1 * 1 * 0,081447964 * 0,143288084 = 0,002429159$$

d) Menentukan probabilitas tertinggi (keputusan akhir)

Karena probabilitas $P(\text{Kurang Laris} \mid X) = 0,002000504$ lebih besar daripada probabilitas $P(\text{Laris} \mid X) = 0,002429159$, maka kelas yang diprediksi adalah "Kurang Laris".

2. Menghitung probabilitas produk pada kode 1

a) Menghitung probabilitas prior $P(H)$

$$P(\text{Laris}) = 633/1.296 = 0,488425926$$

$$P(\text{Kurang Laris}) = 663/1.296 = 0,511574074$$

b) Menghitung probabilitas likelihood $P(X|H)$

$$P(\text{Rasa}) \text{ Kode 4} \mid \text{"Laris"} \text{ Kode 1} = 145/633 = 0,068807339$$

$$P(\text{Rasa}) \text{ Kode 4} \mid \text{"Kurang Laris"} \text{ Kode 2} = 119/663 = 0,035825545$$

$$P(\text{Jenis}) \text{ Kode 1} \mid \text{"Laris"} \text{ Kode 1} = 633/633 = 1$$

$$P(\text{Jenis}) \text{ Kode 1} \mid \text{"Kurang Laris"} \text{ Kode 2} = 663/663 = 1$$

$$P(\text{Kemasan}) \text{ Kode 1} \mid \text{"Laris"} \text{ Kode 1} = 633/633 = 1$$

$$P(\text{Kemasan}) \text{ Kode 1} \mid \text{"Kurang Laris"} \text{ Kode 2} = 663/663 = 1$$

$$P(\text{Bulan}) \text{ Kode 10} \mid \text{"Laris"} \text{ Kode 1} = 50/633 = 0,078988942$$

$$P(\text{Bulan}) \text{ Kode 10} \mid \text{"Kurang Laris"} \text{ Kode 2} = 58/663 = 0,087481146$$

$$P(\text{Tahun}) \text{ Kode 5} \mid \text{"Laris"} \text{ Kode 1} = 78/633 = 0,123222749$$

$$P(\text{Tahun}) \text{ Kode 5} \mid \text{"Kurang Laris"} \text{ Kode 2} = 119/663 = 0,1794487179$$

c) Menghitung probabilitas posterior $P(H|X)$

Probabilitas laris

$$= P(\text{Rasa}) * P(\text{Jenis}) * P(\text{Kemasan}) * P(\text{Bulan}) * P(\text{Tahun})$$

$$= 0,068807339 * 1 * 1 * 0,078988942 * 0,123222749 = 0,002229572$$

Probabilitas kurang laris

$$= P(\text{Rasa}) * P(\text{Jenis}) * P(\text{Kemasan}) * P(\text{Bulan}) * P(\text{Tahun})$$

$$= 0,035825545 * 1 * 1 * 0,087481146 * 0,1794487179 = 0,002818262$$

d) Menentukan probabilitas tertinggi (keputusan akhir)

Karena probabilitas $P(\text{Kurang Laris} \mid X) = 0,002818262$ lebih besar daripada probabilitas $P(\text{Laris} \mid X) = 0,002229572$, maka kelas yang diprediksi adalah "Kurang Laris".

3. Menghitung probabilitas produk pada kode 1

a) Menghitung probabilitas prior $P(H)$

$$P(\text{Laris}) = 633/1.296 = 0,488425926$$

$$P(\text{Kurang Laris}) = 663/1.296 = 0,511574074$$

b) Menghitung probabilitas likelihood $P(X|H)$

$$P(\text{Rasa}) \text{ Kode 5} \mid \text{"Laris"} \text{ Kode 1} = 145/633 = 0,017133956$$

$$P(\text{Rasa}) \text{ Kode 5} \mid \text{"Kurang Laris"} \text{ Kode 2} = 119/663 = 0,032110092$$

$$P(\text{Jenis}) \text{ Kode 1} \mid \text{"Laris"} \text{ Kode 1} = 633/633 = 1$$

$$P(\text{Jenis}) \text{ Kode 1} \mid \text{"Kurang Laris"} \text{ Kode 2} = 663/663 = 1$$

$$P(\text{Kemasan}) \text{ Kode 1} \mid \text{"Laris"} \text{ Kode 1} = 633/633 = 1$$

$$P(\text{Kemasan}) \text{ Kode 1} \mid \text{"Kurang Laris"} \text{ Kode 2} = 633/633 = 1$$

$$P(\text{Bulan}) \text{ Kode 8} \mid \text{"Laris"} \text{ Kode 1} = 55/633 = 0,086887836$$

$$P(\text{Bulan}) \text{ Kode 8} \mid \text{"Kurang Laris"} \text{ Kode 2} = 53/663 = 0,079939668$$

$$P(\text{Tahun}) \text{ Kode 5} \mid \text{"Laris"} \text{ Kode 1} = 83/633 = 0,131121643$$

$$P(\text{Tahun}) \text{ Kode 5} \mid \text{"Kurang Laris"} \text{ Kode 2} = 107/663 = 0,161387632$$

c) Menghitung probabilitas posterior $P(H|X)$

Probabilitas laris

$$= P(\text{Rasa}) * P(\text{Jenis}) * P(\text{Kemasan}) * P(\text{Bulan}) * P(\text{Tahun})$$

$$= 0,137614679 * 1 * 1 * 0,086887836 * 0,079939668 = 0,002609742$$

Probabilitas kurang laris

$$= P(\text{Rasa}) * P(\text{Jenis}) * P(\text{Kemasan}) * P(\text{Bulan}) * P(\text{Tahun})$$

$$= 0,032110092 * 1 * 1 * 0,079939668 * 0,161387632 = 0,002315613$$

d) Menentukan probabilitas tertinggi (keputusan akhir)

Karena probabilitas $P(\text{Laris} \mid X) = 0,002609742$ lebih besar daripada probabilitas $P(\text{Kurang Laris} \mid X) = 0,0023155613$, maka kelas yang diprediksi adalah "Laris".

4. Menghitung probabilitas produk pada kode 1

a) Menghitung probabilitas prior $P(H)$

$$P(\text{Laris}) = 633/1.296 = 0,488425926$$

$$P(\text{Kurang Laris}) = 663/1.296 = 0,511574074$$

b) Menghitung probabilitas likelihood $P(X|H)$

$$P(\text{Rasa}) \text{ Kode 4} \mid \text{"Laris"} \text{ Kode 1} = 138/633 = 0,060747664$$

$$P(\text{Rasa}) \text{ Kode 4} \mid \text{"Kurang Laris"} \text{ Kode 2} = 114/663 = 0,033639144$$

$$P(\text{Jenis}) \text{ Kode 1} \mid \text{"Laris"} \text{ Kode 1} = 633/633 = 1$$

$$P(\text{Jenis}) \text{ Kode 1} \mid \text{"Kurang Laris"} \text{ Kode 2} = 663/663 = 1$$

$$P(\text{Kemasan}) \text{ Kode 1} \mid \text{"Laris"} \text{ Kode 1} = 633/633 = 1$$

$$P(\text{Kemasan}) \text{ Kode 1} \mid \text{"Kurang Laris"} \text{ Kode 2} = 663/663 = 1$$

$$P(\text{Bulan}) \text{ Kode 1} \mid \text{"Laris"} \text{ Kode 1} = 53/633 = 0,08372878$$

$$P(\text{Bulan}) \text{ Kode 1} \mid \text{"Kurang Laris"} \text{ Kode 2} = 55/663 = 0,082956259$$

$$P(\text{Tahun}) \text{ Kode 7} \mid \text{"Laris"} \text{ Kode 1} = 78/633 = 0,123222749$$

$$P(\text{Tahun}) \text{ Kode 7} \mid \text{"Kurang Laris"} \text{ Kode 2} = 119/663 = 0,061387632$$

c) Menghitung probabilitas posterior $P(H|X)$

Probabilitas laris

$$\begin{aligned} &= P(\text{Rasa}) * P(\text{Jenis}) * P(\text{Kemasan}) * P(\text{Bulan}) * P(\text{Tahun}) \\ &= 0,060747664 * 1 * 1 * 0,08372878 * 0,123222749 = 0,002249254 \end{aligned}$$

Probabilitas kurang laris

$$\begin{aligned} &= P(\text{Rasa}) * P(\text{Jenis}) * P(\text{Kemasan}) * P(\text{Bulan}) * P(\text{Tahun}) \\ &= 0,033639144 * 1 * 1 * 0,082956259 * 0,061387632 = 0,0025602 \end{aligned}$$

d) Menentukan probabilitas tertinggi (keputusan akhir)

Karena probabilitas $P(\text{Kurang Laris} \mid X) = 0,0025602$ lebih besar daripada probabilitas $P(\text{Laris} \mid X) 0,002249254$ maka kelas yang diprediksi adalah "Kurang Laris".

5. Menghitung probabilitas produk pada kode 1

a) Menghitung probabilitas prior $P(H)$

$$P(\text{Laris}) = 633/1.296 = 0,488425926$$

$$P(\text{Kurang Laris}) = 663/1.296 = 0,511574074$$

b) Menghitung probabilitas likelihood $P(X|H)$

$$P(\text{Rasa}) \text{ Kode 5} \mid \text{"Laris"} \text{ Kode 1} = 138/633 = 0,020249221$$

$$P(\text{Rasa}) \text{ Kode 5} \mid \text{"Kurang Laris"} \text{ Kode 2} = 114/663 = 0,027522931$$

$$P(\text{Jenis}) \text{ Kode 1} \mid \text{"Laris"} \text{ Kode 1} = 633/633 = 1$$

$$P(\text{Jenis}) \text{ Kode 1} \mid \text{"Kurang Laris"} \text{ Kode 2} = 663/663 = 1$$

$$P(\text{Kemasan}) \text{ Kode 1} \mid \text{"Laris"} \text{ Kode 1} = 633/633 = 1$$

$$P(\text{Kemasan}) \text{ Kode 1} \mid \text{"Kurang Laris"} \text{ Kode 2} = 663/663 = 1$$

$$P(\text{Bulan}) \text{ Kode 6} \mid \text{"Laris"} \text{ Kode 1} = 46/633 = 0,072669826$$

$$P(\text{Bulan}) \text{ Kode 6} \mid \text{"Kurang Laris"} \text{ Kode 2} = 63/663 = 0,095022624$$

$$P(\text{Tahun}) \text{ Kode 8} \mid \text{"Laris"} \text{ Kode 1} = 86/633 = 0,13586079$$

$$P(\text{Tahun}) \text{ Kode 8} \mid \text{"Kurang Laris"} \text{ Kode 2} = 106/663 = 0,159879336$$

c) Menghitung probabilitas posterior $P(H|X)$

Probabilitas laris

$$= P(\text{Rasa}) * P(\text{Jenis}) * P(\text{Kemasan}) * P(\text{Bulan}) * P(\text{Tahun})$$

$$= 0,020249221 * 1 * 1 * 0,095022624 * 0,13586079 = 0,002152406$$

Probabilitas kurang laris

$$= P(\text{Rasa}) * P(\text{Jenis}) * P(\text{Kemasan}) * P(\text{Bulan}) * P(\text{Tahun})$$

$$= 0,027522931 * 1 * 1 * 0,095022624 * 0,159879336 = 0,002612226$$

d) Menentukan probabilitas tertinggi (keputusan akhir)

Karena probabilitas $P(\text{Kurang Laris} | X) = 0,002612226$ lebih besar daripada probabilitas $P(\text{Laris} | X) 0,002152406$ maka kelas yang diprediksi adalah "Kurang Laris".

4.1.3 Evaluasi model

Setelah proses prediksi selesai, hasil prediksi dibandingkan dengan label asli (y_{test}) untuk mengevaluasi performa model. Evaluasi ini dilakukan dengan menghitung jumlah prediksi yang benar dan salah, kemudian ditampilkan dalam bentuk tabel untuk menghitung akurasi model.

Tabel 4. 10 Evaluasi Model

<i>True label \ Predicted label</i>	Laris	Kurang Laris
Laris (TP+FN)	76	87
Kurang Laris (FP+TN)	47	144

Tabel 4.10 menunjukkan evaluasi model berdasarkan perbandingan antara label asli (*true label*) dan label yang diprediksi (*predicted label*), di mana terdapat 76 data yang benar terklasifikasi sebagai "Laris", 87 data "Laris" yang salah diprediksi sebagai "Kurang Laris", 47 data "Kurang Laris" yang salah diprediksi sebagai "Laris," dan 144 data yang benar terklasifikasi sebagai "Kurang Laris".

Dimana :

- TP (*True Positive*) = 76 (prediksi benar sebagai Laris)
- FN (*False Negative*) = 87 (seharusnya laris, tetapi diprediksi kurang laris)
- FP (*False Positive*) = 47 (seharusnya kurang laris, tetapi diprediksi sebagai laris)

- TN (*True Negative*) = 114 (prediksi benar sebagai kurang laris)

Berikut perhitungan dari akurasi, presisi, *recall* dan *f1 – score* :

1. Perhitungan akurasi

Perhitungan akurasi dilakukan dengan membandingkan jumlah prediksi yang benar dengan total data uji, yang dinyatakan dalam persentase menggunakan rumus berikut :

$$akurasi = \frac{76 + 144}{76 + 144 + 87 + 47} = \frac{190}{324} = 0.5864 = 0.59$$

Dengan demikian hasil perhitungan dari akurasi adalah 0.59

2. Perhitungan *precision*

Precision dihitung dengan membagi jumlah prediksi benar pada kelas positif dengan total prediksi untuk kelas tersebut, menggunakan rumus berikut:

Untuk kelas laris dan kurang laris

$$precision_{laris} = \frac{76}{76 + 47} = \frac{76}{123} = 0.6179$$

$$precision_{kuranglaris} = \frac{114}{114 + 87} = \frac{114}{201} = 0.5672$$

Rata – rata *precision* (*macro average*)

$$precision = \frac{0.6179 + 0.5672}{2} = 0.5926 = 0.59$$

Dengan demikian hasil perhitungan dari *precision* adalah 0.59

3. Perhitungan *recall*

Recall dihitung dengan membagi jumlah prediksi benar pada kelas positif dengan total data aktual untuk kelas tersebut, menggunakan rumus berikut:

Untuk kelas laris dan kurang laris

$$Recall_{laris} = \frac{76}{76 + 87} = \frac{76}{163} = 0.4663$$

$$Recall_{kuranglaris} = \frac{114}{114 + 47} = \frac{114}{161} = 0.7081$$

Rata – rata *recall* (*macro average*)

$$Recall = \frac{0.4663 + 0.7081}{2} = 0.5872 = 0.59$$

Dengan demikian hasil perhitungan dari *recall* adalah 0.59

4. Perhitungan *f1-score*

F1-score dihitung sebagai keseimbangan antara precision dan recall dengan menggunakan rata-rata, sesuai rumus berikut:

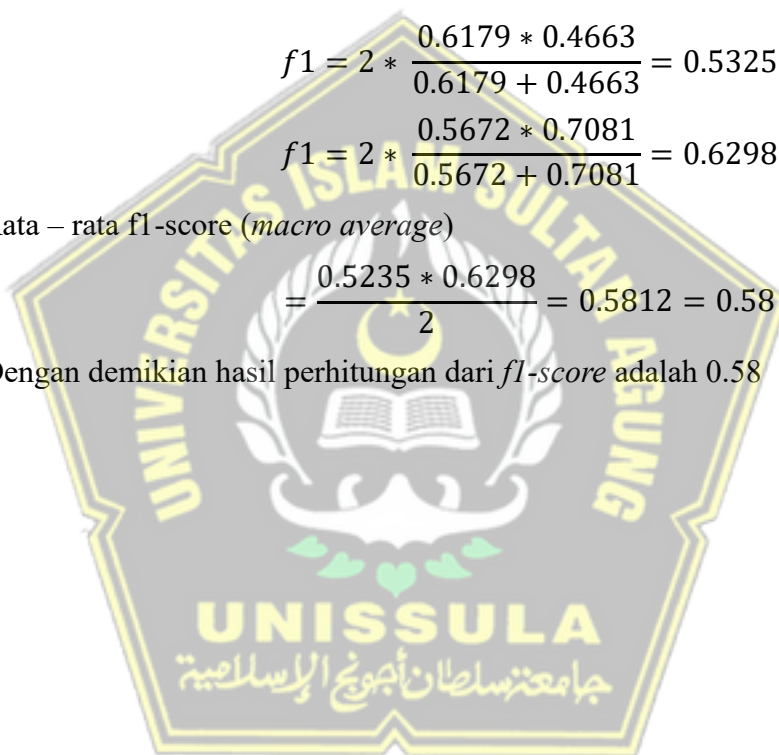
$$f1 = 2 * \frac{0.6179 * 0.4663}{0.6179 + 0.4663} = 0.5325$$

$$f1 = 2 * \frac{0.5672 * 0.7081}{0.5672 + 0.7081} = 0.6298$$

Rata – rata *f1-score* (*macro average*)

$$= \frac{0.5235 * 0.6298}{2} = 0.5812 = 0.58$$

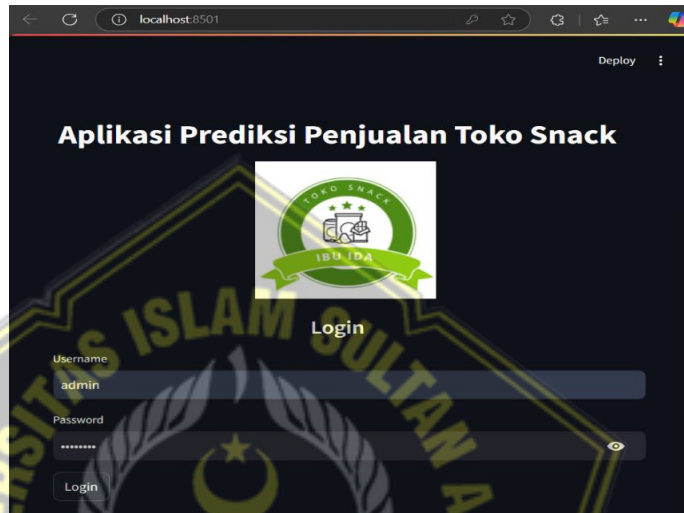
Dengan demikian hasil perhitungan dari *f1-score* adalah 0.58



4.2 Hasil Sistem

Sistem ini menggunakan aplikasi *streamlite* untuk menampilkan prediksi penjualan dengan cara yang sederhana dan interaktif.

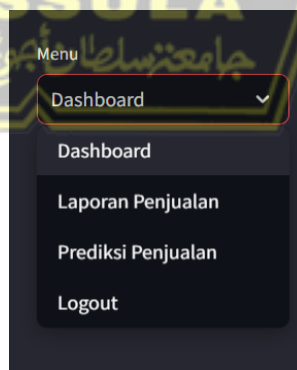
4.2.1 Halaman *Login*



Gambar 4. 1 *Login*

Gambar 4. 1 saat pertama kali membuka sistem, pengguna harus login terlebih dahulu dengan *username* “admin” dan *password* “admin123”.

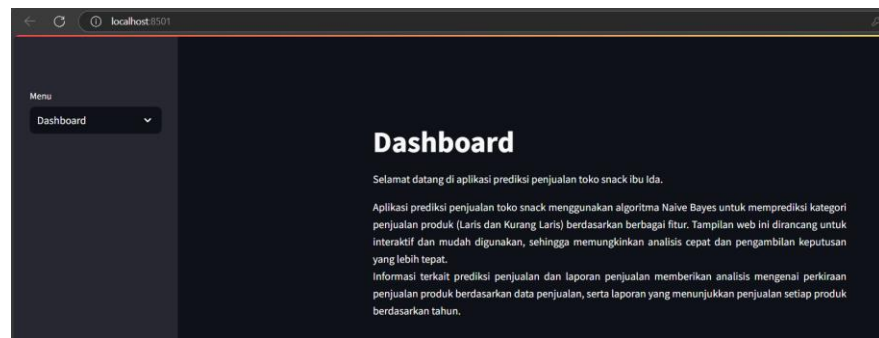
4.2.2 Halaman *Menu*



Gambar 4. 2 *Menu*

Gambar 4. 2 setelah berhasil *login*, pengguna akan melihat menu utama yang berisi dashboard, laporan penjualan, prediksi penjualan dan *logout* untuk mengelola data serta melakukan analisis penjualan.

4.2.3 Halaman Utama (*Dashboard*)



Gambar 4. 3 *Dashboard*

Gambar 4. 3 pengguna dapat memilih menu *dashboard* pada halaman utama untuk melihat deskripsi mengenai aplikasi prediksi penjualan serta informasi terkait fitur sistem.

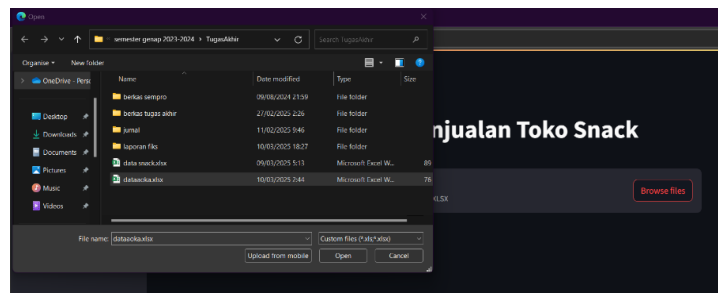
4.2.4 Halaman Laporan Penjualan



Gambar 4. 4 Laporan Penjualan

Gambar 4. 4 pengguna dapat memilih menu laporan penjualan untuk melihat data penjualan yang telah tercatat dalam sistem, termasuk laporan penjualan berdasarkan tahun tertentu.

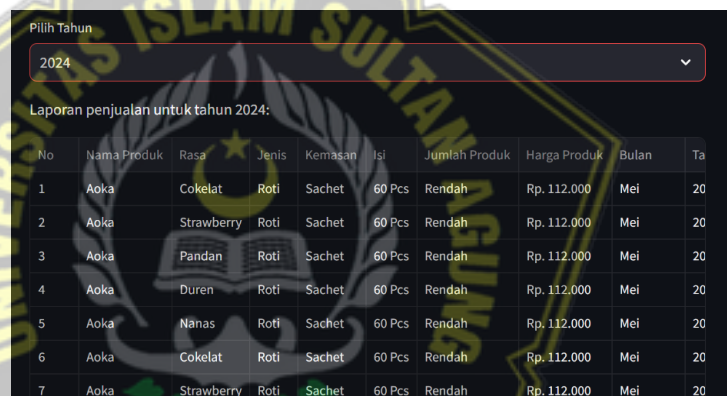
1. Upload data penjualan



Gambar 4. 5 Upload Data

Gambar 4. 5 langkah pertama, pengguna harus mengunggah data penjualan melalui fitur upload data penjualan untuk dapat diproses dalam sistem.

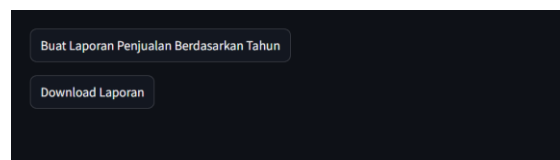
2. Pilih tahun



Gambar 4. 6 Pilih Tahun

Gambar 4. 6 pengguna perlu memilih tahun pada sistem, misalnya memilih tahun 2024, untuk menampilkan dan menganalisis data penjualan berdasarkan periode yang diinginkan.

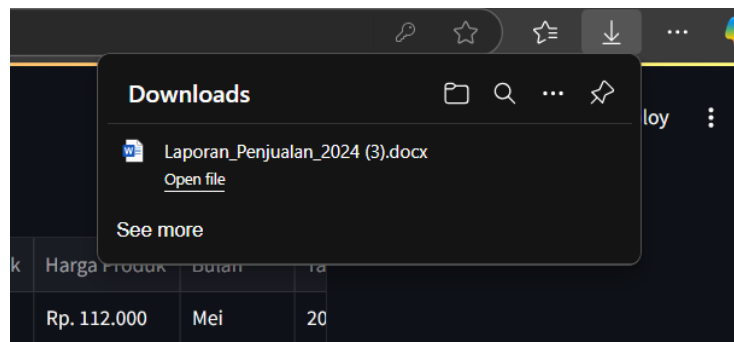
3. Pilih buat laporan



Gambar 4. 7 Buat Laporan dan *Download*

Gambar 4. 7 pengguna dapat memilih fitur buat laporan penjualan berdasarkan tahun untuk menampilkan ringkasan dan analisis penjualan sesuai dengan tahun yang dipilih serta *download* laporan tersebut.

4. Unduh file laporan



Gambar 4. 8 Setelah di *Download*

Pada gambar 4. 8 setelah proses unduh selesai, sistem akan menyimpan laporan penjualan berdasarkan tahun 2024 dalam bentuk *file word*.

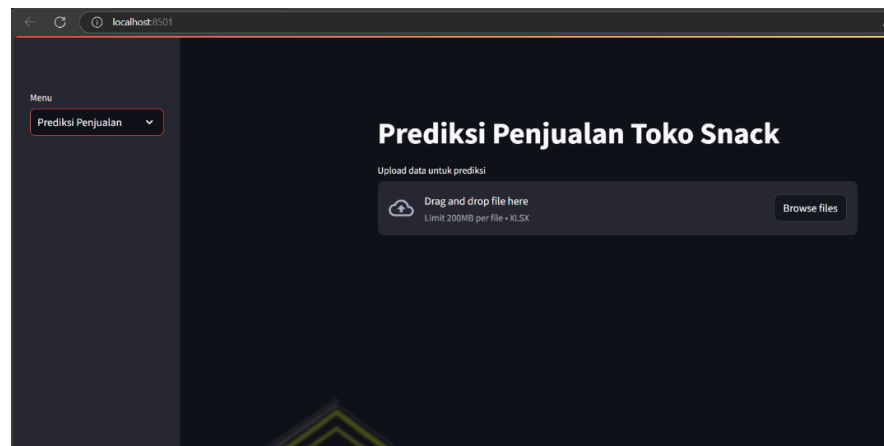
5. Laporan Penjualan

No	Nama Produk	Rasa	Jenis	Kemasan	Isi	Jumlah Produk	Harga Produk	Bulan	Tahun	Jumlah Penjualan
1	Aoka	Cokelat	Roti	Sachet	60 Pcs	Rendah	Rp. 112.000	Mei	2024	Laris
2	Aoka	Strawberry	Roti	Sachet	60 Pcs	Rendah	Rp. 112.000	Mei	2024	Laris
3	Aoka	Pandan	Roti	Sachet	60 Pcs	Rendah	Rp. 112.000	Mei	2024	Laris
4	Aoka	Duren	Roti	Sachet	60 Pcs	Rendah	Rp. 112.000	Mei	2024	Kurang Laris
5	Aoka	Nanas	Roti	Sachet	60 Pcs	Rendah	Rp. 112.000	Mei	2024	Kurang Laris
6	Aoka	Cokelat	Roti	Sachet	60 Pcs	Rendah	Rp. 112.000	Mei	2024	Laris
7	Aoka	Strawberry	Roti	Sachet	60 Pcs	Rendah	Rp. 112.000	Mei	2024	Laris
8	Aoka	Pandan	Roti	Sachet	60 Pcs	Rendah	Rp. 112.000	Mei	2024	Laris
9	Aoka	Duren	Roti	Sachet	60 Pcs	Rendah	Rp. 112.000	Mei	2024	Kurang Laris
10	Aoka	Nanas	Roti	Sachet	60 Pcs	Rendah	Rp. 112.000	Mei	2024	Kurang Laris
11	Aoka	Cokelat	Roti	Sachet	60 Pcs	Rendah	Rp. 112.000	April	2024	Laris
12	Aoka	Strawberry	Roti	Sachet	60 Pcs	Normal	Rp. 112.000	April	2024	Laris
13	Aoka	Pandan	Roti	Sachet	60 Pcs	Normal	Rp. 112.000	April	2024	Laris
14	Aoka	Duren	Roti	Sachet	60 Pcs	Normal	Rp. 112.000	April	2024	Kurang Laris
15	Aoka	Nanas	Roti	Sachet	60 Pcs	Normal	Rp. 112.000	April	2024	Kurang Laris
16	Aoka	Cokelat	Roti	Sachet	60 Pcs	Rendah	Rp. 112.000	April	2024	Laris
17	Aoka	Strawberry	Roti	Sachet	60 Pcs	Normal	Rp. 112.000	April	2024	Laris
18	Aoka	Pandan	Roti	Sachet	60 Pcs	Normal	Rp. 112.000	April	2024	Laris
19	Aoka	Duren	Roti	Sachet	60 Pcs	Normal	Rp. 112.000	April	2024	Kurang Laris
20	Aoka	Nanas	Roti	Sachet	60 Pcs	Normal	Rp. 112.000	April	2024	Kurang Laris

Gambar 4. 9 Laporan Penjualan

Gambar 4. 9 *File* laporan penjualan ini menjelaskan mengenai data penjualan pada tahun 2024, mencakup informasi seperti nama produk, rasa, jenis, kemasan, jumlah produk, harga produk, bulan, serta kategori jumlah penjualan yaitu laris dan kurang laris pada produk.

4.2.5 Halaman Prediksi Penjualan



Gambar 4. 10 Halaman Prediksi Penjualan

Gambar 4. 10 pengguna dapat memilih *menu* prediksi penjualan untuk melakukan prediksi penjualan berdasarkan data yang telah dimasukkan ke dalam sistem.

1. Upload data



Gambar 4. 11 Upload Data

Pada gambar 4. 11 langkah pertama dalam melakukan prediksi adalah mengunggah data penjualan dalam format *excel* ke dalam sistem.

2. Data yang diunggah



Gambar 4. 12 Data yang diunggah

Pada gambar 4. 12 setelah data diunggah, sistem akan menampilkan data yang telah dimasukkan.

3. Pilih grafik penjualan



Gambar 4. 13 Grafik Penjualan

Gambar 4. 13 kemudian pengguna dapat memilih grafik data penjualan yang ditampilkan berdasarkan tahun.

4. Atribut data

☒ Tampilkan Atribut Data

Atribut data:

- value
- No
- Nama Produk
- Rasa
- Jenis
- Kemasan
- Isi
- Jumlah Produk
- Harga Produk
- Bulan
- Tahun

Gambar 4. 14 Atribut Data

Pada gambar 4. 14 selanjutnya, centang atribut data pada tampilan untuk menampilkan value.

5. Menghapus kolom

Hapus Kolom tidak diperlukan:

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun	Jumlah Penjualan
Aoka	Cokelat	Roti	Sachet	Mei	2024	Laris
Aoka	Strawberry	Roti	Sachet	Mei	2024	Laris
Aoka	Pandan	Roti	Sachet	Mei	2024	Laris
Aoka	Duren	Roti	Sachet	Mei	2024	Kurang Laris
Aoka	Nanas	Roti	Sachet	Mei	2024	Kurang Laris
Aoka	Cokelat	Roti	Sachet	Mei	2024	Laris
Aoka	Strawberry	Roti	Sachet	Mei	2024	Laris
Aoka	Pandan	Roti	Sachet	Mei	2024	Laris
Aoka	Duren	Roti	Sachet	Mei	2024	Kurang Laris
Aoka	Nanas	Roti	Sachet	Mei	2024	Kurang Laris

Gambar 4. 15 Hapus Kolom

Gambar 4. 15 sistem telah menghapus beberapa kolom yaitu no, isi, jumlah produk, dan harga produk serta menampilkan kolom yang akan dianalisis dalam sistem yaitu nama produk, rasa, jenis, kemasan, bulan, tahun, dan jumlah penjualan.

6. Cleaning data

Lakukan Cleaning Data

Data setelah dilakukan cleaning:

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun	Jumlah Penjualan
1	1	1	1	1	1	1
1	2	1	1	1	1	1
1	3	1	1	1	1	1
1	4	1	1	1	1	2
1	5	1	1	1	1	2
2	2	1	1	1	1	1
2	1	1	1	1	1	1
2	5	1	1	1	1	2
3	1	2	1	1	1	1
3	6	2	1	1	1	2

Gambar 4. 16 *Cleaning Data*

Gambar 4. 16 setelah data dianalisis dalam sistem, sistem melakukan proses pembersihan data (cleaning) dengan mengubah teks menjadi numerik. Dengan data dalam bentuk numerik, sistem dapat mengelolanya menggunakan algoritma naïve bayes.

7. Tahap klasifikasi algoritma naïve bayes

Tahapan Klasifikasi Algoritma Naïve Bayes

Jumlah data pada variabel X: 1620 (6 kolom)

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun
1	1	1	1	1	1
1	2	1	1	1	1
1	3	1	1	1	1
1	4	1	1	1	1
1	5	1	1	1	1
1	1	1	1	1	1
1	2	1	1	1	1
1	3	1	1	1	1
1	4	1	1	1	1
1	5	1	1	1	1

Gambar 4. 17 Tahap Klasifikasi *Naïve Bayes*

Gambar 4. 17 setelah melalui proses pembersihan data (cleaning), tahap selanjutnya adalah klasifikasi menggunakan algoritma naïve bayes. Pada tahap ini,

data disiapkan dengan variabel x dan y untuk proses prediksi. Jumlah data pada variabel x adalah 1.620 dengan 6 kolom.

Jumlah data pada variabel y: 1620 (1 kolom)

Jumlah Penjualan
1
1
1
2
2
1
1
1
2
2

Gambar 4. 18 Variabel y

Gambar 4. 18 kemudian, variabel y disiapkan dengan 1.620 data dalam 1 kolom, yaitu jumlah penjualan. Variabel ini digunakan untuk menentukan hasil prediksi, apakah termasuk kategori 1 (laris) atau 2 (kurang laris).

8. Pembagian data

Pembagian Data Training dan Testing

Jumlah data untuk training: 1296 (80%)

Jumlah data untuk testing: 324 (20%)

Data Training Variabel x (1296 data, 6 kolom):

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun
1	5	1	1	2	5
1	3	1	1	3	5
1	3	1	1	10	2
1	5	1	1	7	6
1	4	1	1	4	7
1	1	1	1	4	5
1	5	1	1	9	4
1	5	1	1	7	6
1	5	1	1	5	8
1	2	1	1	5	7

Gambar 4. 19 Pembagian Data *Training* dan *Testing*

Gambar 4. 19 selanjutnya, dilakukan pembagian data, dimana data training terdiri dari 1.296 data (80%) dan testing sebanyak 324 data (20%). Data yang

digunakan untuk training data pada variabel x berjumlah 1.296 data dengan 6 kolom.

Data Training Variabel y : 1296

Jumlah Penjualan
1
2
1
2
1
2
1
2
1
2
1
2

Gambar 4. 20 Data Training Variabel y

Pada gambar 4. 20 kemudian, pada variabel y terdapat 1.296 data yang digunakan dalam proses pelatihan model.

9. Testing data

Data Testing (324 data, 6 kolom):

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun
1	3	1	1	7	4
1	4	1	1	10	5
1	5	1	1	8	5
1	4	1	1	1	7
1	5	1	1	6	8
1	3	1	1	5	6
1	4	1	1	9	6
1	4	1	1	7	6
1	4	1	1	4	5
1	1	1	1	4	3

Gambar 4. 21 Testing Data

Gambar 4. 21 setelah itu, data diuji (testing) menggunakan 324 data, yaitu 20% dari keseluruhan data, pada variabel x. proses ini bertujuan untuk memprediksi apakah produk tersebut termasuk dalam kategori 1 (laris) atau 2 (kurang laris).

10. Hasil prediksi

Hasil Prediksi Penjualan Tahun 2025

Data Penjualan (324 data, 8 kolom):

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun	Prediksi Jumlah Penjualan	Kategori Prediksi
1	3	1	1	7	4	2	Kurang Laris
1	4	1	1	10	5	2	Kurang Laris
1	5	1	1	8	5	1	Laris
1	4	1	1	1	7	2	Kurang Laris
1	5	1	1	6	8	2	Kurang Laris
1	3	1	1	5	6	2	Kurang Laris
1	4	1	1	9	6	2	Kurang Laris
1	4	1	1	7	6	2	Kurang Laris
1	4	1	1	4	5	2	Kurang Laris
1	1	1	1	4	3	2	Kurang Laris

Gambar 4. 22 Hasil Prediksi

Gambar 4. 22 setelah data diuji, sistem telah menghasilkan prediksi penjualan untuk tahun 2025 dengan 324 data yang diprediksi masuk ke dalam kategori 1 (laris) dan 2 (kurang laris). Jumlah produk berdasarkan prediksi dalam kategori 1 (laris) sebanyak 123, sedangkan dalam kategori 2 (kurang laris) sebanyak 201.

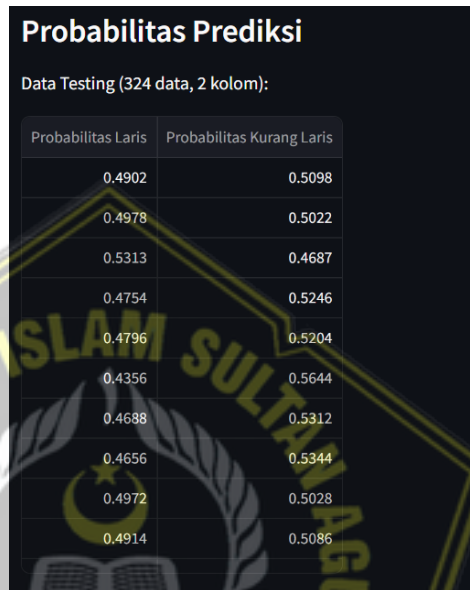
Berikut hasil prediksi penjualan ditahun 2025 antara lain:

Tabel 4. 1 Hasil Prediksi

Nama Produk	Rasa	Jenis	Kemasan	Bulan	Tahun	Jumlah penjualan
Aoka	Pandan	Roti	Sachet	November	2021	Kurang Laris
Aoka	Duren	Roti	Sachet	Agustus	2020	Kurang Laris
Aoka	Nanas	Roti	Sachet	Oktober	2020	Laris
Aoka	Duren	Roti	Sachet	Mei	2018	Kurang laris
Aoka	Nanas	Roti	Sachet	Desember	2017	Kurang laris

Tabel 4.1 menampilkan hasil prediksi penjualan untuk tahun 2025 dengan lima data produk, di mana aoka dengan rasa pandan bulan november, duren agustus dan mei serta nanas bulan september diprediksi sebagai “Kurang Laris”, sementara nanas dibulan oktober diprediksi sebagai “Laris”.

11. Probabilitas prediksi



Data Testing (324 data, 2 kolom):	
Probabilitas Laris	Probabilitas Kurang Laris
0.4902	0.5098
0.4978	0.5022
0.5313	0.4687
0.4754	0.5246
0.4796	0.5204
0.4356	0.5644
0.4688	0.5312
0.4656	0.5344
0.4972	0.5028
0.4914	0.5086

Gambar 4. 23 Probabilitas Prediksi

Gambar 4. 23 hasil prediksi kategori laris dan kurang laris menampilkan probabilitas untuk setiap data dari 324 data yang telah diprediksi.

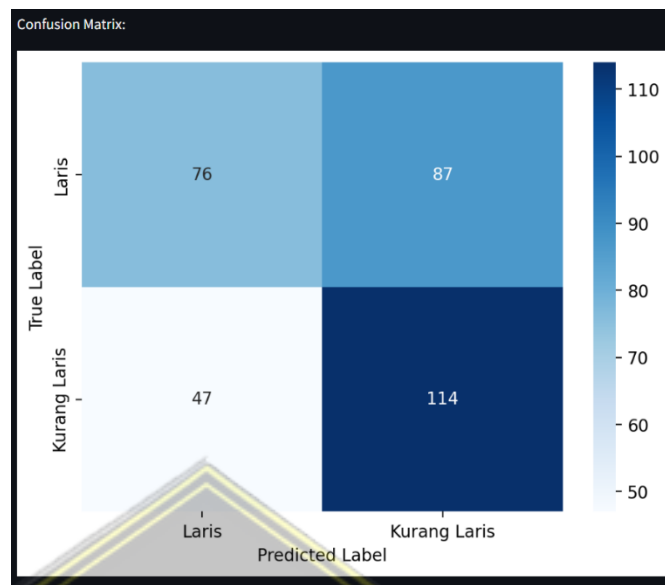
12. Evaluasi model



Akurasi: 0.59
Precision: 0.59
Recall: 0.59
F1-Score: 0.58

Gambar 4. 24 Hasil Evaluasi Model

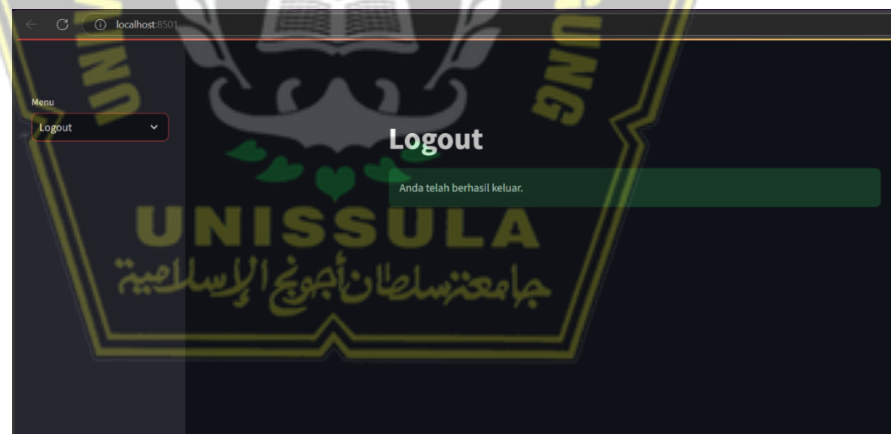
Pada gambar 4. 24 evaluasi model dilakukan setelah memperoleh hasil prediksi dengan menampilkan nilai akurasi 0.59, precision 0.59, recall 0.59 dan f1-score 0.58.



Gambar 4. 25 *Confusion Matrix*

Pada gambar 4. 25 setelah itu, sistem menampilkan visualisasi confusion matrix untuk menganalisis performa hasil prediksi.

4.2.6 Halaman Logout



Gambar 4. 26 *Logout*

Pada gambar 4. 26 setelah selesai menggunakan sistem, pengguna dapat keluar melalui menu logout, dan sistem akan menampilkan notifikasi “Anda telah berhasil keluar”.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Hasil penelitian ini menunjukkan bahwa prediksi penjualan pada toko snack ida menggunakan metode *naïve bayes* belum mencapai tingkat akurasi yang diharapkan. model prediksi telah diimplementasikan dalam sistem berbasis *web* menggunakan *Streamlit* untuk memudahkan prediksi penjualan secara interaktif. Studi kasus pada toko snack ida menunjukkan bahwa penerapan metode *naïve bayes* dalam memprediksi penjualan masih memiliki keterbatasan, sehingga perlu dilakukan perbaikan atau pengembangan untuk meningkatkan performa model. Dari 1.620 data yang diproses, sebanyak 1.296 data yang digunakan sebagai data latih, sementara 324 data digunakan sebagai data uji. Model yang dihasilkan memperoleh akurasi 0,59, presisi 0,59, *recall* 0,59 dan *f1-score* 0,58. hasil prediksi ini masih kurang atau belum maksimal dalam akurasi. Faktor – faktor yang mempengaruhi prediksi menurun :

1. Ketidakseimbangan data terjadi pada jumlah sampel dalam satu kategori jauh lebih sedikit dibandingkan dengan kategori lainnya, sehingga cenderung lebih akurat dalam mengenali kategori yang dominan tetapi mengalami kesulitan dalam memprediksi kategori yang jarang muncul.
2. *Naive bayes* memiliki keterbatasan karena mengasumsikan bahwa setiap fitur tidak saling bergantung, padahal dalam kenyataanya banyak fitur yang memiliki hubungan erat, sehingga ada fitur yang berkorelasi kuat, model dapat menghasilkan prediksi yang akurat.
3. Model yang menggunakan parameter *default* tanpa optimasi tidak sesuai dengan dataset, sehingga perlu dilakukan *hyperparameter tuning* menggunakan *gridsearchcv* atau *randomizedsearchcv* untuk meningkatkan performa.

5.2 Saran

Adapun saran penelitian ini untuk ke depannya sebagai berikut:

1. Untuk menangani ketidakseimbangan data, menambahkan sampel pada kategori yang kurang, mengurangi sampel pada kategori yang dominan atau menyesuaikan bobot agar model lebih seimbang dalam membuat prediksi.
2. Mengatasi keterbatasan *naive bayes*, dapat menggunakan *feature selection* untuk memilih fitur yang relevan, menghapus fitur yang berkorelasi agar tidak menurunkan prediksi atau dengan menggunakan model *random forest* atau *xgboost* yang mampu menangani antar fitur.
3. Meningkatkan performa model lakukan *hyperparameter tuning* dengan *gridsearchcv* atau *randomsearchcv* agar parameter yang digunakan lebih optimal sesuai dengan dataset.



DAFTAR PUSTAKA

- Al Husaini, M., Hermansyah, A., Purwayoga, V., Lukmana, H. H., & Ramadhan, D. (2022). Aplikasi Cerdas Berbasis Website Prediksi Harga Emas dengan Implementasi Algoritma Smoothing Time Series Forecasting. *Data Sciences Indonesia (DSI)*, 2(2), 30–43. <https://doi.org/10.47709/dsi.v2i2.1888>
- Anggraini, F., Rustam, R., Rahmatullah, S., & Supriyanto, S. (2024). PENERAPAN DATA MINING JUMLAH PENJUALAN SEPEDA MOTOR MENGGUNAKAN METODE K-MEANS CLUSTERING (Studi Kasus : PT. Tunas Dwipa Matra Way Kanan). *Jurnal Informatika*, 24(1), 42–51. <https://doi.org/10.30873/ji.v24i1.3685>
- Gulo, M., Sumarno, & Anggraini, F. (2023). Penerapan Algoritma Naïve Bayes dalam Memprediksi Harga Penjualan Makanan Ringan Application of the Naïve Bayes Algorithm in Predicting Sales Prices for Snacks. *JOMLAI: Journal of Machine Learning and Artificial Intelligence Vol.*, 2(2), 125–132. <https://doi.org/10.55123/jomlai.v2i2.2444>
- Hadi, N., & Benedict, J. (2024). Implementasi Machine Learning Untuk Prediksi Harga Rumah Menggunakan Algoritma Random Forest. *Computatio : Journal of Computer Science and Information Systems*, 8(1), 50–61. <https://doi.org/10.24912/computatio.v8i1.15173>
- Hajaroh, H., Suprpti, T., & Narasati, R. (2024). Implementasi Algoritma Naive Bayes Untuk Analisis Sentimen Ulasan Produk Makanan Dan Minuman Di Tokopedia. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 111–118. <https://doi.org/10.36040/jati.v8i1.8237>
- Harahap, F., Fahrozi, W., Adawiyah, R., Siregar, E. T., & Harahap, A. Y. N. (2023). Implementasi Data Mining dalam Memprediksi Produk AC Terlaris untuk Meningkatkan Penjualan Menggunakan Metode Naive Bayes. *Jurnal Unitek*, 16(1), 41–51. <https://doi.org/10.52072/unitek.v16i1.541>

- Harmaja, O. J., Purba, W., Siregar, M. P., Manurung, H. T., & Sirait, F. A. (2022). Analisis Tingkat Kepuasan Pelanggan Pada Rudang Hotel Berastagi Menggunakan Metode Customer Satisfaction Index (Csi). *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, 5(1), 147. <https://doi.org/10.37600/tekinkom.v5i1.511>
- Indah Kusuma Wardhani, N., Hartami S, I., & Dwi Puspitasari, W. (2022). Sistem Forecasting Penjualan Beras Dengan Menerapkan Metode Trend Moment. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 901–907. <https://doi.org/10.36040/jati.v6i2.5780>
- 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.
- Juwita, Safii, M., & Efendi Damanik, B. (2022). Algoritma Naïve Bayes Untuk Memprediksi Penjualan Pada Toko VJCakes Pematang Siantar. *Journal of Machine Learning and Artificial Intelligence*, 1(4), 337–346. <https://doi.org/10.55123/jomlai.v1i4.1674>
- Mahar, N. M. A., Vihi Atina, & Nugroho Arif Sudibyo. (2023). Pemodelan Prediksi Kelulusan Mahasiswa Dengan Metode Naïve Bayes Di Uniba. *Jurnal Manajemen Informatika Dan Sistem Informasi*, 6(2), 148–158. <https://doi.org/10.36595/misi.v6i2.875>
- Metode, P., Algoritma, C., & Naive, D. A. N. (2024). Perbandingan metode algoritma c4.5 dan naive bayes untuk memprediksi penjualan kosmetik pada toko jelita 1,2. 7(2), 220–225.
- Nurhayati aris, N. (2021). Penerapan Algoritma Naive Bayes Dalam Menentukan Kelayakan Nasabah Pada Koperasi Simpan Pinjam. *Journal of Computer and Information System (J-CIS)*, 1(1), 1–9. <https://doi.org/10.31605/jcis.v2i2.811>
- Pramana, I., Sudiarsa, I. W., & ... (2023). Penerapan Algoritma Naive Bayes Untuk

- Prediksi Penjualan Produk Terlaris Pada CV Akusara Jaya Abadi. *JATISI (Jurnal Teknik ...)*, 10(4), 518–534.
<https://jurnal.mdp.ac.id/index.php/jatisi/article/view/6498%0Ahttps://jurnal.mdp.ac.id/index.php/jatisi/article/download/6498/1694>
- Purnama, N., Insani, F., Haerani, E., Afrianty, I., Studi, P., Informatika, T., Islam, U., Sultan, N., Kasim, S., & Riau, U. M. (2024). *Jurnal Computer Science and Information Technology (CoSciTech) Prediction Model for the Number of Machine Lubricant Sales at PT. X With the Naïve Bayes Algorithm*. 5(3), 609–618.
- Putra, F. B., Chulkamdi, M. T., & Febrinita, F. (2024). *IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI DATA STOK FUKUBI OUTFIT MENGGUNAKAN METODE K- NEAREST NEIGHBOR*. 8(5), 9907–9914.
- Putra, P. W. P., Irawan, D., & ... (2022). Sistem Prediksi Persediaan Barang Di Mini Market Mars Menggunakan Algoritma Naïve Bayes Classifier Dengan Pemrograman *JUTIM (Jurnal Teknik ...)*, 7(2), 148–159.
<https://jurnal.univbinainsan.ac.id/index.php/jutim/article/view/2101%0Ahttps://jurnal.univbinainsan.ac.id/index.php/jutim/article/download/2101/1077>
- Rozi, M. F., Siregar, R., & Syahputri, N. I. (2023). *Penerapan Data Mining Menggunakan Metode Naïve Bayes Untuk Klasifikasi Data Penentuan Hasil Penjualan Dalam Strategi Pemasaran*. 2(2), 444–454.
- Science, D., Technology, I., & Analytics, D. (2023). *Journal of Dinda. Journal.Ittelkom-Pwt.Ac.Id/Index.Php/Dinda*, 3(2), 65–70.
- Supendar, H., Suharyanti, N., Bina, U., Informatika, S., & Nusamandiri, U. (2023). *Application Of The Naïve Bayes Algorithm In Determining Sales Of The Month*. 7(2), 873–879.
- Suryana, A., Irma Purnamasari, A., & Ali, I. (2024). Mengoptimalkan Kepuasan Pengguna: Analisis Sentimen Review Aplikasi Grab Di Indonesia. *JATI*

(*Jurnal Mahasiswa Teknik Informatika*), 8(3), 3396–3404.
<https://doi.org/10.36040/jati.v8i3.9688>

Zulkifli, Asmawati.S, & Arnita Irianti. (2022). Penerapan Algoritma Naive Bayes dalam Memprediksi Persediaan Bahan Mebel (Studi Kasus Mebel Usaha Bersama Palipi Soreang). *Journal of Computer and Information System (J-CIS)*, 5(1), 57–64. <https://doi.org/10.31605/jcis.v5i1.1360>

