

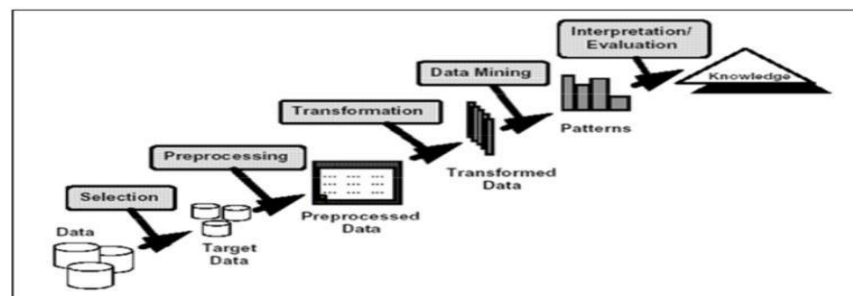
BAB II

LANDASAN TEORI

2.1 *Knowledge Discovery in Databases (KDD)*

Knowledge Discovery in Databases (KDD) adalah sebuah proses yang sistematis dan terstruktur untuk menemukan pengetahuan baru, pola tersembunyi, atau informasi yang bernilai dari kumpulan data dalam jumlah besar. Proses ini berkembang karena meningkatnya volume data yang tersimpan dalam berbagai sistem informasi

KDD bukan sekadar menjalankan algoritma untuk menganalisis data. KDD mencakup semua tahapan mulai dari mengumpulkan data, membersihkan data, mempersiapkannya untuk dianalisis, menjalankan teknik data mining, hingga menafsirkan hasilnya sehingga dapat digunakan sebagai dasar pengambilan keputusan.[1](Virgiana et al., 2025)



**Gambar 21. Tahapan
Proses KDD**

Berikut adalah penjelasan mendetail mengenai tahapan-tahapan dalam Knowledge

Discovery In Database (KDD):

1. Seleksi atau Pemilihan Data

Tahap pertama adalah memilih data dari berbagai sumber. Data yang

dipilih harus relevan dengan masalah atau tujuan analisis. Misalnya, untuk menganalisis kehilangan pelanggan, data yang digunakan mungkin meliputi riwayat transaksi, durasi berlangganan, dan aktivitas pelanggan.(Irwan & Nasution, 2024)

2. Pemrosesan dan Pembersihan Data

Data processing adalah proses mengolah data yang baru di dapat menjadi informasi yang bisa dimengerti dan digunakan. Data yang awalnya masih acak, berantakan, atau tidak jelas akan diproses supaya lebih rapi, teratur, dan berguna. Di dalam proses ini, data biasanya dibersihkan dulu, seperti memperbaiki data yang salah, menghapus data ganda, atau mengisi data yang kosong. Setelah itu, data bisa diurutkan, dikelompokkan, dihitung, atau diubah sesuai kebutuhan(HILDA, 2022)

3. Transformasi

Data transformation adalah proses mengubah data agar bentuknya lebih rapi dan teratur. Tujuan dari data transformation adalah supaya data lebih siap diproses dan hasil analisisnya lebih akurat.(Sofiya Azzara Rafles & Muhammad Irwan Padli Nasution, 2023)

4. Data Mining

Data mining adalah proses mencari pola atau informasi penting dari kumpulan data yang sangat banyak. Dalam data mining, komputer digunakan untuk menemukan hal-hal yang sebelumnya tidak terlihat, seperti kebiasaan pelanggan, kelompok orang yang mirip, hubungan

antar data, atau hal-hal yang tidak wajar. Tujuan data mining adalah mendapatkan pengetahuan baru dari data sehingga bisa membantu membuat keputusan (Rifka Alkhilyatul Ma'rifat, I Made Suraharta, 2024a)

5. Evaluasi

Setelah pola atau model sudah ditemukan maka langkah selanjutnya adalah evaluasi. Pada fase ini model yang ditemukan diuji untuk mengevaluasi efektivitasnya dalam mendeskripsikan data hasil yang diinginkan. Evaluasi bertujuan untuk mengetahui apakah pola atau model yang ditemukan cukup valid dan layak digunakan dalam pengambilan keputusan. (Rifka Alkhilyatul Ma'rifat, I Made Suraharta, 2024b)

2.2 Data Mining

Data mining adalah proses mengambil informasi atau pola penting dari data yang sangat banyak dengan cara yang teratur dan rapi proses ini dimulai dari membersihkan data yang berantakan, lalu mencari pola seperti kelompok data yang mirip, hubungan antar barang yang sering dipakai atau dibeli bersama, memprediksi hal yang akan terjadi ke depan, atau menemukan data yang tidak normal; setelah pola ditemukan, hasilnya disajikan dalam bentuk yang mudah dipahami seperti tabel atau grafik, sehingga data mining dapat membantu berbagai bidang seperti toko online, bank, kesehatan, sekolah, dan lain-lain untuk membuat keputusan yang lebih

tepat dan bermanfaat dari data yang mereka punya.

Mereka menyoroti bahwa data mining merupakan rangkaian langkah-langkah untuk menggali pola-pola dalam data dengan memanfaatkan algoritma seperti Naive Bayes serta teknik pembobotan seperti SAW. Selain itu, mereka mengungkapkan bahwa data mining juga berperan penting dalam mendukung proses pengambilan keputusan yang didasarkan pada data di dalam suatu organisasi

Data mining merupakan suatu metode yang menerapkan pendekatan statistik, matematika, kecerdasan buatan, serta pembelajaran mesin (machine learning) untuk menarik keluar dan menemukan informasi yang berguna serta pengetahuan yang relevan dari data atau sumber terkait terkait.(Wu et al., 2026)

mendefinisikan data mining sebagai serangkaian proses analitik yang menggunakan teknik statistik dan algoritma machine learning untuk mengidentifikasi pola tersembunyi dalam kumpulan data besar. Penelitian ini menggaris bawahi bahwa data mining sangat relevan untuk segmentasi pelanggan dalam layanan pemebel di coffee shop, di mana proses ini memungkinkan pengelompokan pembeli berdasarkan pola tertentu untuk meningkatkan efisiensi layanan penjualan(Santasup & Tengponsathon, 2022)

menjelaskan bahwa data mining, khususnya menggunakan algoritma seperti *Naive Bayes* dan *Chi-Square*, adalah alat yang sangat berguna untuk analisis sentimen berbasis teks. Penerapannya mencakup klasifikasi

opini dan prediksi kepuasan pelanggan di berbagai sektor(Wijaya et al., 2019)

2.3 Algoritma *Naïve Bayes*

Naive Bayes merupakan teknik klasifikasi statistik yang berbasis pada teorema Bayes, di mana ia menghitung probabilitas suatu kelas berdasarkan fitur atau kriteria yang tersedia, lalu memilih kelas dengan probabilitas tertinggi. Dalam penelitian perbandingan algoritma klasifikasi, metode ini terbukti cepat dan memberikan akurasi tinggi saat diterapkan pada dataset berukuran besar. Algoritma ini populer untuk menyelesaikan masalah klasifikasi karena menurut literatur ia memiliki akurasi yang baik serta proses perhitungan yang mudah(Suharyanto, n.d.)

Naïve bayes mengenali nilai penting atau kunci dengan menghitung frekuensi kemunculan variabel dalam data, yang menghasilkan probabilitas untuk elemen tertentu. Untuk menghitung probabilitas posterior, algoritma ini memerlukan probabilitas prior, likelihood, dan probabilitas prediktor sebelumnya. Persamaan 3 menunjukkan bahwa untuk data dengan banyak fitur, pendekatan ini tetap efektif.

,[6](Saputra & Hasan, 2023)

Naïve bayes Sebagai algoritma dalam data mining yang menggunakan teori Bayes untuk klasifikasi, *Naive Bayes* dikenal dengan tingkat kesalahan yang sangat rendah, serta perhitungan yang simpel, cepat, dan akurat. Metode ini sering dianggap lebih unggul dibandingkan

pengklasifikasi lain karena akurasi yang superior. Namun, ia memiliki kekurangan, seperti performa prediksi probabilitas yang kurang optimal dan akurasi yang menurun jika pemilihan fitur tidak relevan dengan klasifikasi. (Pebdika et al., 2023)

Dalam aplikasinya, Naive Bayes banyak diterapkan di berbagai sektor, seperti analisis sentimen, penyaringan email spam, dan pengelompokan dokumen. Sebagai ilustrasi, dalam bidang analisis sentimen, algoritma ini mampu mengkategorikan pandangan konsumen mengenai suatu produk atau jasa dengan menganalisis teks dari platform media sosial (Nursatika Kusuma & Ali, 2024)

Selain itu, performa algoritma dapat ditingkatkan melalui metode tambahan, misalnya dengan memilih fitur menggunakan Information Gain, yang bertujuan untuk memperbaiki tingkat keakuratan klasifikasi. [9]. Perbandingan dengan Algoritma Lain Studi menunjukkan bahwa Naive Bayes mampu bersaing dengan algoritma seperti K-Nearest Neighbor (K-NN) dan Decision Tree dalam aspek akurasi klasifikasi [10].

Untuk perhitungan pada metode Naive Bayes menggunakan rumus sebagai berikut.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$P(A|B)$ = Probabilitas A bersyarat yang diberikan oleh B
 $P(B|A)$ = Probabilitas B bersyarat yang diberikan oleh A
 $P(A)$ = Probabilitas kejadian A

$P(B)$ = Probabilitas kejadian B

2.3.1 Kepuasan Pelayanan di *Coffee Shop*

Kepuasan layanan di *coffee shop* merupakan salah satu indikator utama untuk menilai kualitas layanan yang disediakan. Tingkat kepuasan pelanggan dipengaruhi oleh berbagai aspek, termasuk kualitas interaksi dengan pelayan, fasilitas yang ada, dan efisiensi sistem administrasi.

kepuasan pelanggan di *coffee shop* sangat dipengaruhi oleh mutu layanan yang diberikan. *Coffee shop* tersebut melibatkan elemen-elemen seperti ketanggapan staf, kehandalan sistem, dan kepedulian terhadap konsumen. Studi ini mengungkapkan bahwa *coffee shop* dengan karyawan yang tanggap dan bersahabat menunjukkan tingkat kepuasan pembeli yang lebih besar jika dibandingkan dengan *coffee shop* yang memiliki staf pelayan kurang komunikatif.[11]

menyoroti pentingnya fasilitas fisik dan kebersihan dalam meningkatkan kenyamanan pelanggan. Hasil penelitian mengungkapkan bahwa kafe yang menjaga kebersihan lokasi dan menyediakan ruang yang nyaman cenderung mendapatkan tingkat kepuasan yang lebih tinggi. Di samping itu, layanan administrasi yang efisien dan sederhana juga berkontribusi besar terhadap pengalaman pelanggan secara keseluruhan..[12]

Penelitian oleh [13]. Mengindikasikan bahwa kemampuan penyedia layanan untuk merespons dengan cepat sangat mempengaruhi tingkat kepuasan pelanggan. Elemen-elemen ini biasanya dinilai melalui survei

yang mencakup dimensi seperti empati, kecepatan pelayanan, dan keandalan layanan di coffee shop.

Indikator utama kepuasan pelanggan dalam suatu layanan memainkan peran krusial untuk mengukur kualitas pelayanan yang disediakan. Berbagai studi mengungkapkan bahwa beberapa aspek turut berkontribusi pada kepuasan pelanggan, seperti kualitas layanan, kualitas informasi, dan kualitas sistem. Di antara itu, kualitas layanan yang meliputi elemen seperti kecepatan, ketepatan, dan empati dalam pelayanan merupakan elemen utama yang paling berpengaruh terhadap kepuasan pelanggan.[14]

Pengukuran kepuasan pengguna juga dapat dilakukan dengan menggunakan berbagai metode, seperti servqual dan eucs, yang masing-masing mengukur dimensi kualitas layanan dari perspektif pengguna. Metode servqual, misalnya, mencakup lima dimensi yaitu tangibles, reliability, responsiveness, assurance, dan empathy, yang semuanya berkontribusi terhadap persepsi pengguna mengenai kualitas layanan yang diterima[15]. Sementara itu, EUCS lebih fokus pada aspek kepuasan pengguna terhadap sistem informasi yang digunakan[16].

Salah satu pendekatan yang umum digunakan dalam survei kepuasan adalah Survei Kepuasan Masyarakat (SKM), yang bertujuan untuk mengukur tingkat kepuasan pembeli. Penelitian menunjukkan bahwa variabel yang digunakan dalam survei harus relevan, valid, dan reliabel agar dapat mencerminkan faktor-faktor yang benar-benar mempengaruhi kepuasan pembeli[18].

2.3.2 Strategi Penjualan di *Coffee Shop*

Strategi penjualan di *coffee shop* merupakan salah satu faktor penting dalam meningkatkan pendapatan dan mempertahankan loyalitas pelanggan. Strategi ini mencakup berbagai upaya yang dilakukan oleh pengelola *coffee shop* untuk menarik minat konsumen, meningkatkan frekuensi pembelian, serta menyesuaikan produk dan layanan dengan preferensi pelanggan. Dalam penelitian ini, strategi penjualan dianalisis menggunakan metode Naive Bayes untuk mengklasifikasikan dan memprediksi pola pembelian pelanggan berdasarkan data historis penjualan.

1. Harga produk

Harga merupakan salah satu faktor utama yang memengaruhi keputusan pembelian pelanggan di *coffee shop*. Penetapan harga yang sesuai dengan daya beli dan persepsi kualitas pelanggan dapat meningkatkan volume penjualan. Penelitian sebelumnya menunjukkan bahwa harga yang kompetitif dan seimbang dengan kualitas produk memiliki pengaruh signifikan terhadap minat beli pelanggan di *coffee shop*.

2. Promosi penjualan

Promosi penjualan meliputi pemberian diskon, paket bundling, voucher, dan program loyalitas pelanggan. Strategi promosi yang tepat dapat menarik pelanggan baru sekaligus mempertahankan pelanggan lama. Berdasarkan penelitian terdahulu, promosi yang

dilakukan secara berkala dan terarah mampu meningkatkan penjualan, terutama pada jam-jam sepi atau hari tertentu.

3. Kualitas produk

Kualitas produk, khususnya cita rasa kopi dan konsistensi penyajian, menjadi faktor penting dalam menentukan kepuasan dan keputusan pembelian ulang pelanggan. Coffee shop yang mampu menjaga kualitas produknya cenderung memiliki pelanggan yang loyal. Data kualitas produk ini dapat dianalisis untuk melihat pengaruhnya terhadap tingkat penjualan.

4. Pelayanan

Pelayanan yang ramah, cepat, dan profesional dapat memberikan pengalaman positif bagi pelanggan. Pelayanan yang baik sering kali menjadi alasan pelanggan untuk kembali berkunjung. Dalam konteks analisis data, aspek pelayanan dapat dikaitkan dengan tingkat kepuasan pelanggan yang berdampak pada peningkatan penjualan.

Melalui analisis faktor-faktor tersebut menggunakan algoritma Naive Bayes, coffee shop dapat mengklasifikasikan data penjualan ke dalam kategori tertentu, seperti penjualan tinggi dan penjualan rendah. Hasil klasifikasi ini kemudian digunakan sebagai dasar dalam menentukan strategi penjualan yang lebih efektif dan berbasis data, sehingga dapat membantu pengelola coffee shop dalam mengambil keputusan bisnis secara tepat dan terukur.

2.3.3 Analisis Sentimen Pelanggan Coffee Shop

Analisis sentimen adalah teknik untuk mengklasifikasikan opini pelanggan terhadap produk dan layanan coffee shop berdasarkan data teks, seperti ulasan pelanggan atau hasil kuesioner. Analisis ini digunakan untuk mengetahui persepsi pelanggan terhadap kualitas produk, harga, pelayanan, dan suasana coffee shop.

Algoritma Naive Bayes digunakan untuk mengelompokkan sentimen pelanggan ke dalam kategori positif, negatif, dan netral berdasarkan probabilitas kemunculan kata atau atribut tertentu. Metode ini dipilih karena proses perhitungannya sederhana dan memiliki tingkat akurasi yang baik.

Hasil analisis sentimen dapat dimanfaatkan sebagai dasar dalam menentukan strategi penjualan, seperti peningkatan kualitas menu, perbaikan pelayanan, dan penentuan promosi yang sesuai dengan kebutuhan pelanggan. Dengan demikian, analisis sentimen membantu pengelola coffee shop dalam mengambil keputusan penjualan secara tepat dan berbasis data.

2.4. Tahapan Penelitian

Tahapan penelitian ini menjelaskan langkah-langkah yang dilakukan dalam menganalisis dan menentukan strategi penjualan di coffee shop menggunakan metode Naive Bayes. Tahapan penelitian disusun secara sistematis agar hasil yang diperoleh sesuai dengan tujuan penelitian. Metodologi dalam penerapan algoritma ini meliputi langkah-

langkah berikut:

1. Pengumpulan Data

Mengumpulkan data penjualan dan data pelanggan coffee shop.

2. Preprocessing Data

Membersihkan dan menyiapkan data agar siap dianalisis.

3. Pembagian Data

Membagi data menjadi data latih dan data uji

4. Penerapan Naïve Bayes

Mengolah data untuk mengklasifikasikan strategi penjualan

5. Evaluasi Model

Mengukur akurasi hasil klasifikasi.

6. Kesimpulan

Menentukan strategi penjualan terbaik berdasarkan hasil analisis.

2.5 Penelitian Terdahulu

Penelitian sebelumnya menunjukkan bahwa algoritma Naive Bayes memiliki performa yang baik dalam berbagai aplikasi, termasuk analisis sentimen dan klasifikasi teks. Sebagai contoh:

2.1 Tabel Penelitian Terdahulu

No	Judul	Penulis	Hasil	Hubungan	Kekurangan
1	Analisis Strategi Penjualan Menggunakan Algoritma	Andi Pratama	Naive Bayes mampu mengklasifikasi data penjualan dengan tingkat	Menjadi dasar penerapan Naive Bayes dalam	Data penelitian terbatas pada satu periode waktu.

	Naïve Bayes		akurasi yang baik.	menentukan strategi penjualan coffee shop.	
2	Prediksi Minat Beli Pelanggan Menggunakan Metode Naïve Bayes	Siti Rahmawati	Metode Naive Bayes efektif dalam memprediksi minat beli pelanggan.	Relevan untuk menganalisis perilaku pembelian pelanggan coffee shop.	Tidak mempertimbangkan faktor promosi secara detail.
3	Analisis Kepuasan Pelanggan Menggunakan Naïve Bayes Classifier	Budi Santoso	Algoritma Naive Bayes mampu mengklasifikasi tingkat kepuasan pelanggan dengan baik	Mendukung analisis kepuasan pelanggan sebagai dasar strategi penjualan.	Data kuesioner bersifat subjektif.
4	Penerapan Data Mining Untuk Analisis Penjualan Retail	Dewi Lestari	Data mining dengan Naive Bayes dapat mengidentifikasi pola penjualan. Digunakan sebagai acuan analisis data penjualan coffee shop.	Digunakan sebagai acuan analisis data penjualan coffee shop	Tidak membahas evaluasi performa model secara mendalam.
5	Klasifikasi Produk Terlaris Menggunakan Metode Naïve	Rian Hidayat	Naive Bayes berhasil menentukan produk terlaris berdasarkan data penjualan.	Relevan untuk menentukan menu unggulan di coffee shop.	Dataset yang digunakan relatif kecil.

	Bayes				
6	Perbandingan Metode Naïve Bayes dan Decision Tree pada Data Penjualan	Ahmad Fauzi	Naive Bayes lebih cepat dan memiliki akurasi yang kompetitif.	Relevan untuk menentukan menu unggulan di coffee shop.	Metode perbandingan terbatas.
7	Analisis Sentimen Pelanggan Menggunakan Naïve Bayes	Fitriani	Naive Bayes memiliki akurasi yang baik dalam klasifikasi sentimen pelanggan.	Mendukung analisis sentimen pelanggan coffee shop	Tidak menggabungkan data numerik penjualan.
8	Analisis Pola Pembelian Konsumen Menggunakan Naïve Bayes	Lina Marlina	Pola pembelian konsumen dapat diklasifikasikan secara efektif.	Membantu memahami perilaku pelanggan coffee shop	Tidak membahas pengaruh waktu pembelian.
9	Penerapan Naïve Bayes dalam Sistem Pendukung Keputusan Penjualan	Dedi Kurniawan	Sistem mampu memberikan rekomendasi strategi penjualan berbasis data	Menjadi referensi sistem penentuan strategi penjualan.	Sistem belum diuji dalam jangka panjang.
10	Analisis Strategi Pemasaran Menggunakan Data Mining Naïve Bayes	Nur Aisyah	Strategi pemasaran dapat ditentukan berdasarkan hasil klasifikasi data.	Mendukung analisis strategi penjualan berbasis data pada coffee shop.	Tidak membahas loyalitas pelanggan secara spesifik.

2.6 Kelebihan dan Kekurangan Penelitian

1. Naive Bayes mampu menangani dataset besar dengan waktu komputasi yang relatif singkat.
2. Algoritma ini bekerja baik dengan data berdimensi tinggi karena proses perhitungannya yang sederhana.
3. Dapat diterapkan pada berbagai jenis data, baik kategorikal maupun numerik kekurangan
4. Asumsi ini sering tidak terpenuhi dalam data dunia nyata, sehingga dapat menurunkan akurasi prediksi.
5. Ketidakseimbangan dalam distribusi kelas dapat menyebabkan bias dalam hasil klasifikasi.
6. Tidak mempertimbangkan hubungan antar fitur secara mendalam.

2.7 Evaluasi Metode Naïve Bayes

Evaluasi metode Naive Bayes menggunakan confusion matrix adalah salah satu cara yang efektif untuk menilai kinerja model klasifikasi. Confusion matrix memberikan gambaran yang jelas tentang bagaimana model memprediksi kelas- kelas yang ada dalam dataset, dengan membandingkan prediksi model dengan label sebenarnya. Terdapat empat elemen utama dalam confusion matrix, yaitu True Positives (TP), False Positives (FP), True Negatives (TN), dan False Negatives (FN). Dari elemen-elemen ini, kita dapat menghitung berbagai metrik penting seperti

akurasi, precision, recall, dan F1-score, yang masing-masing memberikan informasi berbeda tentang kinerja model dalam mengklasifikasikan data. Penggunaan confusion matrix memudahkan untuk mengidentifikasi masalah pada model, seperti ketidakseimbangan kelas yang dapat memengaruhi akurasi dan interpretasi hasil

Tabel . 2.2. Evaluasi

Kelas Atribut	Kelas Prediksi		
	Kelas	Benar	Salah
	Benar	True Positive (TP)	False Positive (FP)
	Salah	False Negative (FN)	True Negative (TN)

Dimana tabel 1 berisi:

1. TP (True Positive), yaitu jumlah data positif yang memiliki nilai benar.
2. TN (True Negative), yaitu jumlah data negatif yang memiliki nilai benar.
3. FN (False Negative), yaitu jumlah data negatif tetapi yang memiliki nilai salah.
4. FP (False Positive), yaitu jumlah data yang positif tetapi yang memiliki nilai.

Naive Bayes dalam mengklasifikasikan data yang ada, serta mengungkapkan apakah model cenderung menghasilkan kesalahan prediksi pada salah satu kategori, seperti mengklasifikasikan orang yang puas sebagai tidak puas atau sebaliknya.