

BAB II

TINJAUAN PUSTAKA

2.1 Teori Dasar

Teori dasar adalah teori-teori yang bersumber dari jurnal dan buku yang dijadikan sebagai referensi untuk mendukung penelitian. Beberapa teori yang digunakan dalam penelitian ini:

2.1.1 *E-commerce*

E-commerce singkatan dari *elektronik commerce*, yang diartikan sebagai sistem pembelanjaan yang dilakukan secara elektronik atau media *elektronik*. *E-commerce* juga adalah proses pembelian dan penjualan produk, jasa dan informasi melalui jaringan komputer atau jaringan internet (Kolbia and Dahliyanti 2024). Perkembangan teknologi informasi dan komunikasi yang sangat cepat telah mengubah konsumen dan pelaku usaha berinteraksi, Dimana *platform online* mengubah arah untuk memenuhi kebutuhan sehari-hari. Salah satu daya tarik *e-commerce* ini adalah kemudahan akses untuk mencari barang yang diinginkan, dan tanpa batas waktu, lokasi yang memberikan fleksibel dari pengguna untuk berbelanja kapan dan Dimana saja.

Salah satu *platform e-commerce* yang berkembang pesat saat ini adalah Tokopedia. Tokopedia menjadi salah satu yang sukses dalam memanfaatkan potensi *platform e-commerce*. Dibantu dengan teknologi yang canggih dan strategi inovatif, Tokopedia memainkan peran yang signifikan dalam mendorong pertumbuhan

ekonomi digital Indonesia, kehadiran Tokopedia ini tidak hanya untuk mempengaruhi cara konsumen berbelanja, tetapi juga memberikan dampak bagi UKM untuk bersaing dipasar global yang semakin maju saat ini (Kolbia and Dahliyanti 2024). Tokopedia juga menyediakan fitur untuk mempermudah pengalaman berbelanja online, mulai dari pencarian produk, hingga sistem pembayaran dan pengiriman yang terintegrasi. Rekomendasi berperan sebagai alat bantu untuk terhubung dengan pengguna dalam menemukan produk yang relevan, tujuannya untuk meningkatkan kenyamanan pengguna sekaligus mendorong konversi penjual. Sistem rekomendasi ini diterapkan dalam *e-commerce* seperti Tokopedia secara signifikan mempengaruhi Keputusan pembelian konsumen, terutama dalam kategori produk yang memiliki banyak varian dan pilihan.

2.1.2 Perilaku pengguna *e-commerce*

Perilaku pengguna adalah suatu Tindakan dari seseorang yang dapat secara langsung untuk melakukan pencarian, membeli, mengevaluasi, menggunakan, dan membelanjakan produk dan jasa yang di minatkan untuk memenuhi kebutuhan nya (Siregar 2023). Perilaku pengguna *e-commerce* ini ada dua yang terlibat, penjual dan konsumen. Dimana penjual memanfaatkan perkembangan *e-commerce* salah satunya Tokopedia, dengan menjual barang melalui Tokopedia maka konsumen dapat melihat dan membeli barang sesuai kebutuhannya. Sedangkan konsumen adalah seseorang yang terlibat dalam suatu kegiatan pembelian ataupun meriview produk yang di pasarkan di dalam platform tersebut.

1. Konsumen

Konsemen adalah individua tau kelompok yang membeli, menggunakan, atau mengonsumsi barang dan/atau jasa untuk memenuhi kebutuhan dan keinginan nya. Dengan ringkasan pemesanan mendorong konsemen untuk berbelanja secara online dan produknya banyak pilihan, harga murah, pengiriman cepat, promo yang menarik, respon penjual bagus dan review produk sangat bagus sehingga membuat konsumen tertarik untuk berbelanja Perilaku konsumen seringkali berbelanja di saat banyak promo seperti tanggal yang sama dengan bulan biasanya banyak promo gratis ongkir dan diskon membuat konsumen tertarik untuk berbelanja.

2. Penjual

Penjual adalah individua tau pihak badan usaha maupun perorangan yang menawarkan barang atau jasa Kepada konsumen dengan tujuan untuk mendapatkan keuntungan. Penjual memiliki peran penting untuk menarik konsumen agar produk yang di pasarkan dibeli, dengan menampilkan foto produk yang jelas, memberikan informasi yang jelas pada produk mengikuti program promosi, diskon, strategi harga, dan membangun reputasi toko melalui ulasan dan reteng.

2.1.3 Tokopedia

Tokopedia adalah platform *e-commerce* di indonesia dengan model bisnis marketplace dan menjadikan tempat jual beli produk. Tokopedia memberikan tempat bagi para penjual untuk memasarkan produknya dan menyediakan juga fasilitas bagi para pembeli untuk menemukan produk yang diminatin sesuai kebutuhannya (Nasywa Zahira Ramadhani et al. 2023). Hadirnya aplikasi Tokopedia ini memberikan kemudahan bagi masyarakat Indonesia untuk

mendapatkan barang atau jasa sesuai dengan yang di inginkan. Aplikasi Tokopedia memberikan fitur untuk memudahkan pelanggan untuk mencaari barang dan sudah bekerja sama dengan jasa pengiriman seperti *JNE*, *J&T*, dan lain sebagainya, termasuk keaaman data pribadi saat melakukan transaksi sangat aman.

Kepercayaan konsumen terhadap Tokopedia berbentuk melalui pengalaman positif dalam transaksi dan citra merek yang baik. Dalam hal ini bahwa kualitas produk dan promosi memiliki pengaruh positif terhadap Keputusan pembelian konsumen di Tokopedia. Konsumen dimudahkan dalam pencarian produk, perbandingan harga, dan membaca ulasan sebelum membeli dan penjual mendapat akses ke jutaan pembeli serta alat bantu digital untuk mengelola bisnis mereka.

2.2 *Machine learning*

Machine learning atau pembelajaran mesin adalah bagian dari kecerdasan buatan/*Artificial Intelligence* (AI) yang di kembangkan sistem komputer dapat meniru perilaku manusia, dan dapat memproses Bahasa alami, pengenalan suara, presepsi visual, dan pengambilan Keputusan (Boukrouh and Azmani 2025).

Kecerdasan buatan (AI) yaitu bidang ilmu komputer yang berfokus untuk pembuatan agen cerdas, yang sistem dapat merasakan lingkungan, belajar, mengambil Keputusan, dan mengambil tindakan agar dapat mencapai tujuan tertentu. AI ini bertujuan untuk membuat mesin dapat beradaptasi, berpikir, dan bertindak seperti hal nya dengan manusia (Kolbia and Dahliyanti 2024). Dalam rekomendasi produk *e-commerce* pada Tokopedia, kecerdasan buatan (AI) dapat dimanfaatkan untuk menemukan produk yang relevan berdasarkan hasil dari

preferensi dan pengguna sebelumnya serta memberikan peningkatan pengalaman berbelanja secara daring terhadap pengguna baru.

Machine learning memungkinkan mesin belajar dari data sebelum nya (data historis) untuk melakukan perintah tertentu. Ada tiga bentuk algoritma dalam *machine learning* yaitu pengajaran yang diawasi, pengajaran yang tidak diawasi, dan penajaran pendukung. Setiap bentuk algoritma mempunyai prosedur pengajaran yang berbeda-beda. Proses pembelajaran ini mencakup pengumpulan data, ekstraksi data, pemilihan model atau algoritma pembelaran, pelatihan model atau algoritma tersebut, dan evaluasi hasil pembelajaran mesin (Kolbia and Dahliyanti 2024). Beberapa keleihan dari *machine learning* adalah:

1. Otomatisasi yaitu dapat melakukan hal-hal yang secara otomatis tanpa bantuan dari manusia. *Machine learning* ini memiliki kemampuan untuk mempelajari pola tertentu, hingga menyerupai manusia dan berkembang tanpa pengkodean.
2. Peningkatan akurasi yaitu Dimana pengguna data yang lebih besar dan tersusun rapi akan membuat hasil yang lebih akurat. Dalam *machine learning* ini akan membantu berbagai industri, medis, keuangan dan lain-lain yang dapat meningkatkan akurasi pengambilan Keputusan yang di proses pada pembelajaran sebelumnya.
3. Pengguna data yang banyak akan menangani data yang sangat besar dan kompleks. *Machine learning* dimanfaatkan dalam Perindustrian, analisis data, pengembangan produk, dan lain-lainnya.

4. Meningkatkan efesiensi, *machine learning* ini dapat menganani tugas seperti pengolahan data, pengembangan produk, dan lainnya, maka bermanfaat untuk berbagai industri seperti teknologi, keuangan dan yang lain nya.

Beberapa algoritma machine learning yang sering digunakan dalam mengklasifikasi e-commerce.

2.2.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma supervised learning yang digunakan untuk klasifikasi maupun regresi. SVM bekerja dengan mencari hyperplane terbaik yang memisahkan data ke dalam kelas berbeda dengan margin maksimal (Zhang et al. (2020). Konsep utamanya adalah memaksimalkan jarak antara support vectors (data terdekat dengan hyperplane) sehingga model lebih robust terhadap kesalahan klasifikasi. Kelebihan SVM adalah kemampuannya menangani data berdimensi tinggi, seperti teks pada ulasan produk e-commerce. SVM dapat digunakan untuk mengklasifikasikan ulasan produk menjadi positif, negatif, atau netral. Dengan menggunakan representasi teks berbasis TF-IDF, SVM dapat memisahkan ulasan relevan dan tidak relevan dengan akurasi yang tinggi. Misalnya, ulasan pelanggan yang berisi kata-kata seperti “*cepat, bagus, berkualitas*” akan dikategorikan sebagai ulasan positif/relevan, sementara kata-kata seperti “*rusak, lambat, mengecewakan*” dikategorikan sebagai ulasan negatif/tidak relevan.

2.2.2 Logistic Regression

Logistic Regression merupakan algoritma *supervised learning* yang digunakan untuk memprediksi probabilitas suatu kejadian dalam bentuk kelas biner maupun multikelas. Algoritma ini menggunakan fungsi logit atau sigmoid untuk mengubah output linear menjadi nilai probabilitas (Hosmer et al., 2019). Meskipun sederhana, logistic regression efektif digunakan dalam klasifikasi teks, termasuk analisis sentimen ulasan produk.

2.2.3 Decesion Tree

Decision Tree adalah algoritma klasifikasi yang berbentuk struktur pohon, di mana setiap simpul (node) merepresentasikan suatu atribut, cabang merepresentasikan keputusan, dan daun (leaf node) menunjukkan kelas prediksi. Algoritma ini sederhana, mudah dipahami, serta mampu menangani data kategorikal maupun numerik (Han et al., 2022). Kelebihannya adalah interpretabilitas, sedangkan kelemahannya adalah rentan terhadap overfitting.

2.2.4 Random Forest

Random Forest merupakan algoritma ensemble learning yang membangun banyak pohon keputusan (decision trees) kemudian menggabungkan hasil prediksi setiap pohon untuk menghasilkan keputusan akhir (Nuraini & Hidayat (2022). Konsep bagging yang digunakan Random Forest membuat algoritma ini lebih tahan terhadap overfitting dibanding Decision Tree tunggal.

2.2.5 *Naive Bayes*

Naive Bayes adalah algoritma *machine learning* yang digunakan untuk klasifikasi. Metode ini didasarkan pada teorema bayes, yang digunakan untuk menghitung probabilitas suatu kejadian yang berdasarkan informasi sebelumnya. (Hermiani et al. 2023). Disebut *naive* karena metode ini mengasumsi setiap fitur (atribut) yang bersifat independent satu sama lain, yang kenyataannya sering tidak benar, tetapi selalu memberikan hasil yang baik dalam banyak kasus.

Berikut adalah Langkah-langkah pada metode *Naive Bayes*:

1. Baca data training: Langkah awal adalah membaca data training, adalah data yang digunakan untuk melatih model *Naive Bayes*. Data training terdiri dari lebel, dan kategori.
2. Hitung jumlah dan probabilitas: setelah membaca data training, Langkah selanjutnya yaitu menghitung jumlah data pada masing-masing kategori atau lebel dan menghitung probabilitas dari masing-masing kategori. Probabilitas ini melanjutkan seberapa sering data muncul dalam kategori tertentu dalam data training
3. Data numerik: selanjutnya jika data yang digunakan adalah data numerik, maka akan dicari nilai mean (rata-rata) dan standar deviasi (deviasi standar) dari setiap parameter yang merupakan data numerik.

Menurut (Fatiah Rahmadanti, Jajuli, and Purnamasari 2021) Algoritma *Naive Bayes* adalah salah satu pengklasifikasi statistik yang digunakan untuk memprediksi sebuah kelas. Pengklasifikasi dilakukan perhitungan nilai probabilitas dari

setiap lebel kelas yang ada terhadap masukan yang diberikan lebel kelas diambil dari nilai probabilitas yang sangat besar. *Naive bayes* terdapat kelebihan dalam perhitungan yang cepat, berakursi tinggi dan algoritma yang sederhana. Algoritma *naive bayes* mudah digunakan karena memiliki perhitungan yang tidak panjang (Putro, Vlandari, and Saptomo 2020).

Rumus *Naive Bayes*

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

Rumus 1 *Naive Bayes*

Sumber: (Putro, Vlandari, and Saptomo 2020)

Keterangan:

- $P(H|X)$: probabilitas hipotesis H (misalnya, kelas tertentu) diberikan data X (*posterior probability*).
- $P(H)$: probabilitas awal dari hipotesis H sebelum melihat data (*prior probability*)
- $P(X|H)$: probabilitas data X diberikan hipotesis H (*likelihood*)
- $P(X)$: probabilitas data X (*evidence*)

2.2.6 Klasifikasi

Klasifikasi adalah proses yang mencari model atau fungsi yang dapat membedakan dan menjelaskan objek atau kelas-kelas data berdasarkan atribut atau fitur-fitur yang dimilikinya. Model atau fungsi tersebut kemudian digunakan untuk memprediksi lebel atau kelas dari objek yang lebel kelasnya tidak di ketahuai.

Tujuan nya adalah mengelompokkan data ke dalam kelas atau kategori tertentu berdasarkan pola-pola yang terdapat dalam data pelatihan. Algoritma ini mencari pola atau relasi dalam data untuk membuat Keputusan tentang kategori mana yang paling cocok untuk data baru.

Klasifikasi mencakup pengenalan tulis tangan, analisis sentiment, deteksi penyakit melalui gambar medis pemisahan spam pada email dan prediksi harga saham. Dalam klasifikasi ini menggunakan algoritma naive bayes untuk memprediksi produk *e-commerce* Tokopedia (Pangestu, Novita, and Mustakim 2023).

2.2.7 Confusion matrix

Menurut (K. S. Putri, Setiawan, and Pambudi 2023) *Confusion matrix* bertujuan untuk mengetahui performa algoritma *naive bayes* dalam mengklasifikasi produk apakah relevan atau tidak relevan bagi pengguna. Matrix ini memberikan gambaran keseluruhan performa model dengan menampilkan TN, TP, FN dan FP pada konteks klasifikasi rekomendasi produk, confusion matrix dapat di jelaskan dalam tabel dibawah ini.

Tabel 2. 1 Confusion Matrix

TP	FP
----	----

FN	TN

Sumber: Penelitian 2025

Penjelasan:

1. TP (*True Positive*): Produk yang memang relevan dan berhasil di prediksi sebagai relevan
2. TN (*True Negative*): produk yang tidak relevan dan berhasil di prediksi tidak relevan
3. FP (*Fals Posotive*): produk yang tidak relevan tetapi salah diklasifikasi sebagai relevan
4. FN (*Fals Negative*): produk yang relevan tatapi salah diklasifikasi sebagai tidak relevan

2.2.8 Evaluasi Model

Dari *confusion matrix* diatas, maka bisa di hitung dengan matrix evaluasi seperti akurasi, presisi, recall, dan F1-Score. Berikut penjelasan evaluasi matrix di bawah ini.

1. Akurasi: Akurasi total dari presentase prediksi yang benar

Rumus dari Akurasi:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

Rumus 2 Akurasi

2. Presisi: Presisi mengukur prediksi relevan yang benar-banar relevan

Rumus dari Presisi:

$$Presisi = \frac{TP}{TP + FP}$$

Rumus 3 presisi

3. Recall: Recal presentasi produk relevan yang berhasil direkomendasikan

Rumus dari Recall:

$$Recal = \frac{TP}{TP + FN}$$

Rumus 4 Recall

4. F1_Score: F1_Score adalah rata-rata harmonis antara presisi dan recall.

Rumus dari F1_Score:

$$F1\ Score = 2x \frac{Presisi \times Recall}{Presisi + Recal}$$

Rumus 5 F1_Score

2.3 Software pendukung

Software pendukung adalah perangkat lunak yang digunakan pada penelitian ini yang mendukung proses penyelesaian penelitian. Ada beberapa perangkat lunak yang digunakan di antaranya:

2.3.1 Jupyter Notebook

Jupyter notebook adalah aplikasi berbasis web yang digunakan untuk membuat dan berbagi dokumen yang berisi kode program, visualisasi, dan teks naratif. Dalam konteks pemrograman Python, Jupyter sangat populer untuk analisis

data, pembelajaran mesin, pendidikan, dan eksperimen ilmiah karena memungkinkan pengguna menulis dan menjalankan kode Python dalam sel interaktif

Python adalah bahasa pemrograman yang dapat digunakan untuk tujuan umum, khususnya untuk membuat kode sumber yang mudah dibaca. Selain itu, Python memiliki library lengkap yang memungkinkan programmer untuk membuat aplikasi yang mutakhir menggunakan kode sumber yang tampak sederhana (Hasan, Abdullah, and Ambarita 2024).



Gambar 2. 1 *Jupyter Notebook*

Sumber: (Hasan, Abdullah, and Ambarita 2024)

2.3.2 *Microsoft Excel*

Microsoft excel adalah bagian dari *Microsoft office* yang bertugas untuk menghitung yang berup angka dan mengelolah data. Excel ini digunakan untuk

menganalisis data, dapat menghitung secara matematika dan menyajikan data dalam bentuk table, grafik, dan laporan.

Pada penelitian ini digunakan *software excel* untuk pengumpulan data, dan akan diolah dengan menggunakan aplikasi *jupyter*. *Microsoft excel* merupakan aplikasi bagian dari *Microsoft office*, yang memiliki berbagai fungsi sebagai pengolahan bentuk angka menggunakan *spreadsheet* yang terdiri dari baris dan kolom untuk mengeksekusi perintah (Baharuddin and Anggraini 2021).



Gambar 2. 2 *Microsoft Excel*

Sumber: (Baharuddin and Anggraini 2021)

2.4 Penelitian terdahulu

Penelitian terdahulu adalah penelitian yang telah dilakukan oleh peneliti sebelum nya dan kemudian di jadikan sebagai referensi pada penelitian ini.

1. Dari jurnal (Boukrouh and Azmani 2025) “Model Pembelajaran Mesin Yang Dapat Di Terapkan Untuk Memprediksi Peputaran Pelanggan Untuk E-Commerce” *IAES International Journal of Artificial Intelligence* 14(1):286–97. Memberikan kesimpulan bahwa model pembelajaran masin klasifikasi

untuk memprediksi churn pelanggan dan mengeksplorasi penerapan XAI dalam sektor e-commerce.

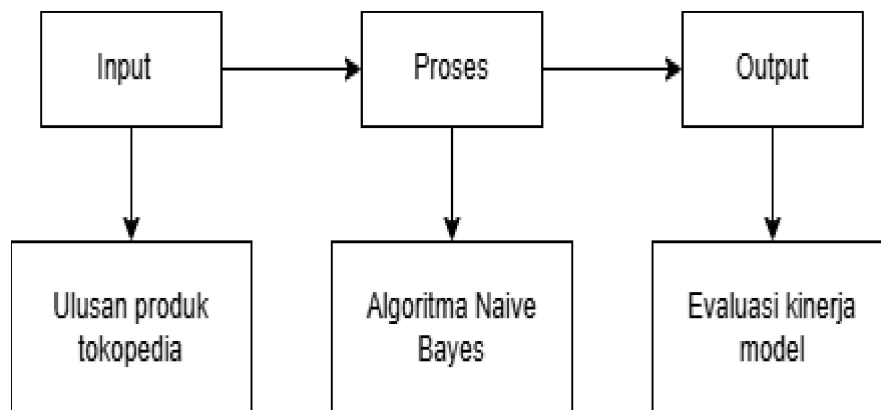
2. Dari jurnal penelitian Putro, Hakam Febtadianrano, Retno Tri Vulandari, and Wawan Laksito Yuly Saptomo. 2020. “Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan.” *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)* 8 (2). Menarik kesimpulan bahwa penerapan metode naïve bayes untuk klasifikasi pelanggan dapat membantu pemilik memberikan tambahan terhadap pelanggan yang jujur dan meningkatkan kualitas yang sempurna terhadap pelanggan
3. Dari jurnal penelitian (Sumual et al., n.d.) dengan judul “Klasifikasi Kategori Produk Terlaris Pada E-Commerce.” 5(2):51–60. Menarik kesimpulan bahwa algoritma naïve bayes dapat dipakai untuk penelitian klasifikasi produk terlaris dengan menghasilkan nilai accuracy jumlah kelas status yang laris dan tidak laris yang diprediksi dengan benar adalah 92,64% atau setara dengan 0,9264 yang termasuk kedalam kategori klasifikasi excellent.
4. Jurnal dari (Saputro and Utomo 2024). Dengan judul “Rekomendasi Produk E-Commerce Berbasis Klasifikasi Ulasan Menggunakan Ensemble Random Forest Dan Teknik Boosting.” Memberikan kesimpulan rekomendasi produk berbasis klasifikasi dengan ensemble random forest dengan xgboost memberikan nilai akurasi tertinggi atau terbaik dibandingkan dengan ensemble random forest-GB dan ensemble random forest-LightGBM.
5. Dari jurnal (Fatiah Rahmadanti, Tania, Mohamad Jajuli, and Intan Purnamasari. 2021). “Klasifikasi Pengguna Shopee Berdasarkan Promosi

Menggunakan Naïve Bayes.” *Generation Journal* 5 (2): 81–90. Dengan memberikan kesimpulan bahwa hasil perhitungan klasifikasi menggunakan algoritma naïve bayes promosi free ongkir yaitu promosi yang selalu di tunggu oleh konsumen untuk membeli barang yang di inginkan.

6. Dari jurnal (Putri, Kaswili Sriwenda, Iwan Rizal Setiawan, and Agung Pambudi. 2023). “Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan Naïve Bayes Classifier.” *Technologia : Jurnal Ilmiah* 14 (3): 227.
7. Dari jurnal (Hermiani, Monika, Yolanda Simamora, Ida Bagus, Made Mahendra A2 A. 2023). “Analisis Ulasan Produk Menggunakan Metode Naive Bayes Classifier.” *Jnatia* 2 (1): 177–84.

2.5 Kerangka pikir

Kerangka pikir adalah Gambaran penelitian dalam bentuk kerangka yang telah dibuat oleh peneliti dari awal sampai akhir, sehingga dapat membantu peneliti dalam menyelesaikan penelitian tepat sasaran.



Gambar 2. 3 Kerangka pikir
Sumber : Penelitian 2025

1. Input Ulasan Produk Tokopedia

Inputan dalam pengolahan data ini adalah ulasan produk Tokopedia sebagai data mentah Dimana terdiri dari ulasan, reting, kategori, nama produk, id_produk, yang terjual, id_toko dan produk_url. Akan menjadi sumber utama untuk klasifikasi

2. Proses Algoritma Naive Bayes

Pada tahap ini, data ulasan produk Tokopedia akan di proses menggunakan algoritma naïve bayes yang dapat mengklasifikasi berbasisi probabilitas. Beberapa langkah untuk proses algoritma naïve bayes yaitu preprocessing teks, ekstraksi fitur, pelatihan model dan prediksi kategori.

3. Output Evaluasi Kinerja Model

Output ini akan menghasilkan akhir dari sistem evaluasi performa model matrix seperti akurasi, presisi, recall, F1_score dan *confusion matrix*.