

BAB II

TINJAUAN PUSTAKA

Bab ini bertujuan untuk meninjau temuan dan konsep yang telah dipresentasikan sebelumnya, dengan tujuan untuk meningkatkan dan memberikan landasan yang kokoh bagi penelitian yang sedang dilakukan oleh peneliti.

2.1 Pencapaian Terdahulu

Penelitian terdahulu telah menunjukkan keberhasilan penerapan algoritma Naïve Bayes dalam berbagai domain klasifikasi dan sistem rekomendasi. Wijaya & Hajjah (2023) menerapkan algoritma ini dalam penentuan atribut genset—kapasitas, sistem starter, dan tegangan—dengan akurasi tinggi terhadap preferensi teknis pengguna. Pratama & Yulmaini (2018) mengembangkan sistem penentuan konsentrasi skripsi menggunakan empat belas atribut masukan, mencapai akurasi 96,55 % pada 29 data mahasiswa. Salma & Nugroho (2016) merancang sistem berbasis web untuk rekomendasi pemilihan SMA berdasarkan jarak, akreditasi, nilai UAN, dan biaya pendidikan menggunakan pendekatan probabilistik Naïve Bayes. Mahardhika et al. (2016) menggabungkan Naïve Bayes dan Cosine Similarity guna mengelompokkan keluhan pelanggan UPT Puskom UNS, dengan akurasi pelatihan 86,67 %–100 % meskipun hasil kluster tidak seragam pada data berfitur heterogen. Fikrillah & Kurniadi (2023) memanfaatkan Naïve Bayes untuk merekomendasikan program studi bagi siswa STT Garut; akurasi terbaik tercatat 73,4 % dari 90 data yang dianalisis.

Pada ranah e-commerce berskala butik daring independen, sejumlah studi memperlihatkan efektivitas pendekatan probabilistik ini. Situngkir et al. (2025) menerapkan Multinomial Naïve Bayes pada 921 ulasan produk MSI Official Store dan meraih akurasi uji 71 %, membuktikan kelayakan integrasi ke aplikasi web PHP sederhana. Hayuningtyas (2019) mengolah empat belas atribut pada ± 600 entri produk pakaian wanita dan memperoleh akurasi ≥ 80 %. Avianto et al. (2019) menggunakan tujuh atribut demografis dan perilaku untuk merekomendasikan baju

wanita; akurasi akhir juga berada di kisaran 80 %. Wulandari et al. (2022) meneliti sentimen pelanggan beberapa merek hijab di Shopee dan melaporkan F1-score 0,856, menunjukkan bahwa Naïve Bayes tetap kompetitif pada segmen fesyen muslim niche.

Berbeda dari model hybrid berbasis deep learning yang diimplementasikan pada marketplace berskala besar, penelitian ini memanfaatkan algoritma Naïve Bayes tunggal dengan empat belas atribut demografis dan perilaku pengguna, serta dataset $\pm 1\,500$ baris. Pendekatan terstruktur dan ringan ini memungkinkan personalisasi rekomendasi produk secara efisien tanpa kebutuhan infrastruktur komputasi kompleks, sehingga praktis diterapkan pada platform e-commerce yang masih berkembang.

2.2 Tinjauan Teoritis

Pada bab tinjauan teoritis, digunakan oleh peneliti untuk memaparkan teoriteori dan segala informasi yang terkait secara erat dalam penelitian, sehingga mendukung pemahaman yang lebih mendalam mengenai konteks dan implementasi dari penelitian yang dilakukan.

2.2.1 Fashion

Fashion merupakan bentuk representasi diri yang ditampilkan melalui pilihan pakaian, sepatu, aksesoris, dan gaya hidup. Dalam konteks digital, khususnya e-commerce, pemahaman atas preferensi fashion pengguna menjadi elemen penting dalam menciptakan pengalaman yang dipersonalisasi. Tren dalam dunia fashion sangat dinamis dan dipengaruhi oleh faktor usia, budaya, musim, dan arus media sosial. Oleh karena itu, sistem rekomendasi pada platform fashion online harus mampu beradaptasi terhadap perilaku konsumen yang berubah-ubah. Penelitian oleh Liu et al. (2021) menyimpulkan bahwa pendekatan berbasis analisis data perilaku pengguna mampu meningkatkan akurasi sistem rekomendasi real-time di industri fashion digital.

2.2.2 Website

Website adalah kumpulan halaman digital yang saling terhubung dan diakses melalui internet. Kontennya dapat berupa teks, gambar, audio, dan video. Website menjadi sarana interaktif utama untuk menyampaikan informasi. Lee Marvin & Humdiana (2021) menyatakan bahwa peran website sangat vital dalam mendistribusikan informasi di era digital.

2.2.3 E-commerce

E-commerce merujuk pada aktivitas jual beli produk atau layanan secara daring. Transaksi dapat dilakukan kapan saja melalui perangkat digital. Konsep mobile commerce berkembang seiring meningkatnya penggunaan ponsel. Menurut Lee Marvin & Humdiana (2021), e-commerce telah menjadi tulang punggung perdagangan modern.

2.2.4 Machine Learning

Machine learning adalah cabang dari kecerdasan buatan yang mendukung sistem belajar dari data untuk mengambil keputusan tanpa pemrograman eksplisit. Dalam sistem rekomendasi, algoritma pembelajaran diawasi meliputi *Naïve Bayes* digunakan untuk menganalisis perilaku pengguna. Amruthnath & Gupta (2020) menyatakan bahwa metode ini efektif untuk dataset kecil hingga menengah.

2.2.5 Algoritma Naïve Bayes

Naïve Bayes adalah metode klasifikasi berbasis Teorema Bayes dengan asumsi independensi bersyarat di antara fitur-fiturnya. Keunggulannya terletak pada kompleksitas komputasi yang rendah, kemudahan interpretasi hasil, serta kemampuannya beradaptasi pada data berukuran kecil maupun besar (Amzil et al., 2025; Capote, 2023). Persamaan umum yang digunakan dituliskan sebagai berikut

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

Dimana :

- $P(C|X)$ probabilitas kelas C diberikan vektor fitur X.
- $P(X|C)$ peluang kemunculan fitur X apabila kelas C diketahui.
- $P(C)$ prior atau probabilitas awal kelas C.
- $P(X)$ total probabilitas dari kombinasi fitur X.

Contoh Perhitungan (Data Pemisalan Netral)

Sebagai ilustrasi, misalkan sistem ingin memutuskan apakah aktivitas luar ruangan disarankan (Ya/Tidak) berdasarkan tiga atribut: Cuaca, Suhu, dan Kelembapan. Tabel 2.1 menyajikan empat baris data latih pemisalan (disederhanakan dari 14 baris) guna memudahkan penjabaran manual.

Tabel 2.1 Contoh Data Perhitungan Naive Bayes

No	Cuaca	Suhu	Kelembapan	Aktivitas
1	Cerah	Hangat	Normal	Ya
2	Cerah	Hangat	Tinggi	Ya
3	Hujan	Dingin	Tinggi	Tidak
4	Mendung	Hangat	Normal	Ya

Dari tabel tersebut diperoleh prior

$$P(\text{Ya}) = \frac{3}{4} = 0,75 \quad \text{dan} \quad P(\text{Tidak}) = \frac{1}{4} = 0,25.$$

Kasus uji: Cuaca = Cerah, Suhu = Hangat, Kelembapan = Normal.

Apakah aktivitas luar ruangan disarankan?

Perhitungan likelihood (dengan Laplace smoothing $\alpha=1$ untuk menghindari probabilitas nol) menghasilkan

$$\begin{aligned}
P(\text{Cerah} \mid Y_a) &= \frac{2+1}{3+|V|} = \frac{3}{5}, \\
P(\text{Hangat} \mid Y_a) &= \frac{3+1}{3+|V|} = \frac{4}{5}, \\
P(\text{Normal} \mid Y_a) &= \frac{2+1}{3+|V|} = \frac{3}{5}, \\
P(\text{Cerah} \mid \text{Tidak}) &= \frac{0+1}{1+|V|} = \frac{1}{5}, \\
P(\text{Hangat} \mid \text{Tidak}) &= \frac{0+1}{1+|V|} = \frac{1}{5}, \\
P(\text{Normal} \mid \text{Tidak}) &= \frac{0+1}{1+|V|} = \frac{1}{5},
\end{aligned}$$

dengan $|V|=3$ (banyaknya nilai unik tiap atribut).

Probabilitas posterior (konstanta pembagi $P(X)$ diabaikan karena sama untuk kedua kelas):

$$\begin{aligned}
P(Y_a \mid X) &\propto 0,75 \times \frac{3}{5} \times \frac{4}{5} \times \frac{3}{5} = 0,108, \\
P(\text{Tidak} \mid X) &\propto 0,25 \times \frac{1}{5} \times \frac{1}{5} \times \frac{1}{5} = 0,002.
\end{aligned}$$

Karena $P(Y_a|X) > P(\text{Tidak}|X)$, sistem **merekomendasikan** aktivitas luar ruangan.

2.2.6 Evaluasi data: Confusion matrix

Confusion matrix digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan hasil prediksi dan data aktual. Empat elemen utamanya: True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN). Dari elemen tersebut, metrik evaluasi meliputi akurasi, presisi, recall, dan F1-score dapat dihitung. Kaur & Singh (2021) menegaskan bahwa evaluasi meliputi ini penting untuk memastikan bahwa sistem rekomendasi yang dibangun memiliki akurasi dan konsistensi yang tinggi.

Contoh Confusion Matrix:

Sistem diuji dengan 20 data dan menghasilkan:

Tabel 2.2 Contoh Perhitungan Confusion Matrix

Kelas Aktual	Prediksi Kasual	Prediksi Formal
--------------	-----------------	-----------------

Kasual	8 (TP)	2 (FN)
Formal	1 (FP)	9 (TN)

Dari tabel tersebut:

- $\text{Akurasi} = (\text{TP} + \text{TN}) / \text{Total} = (8 + 9) / 20 = 85\%$
- $\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) = 8 / (8 + 1) = 88.9\%$
- $\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) = 8 / (8 + 2) = 80\%$
- $\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) = 84.2\%$

Evaluasi ini menunjukkan bahwa sistem memiliki performa klasifikasi yang baik, dengan tingkat akurasi 84.2% dan keseimbangan antara presisi serta sensitivitas yang mendekati 84.2%.