

BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurut APJII dalam *Survey Internet Indonesia 2023*, pengguna *e-commerce* di Indonesia sudah mencapai 77,02% dari total populasi, diperkirakan pengguna *e-commerce* akan terus bertambah secara signifikan hingga tahun 2025. Perkembangan *e-commerce* di Indonesia telah menjadi salah satu katalis utama dalam transformasi digital nasional yang berdampak luas pada berbagai sektor, khususnya industri teknologi. Fenomena ini tidak hanya mencerminkan perubahan pola konsumsi masyarakat yang kian mengandalkan platform digital untuk memenuhi kebutuhan sehari-hari, tetapi juga menandai pergeseran struktural dalam sistem distribusi dan mekanisme pasar. Dengan semakin meluasnya akses terhadap internet dan penetrasi perangkat mobile yang tinggi, *e-commerce* telah membuka jalan bagi pelaku usaha dari berbagai skala, terutama Usaha Mikro, Kecil, dan Menengah (UMKM), keberadaan platform *e-commerce* turut mendorong inovasi dalam model bisnis, sistem logistik, dan metode pembayaran digital, yang semuanya berkontribusi terhadap peningkatan efisiensi operasional dan perluasan jangkauan layanan. Kolaborasi antara pemerintah, penyedia layanan digital, dan sektor swasta juga memainkan peran penting dalam menciptakan ekosistem ekonomi digital yang berkelanjutan dan adaptif terhadap perubahan teknologi (Abdan Sifa et al., 2024).

Perkembangan *e-commerce* di Indonesia dalam beberapa tahun terakhir menunjukkan pertumbuhan yang sangat pesat dan menjadi salah satu pilar utama dalam transformasi ekonomi digital nasional. Shopee sebagai salah satu platform *e-commerce* terbesar di Asia Tenggara khususnya di Indonesia, berhasil menempati posisi dominan dengan mencapai lebih dari 237 juta kunjungan pengguna pada September 2023, mengalami peningkatan sebesar 38% dibandingkan awal tahun yang sama

(Inayah et al., 2024). Data ini mencerminkan tingginya minat dan kepercayaan masyarakat Indonesia terhadap Shopee sebagai media belanja daring yang menyediakan beragam produk serta layanan yang inovatif, seperti pembelian barang, pulsa, dan layanan *ShopeeFood* (Inayah et al., 2024).

Ulasan pengguna yang dihasilkan secara daring telah menjadi salah satu sumber informasi yang paling berpengaruh dalam proses pengambilan keputusan konsumen, khususnya dalam ranah perdagangan elektronik. Kegunaan informasi yang terkandung dalam ulasan pengguna sangat dipengaruhi oleh kualitas, relevansi, dan keakuratan konten yang disampaikan. Konsumen tidak hanya mencari data faktual mengenai produk, tetapi juga menghargai pengalaman dan opini yang jujur yang mampu memberikan gambaran menyeluruh tentang produk atau layanan tersebut. Informasi yang dianggap berguna tersebut pada akhirnya menjadi landasan bagi pembentukan persepsi konsumen terhadap merek serta memengaruhi niat pembelian secara signifikan (Coursaris et al., 2018).

Ulasan pengguna tidak hanya berperan sebagai *feedback* biasa, melainkan juga sebagai sumber data yang kaya untuk pengembangan teknologi pencarian dan rekomendasi produk di platform *e-commerce*. Dengan mengolah data ulasan tersebut secara cermat, sistem dapat menghasilkan rekomendasi yang lebih personal dan relevan, sekaligus meningkatkan pengalaman belanja secara keseluruhan. Hal ini memperlihatkan betapa pentingnya pengelolaan data ulasan secara optimal untuk menunjang interaksi konsumen dengan platform (Lien et al., 2022).

Konsumen cenderung menaruh kepercayaan yang lebih tinggi pada opini sesama pengguna dibandingkan dengan ulasan yang diberikan oleh ahli atau pakar. Kepercayaan ini meningkat seiring dengan jumlah dan variasi ulasan yang tersedia, yang memberikan gambaran lebih autentik mengenai pengalaman nyata pengguna produk. Fenomena ini menegaskan posisi sentral ulasan pengguna sebagai media komunikasi yang efektif dalam membentuk persepsi pasar dan memandu perilaku pembelian

(Ackerman et al., 2024). Ulasan pengguna tidak hanya berfungsi sebagai mekanisme *feedback* yang sederhana, tetapi juga merupakan aset strategis bagi pelaku bisnis *e-commerce* dalam memperbaiki kualitas produk dan layanan. Pengelolaan ulasan yang efektif dapat meningkatkan kepuasan pelanggan, membangun kepercayaan, serta menciptakan loyalitas konsumen jangka panjang.

Pertumbuhan volume ulasan pengguna di platform *e-commerce*, termasuk Shopee, menghadirkan tantangan besar dalam pengelolaan dan pemanfaatan data tersebut secara efektif. Aplikasi populer mampu menerima ribuan hingga jutaan ulasan setiap harinya, dimana sebagian besar ulasan mengandung informasi yang beragam dan terkadang ambigu. Proses analisis manual terhadap ulasan-ulasan ini tidak hanya membutuhkan sumber daya manusia yang besar dan waktu yang lama, tetapi juga rentan terhadap bias subjektif serta inkonsistensi dalam interpretasi data. Hal ini tentu saja menjadi kendala signifikan bagi perusahaan yang ingin merespons masukan pelanggan secara cepat dan akurat untuk meningkatkan kualitas produk dan layanan (Dhinakaran, 2019).

Beban kerja yang tinggi dalam meninjau ulasan secara manual seringkali menyebabkan lambatnya proses pengolahan *feedback* pelanggan, yang berdampak pada menurunnya tingkat kepuasan dan loyalitas konsumen. Dalam konteks ini, keterbatasan analisis manual menjadi penghambat utama dalam memaksimalkan potensi data ulasan sebagai alat pengambilan keputusan strategis. Oleh karena itu, perlunya metode otomatisasi dalam menganalisis sentimen menjadi sangat mendesak, terutama metode yang dapat mengolah data dalam jumlah besar dengan akurasi yang tinggi dan waktu proses yang singkat (Ernawati et al., 2023).

Pendekatan berbasis machine learning menawarkan solusi efektif untuk mengatasi keterbatasan tersebut. Dengan algoritma seperti *Naïve Bayes* dan *Support Vector Machine*, serta model *deep learning* seperti *BERT*, analisis sentimen dapat dilakukan secara otomatis dengan tingkat ketelitian yang meningkat seiring dengan proses pelatihan model.

Penggunaan teknologi ini tidak hanya mempercepat proses analisis ulasan, tetapi juga mampu menghasilkan wawasan yang lebih objektif dan sistematis. Dengan demikian, pelaku bisnis dapat lebih cepat menanggapi kebutuhan dan keluhan pelanggan, sekaligus meningkatkan kualitas layanan dan daya saing di pasar *e-commerce* yang semakin kompetitif (Ernawati et al., 2023). Meningkatnya volume ulasan pengguna yang sulit dikelola secara manual menjadi landasan penting dalam penelitian yang bertujuan mengoptimalkan analisis sentimen dengan metode *machine learning* dan *deep learning*. Penelitian ini tidak hanya memberikan kontribusi dalam ranah akademik, tetapi juga relevan secara praktis untuk mendukung pengambilan keputusan bisnis yang berbasis *data real-time* dan meningkatkan pengalaman konsumen secara keseluruhan.

Teknologi analisis sentimen telah berkembang menjadi salah satu solusi utama dalam mengelola dan memahami opini konsumen yang tersebar luas di platform *e-commerce*. Dengan meningkatnya jumlah ulasan pengguna yang beragam dan berjumlah sangat besar, pendekatan manual untuk mengkategorikan sentimen menjadi tidak praktis dan kurang efisien. Analisis sentimen memanfaatkan algoritma pembelajaran mesin yang mampu mengklasifikasikan teks ulasan secara otomatis ke dalam kategori sentimen positif, negatif, atau netral. Proses otomatis ini memungkinkan pengolahan data dalam skala besar dengan kecepatan dan konsistensi yang tinggi, sekaligus meminimalisasi kesalahan manusia yang mungkin terjadi dalam analisis manual (Indrani et al., 2024). Peran penting teknologi otomatisasi seperti *Robotic Process Automation (RPA)* dan kecerdasan buatan dalam meningkatkan akurasi dan efektivitas analisis sentimen, khususnya pada data yang bersifat tidak terstruktur dan mengandung bahasa alami yang kompleks. Teknologi ini mampu menangkap nuansa emosional dan konteks bahasa yang seringkali sulit diinterpretasikan secara manual, sehingga menghasilkan klasifikasi yang lebih tepat dan bernilai (Sativilla Mohan Kumar & Advin Manhar, 2024). Kemampuan analisis sentimen otomatis tidak hanya memberikan gambaran sekilas mengenai opini

konsumen, tetapi juga menyediakan wawasan strategis yang sangat diperlukan bagi pengambilan keputusan bisnis yang responsif dan berbasis data nyata. Dengan demikian, teknologi ini berperan sebagai alat bantu yang strategis untuk perusahaan *e-commerce* dalam memperbaiki produk, meningkatkan layanan, dan memperkuat hubungan dengan pelanggan secara berkelanjutan di tengah persaingan pasar yang semakin ketat.

Metode *Naive Bayes* merupakan algoritma klasifikasi probabilistik yang beroperasi berdasarkan asumsi independensi kondisional antar fitur. Dalam konteks klasifikasi teks, *Naive Bayes* mengestimasi probabilitas kelas dengan menggunakan rumus Bayes dan memilih kelas dengan probabilitas tertinggi sebagai hasil klasifikasi. Keunggulan utama dari metode ini terletak pada kesederhanaannya, efisiensi komputasi, serta kemampuan bekerja dengan baik meskipun data pelatihan tidak terlalu besar (Rajvanshi & Chowdhary, 2017). Meskipun memiliki asumsi independensi yang sering kali tidak terpenuhi secara sempurna dalam data bahasa alami, *Naive Bayes* tetap menjadi metode yang andal dan sering dijadikan baseline dalam berbagai studi klasifikasi teks.

Sementara itu, *Support Vector Machine (SVM)* berfungsi sebagai algoritma klasifikasi yang membangun *hyperplane* optimal dalam ruang berdimensi tinggi untuk memisahkan kelas-kelas data secara maksimal. Dalam penerapannya pada klasifikasi teks, *SVM* dapat menggunakan berbagai fungsi *kernel*, seperti *linear* dan *radial basis function (RBF)*, untuk menangani data yang tidak dapat dipisahkan secara *linear*. *SVM* dikenal memiliki performa yang sangat baik dengan tingkat akurasi tinggi dan cenderung tidak mudah mengalami *overfitting* ketika model disetel dengan tepat (Rajvanshi & Chowdhary, 2017). Perpaduan metode *Naive Bayes* dan *SVM* juga telah diteliti untuk menggabungkan kekuatan probabilistik *Naive Bayes* dalam pemberian bobot fitur dengan kemampuan *SVM* dalam klasifikasi, sehingga menghasilkan performa yang lebih optimal pada tugas klasifikasi teks (Zain & Sibaroni, 2019).

BERT (Bidirectional Encoder Representations from Transformers) merupakan model *deep learning* berbasis arsitektur transformer yang secara signifikan mengubah paradigma pemrosesan bahasa alami (*Natural Language Processing/NLP*). *BERT* mampu memahami konteks kata secara *bidirectional*, artinya model ini menganalisis konteks kata dari kedua sisi kiri dan kanan sekaligus, sehingga mampu menangkap makna yang lebih dalam dan tepat dibandingkan model-model sebelumnya yang hanya *unidirectional* (Devlin et al., 2018).

Analisis sentimen telah menjadi fokus penelitian yang berkembang pesat dalam beberapa tahun terakhir, terutama dalam konteks *e-commerce* dan ulasan pengguna. penelitian yang dilakukan oleh (Indrani et al., 2024) yang mengaplikasikan teknik analisis sentimen pada ulasan produk dengan memanfaatkan metode *machine learning* untuk mengklasifikasikan opini konsumen menjadi kategori positif, negatif, dan netral. Studi ini menekankan pentingnya preprocessing data dan pemilihan fitur yang tepat agar hasil klasifikasi dapat akurat dan efisien.

Penelitian (Sativilla Mohan Kumar & Advin Manhar, 2024) mengembangkan pendekatan yang mengintegrasikan *Robotic Process Automation (RPA)* dan teknik kecerdasan buatan untuk menangani pengolahan ulasan dalam skala besar secara otomatis. Penelitian (Rajvanshi & Chowdhary, 2017) melakukan studi perbandingan antara *Naive Bayes* dan *Support Vector Machine (SVM)* dalam klasifikasi teks, di mana keduanya memiliki keunggulan tersendiri dalam hal kecepatan dan akurasi, bergantung pada karakteristik dataset yang digunakan. Penelitian (Zain & Sibaroni, 2019) menambahkan bahwa pemberian bobot dari *SVM* pada metode *Naive Bayes* dapat meningkatkan performa klasifikasi sentimen, membuka jalan bagi optimasi yang lebih efektif. Penelitian (J. Wang et al., 2024) menyoroti keunggulan *BERT (Bidirectional Encoder Representations from Transformers)* yang mampu menangkap konteks bahasa alami secara mendalam. Model ini terbukti unggul dalam berbagai tugas pemrosesan bahasa alami, termasuk analisis sentimen bahasa Indonesia yang kompleks

dan kaya variasi. (J. Wang et al., 2024) menegaskan bahwa dengan embedding kontekstual yang lebih kaya, *BERT* dapat secara signifikan meningkatkan akurasi klasifikasi dibandingkan metode tradisional, sehingga menjadi pilihan utama dalam pengembangan teknologi analisis sentimen modern. Meskipun banyak penelitian telah menerapkan masing-masing metode tersebut secara terpisah, terdapat kekurangan dalam studi yang secara sistematis membandingkan performa ketiga metode *Naive Bayes*, *SVM*, dan *BERT* khususnya dalam konteks ulasan pengguna aplikasi Shopee yang memiliki karakteristik linguistik dan konteks budaya tersendiri. Kesenjangan ini menimbulkan kebutuhan mendesak untuk melakukan evaluasi komparatif yang holistik guna menentukan algoritma mana yang paling efektif dan efisien untuk analisis sentimen ulasan Shopee.

Penelitian ini bertujuan untuk mengoptimalkan analisis sentimen ulasan pengguna Shopee dengan membandingkan performa ketiga metode tersebut. Penelitian ini juga bertujuan untuk menentukan metode terbaik yang dapat memberikan klasifikasi sentimen secara akurat dan efisien, sehingga dapat menjadi rekomendasi yang tepat bagi pengelola platform *e-commerce* maupun peneliti dalam pengembangan teknologi pengolahan bahasa alami di Indonesia.

1.2 Identifikasi Masalah

Berdasarkan latar belakang tersebut, maka dapat diidentifikasi beberapa masalah yang menjadi fokus penelitian ini, antara lain:

- 1) Banyaknya ulasan pengguna aplikasi Shopee yang sulit dianalisis secara manual karena jumlah dan variasinya sangat besar.
- 2) Belum adanya studi komparatif yang menyeluruh mengenai performa metode *Naive Bayes*, *Support Vector Machine*, dan *BERT* dalam mengklasifikasikan sentimen ulasan Shopee.
- 3) Perlunya optimasi metode analisis sentimen untuk mendapatkan hasil yang lebih akurat dan efisien, khususnya dalam konteks bahasa Indonesia yang memiliki keragaman ekspresi.

1.3 Rumusan Masalah

Dari identifikasi masalah di atas, rumusan masalah dalam penelitian ini adalah:

- 1) Bagaimana performa metode *Naive Bayes* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Shopee?
- 2) Bagaimana performa metode *Support Vector Machine* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Shopee?
- 3) Bagaimana performa model *BERT* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Shopee?
- 4) Metode mana yang paling efektif dan efisien dalam mengklasifikasikan sentimen ulasan pengguna Shopee?

1.4 Batasan Masalah

Agar penelitian ini tetap fokus dan terarah, maka batasan masalah yang diambil adalah sebagai berikut:

- 1) Penelitian ini menggunakan dataset ulasan pengguna aplikasi Shopee yang diperoleh dari platform Kaggle.
- 2) Klasifikasi sentimen hanya dibagi menjadi tiga kategori, yaitu positif, negatif, dan netral.
- 3) Metode yang digunakan terbatas pada *Naive Bayes*, *Support Vector Machine*, dan *BERT* tanpa melakukan penggabungan antar metode.
- 4) Bahasa yang digunakan dalam ulasan adalah bahasa Indonesia.

1.5 Maksud Penelitian

Penelitian ini dimaksudkan untuk melakukan evaluasi dan perbandingan yang objektif terhadap metode *Naive Bayes*, *Support Vector Machine*, dan *BERT* dalam konteks analisis sentimen ulasan pengguna aplikasi Shopee. Maksud penelitian ini adalah untuk mengidentifikasi metode mana yang paling sesuai dan optimal diterapkan pada data ulasan tersebut guna mendukung pengambilan keputusan bisnis yang lebih baik.

1.6 Tujuan Penelitian

Tujuan penelitian ini secara spesifik adalah:

- 1) Untuk mengetahui Performa Metode *Naive Bayes* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Shopee.
- 2) Untuk mengetahui Performa Metode *SVM* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Shopee.
- 3) Untuk mengetahui Performa Metode *BERT* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Shopee.
- 4) Untuk mengetahui metode mana yang paling efektif dan efisien dalam mengklasifikasikan sentimen ulasan pengguna Shopee.

1.7 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

- Manfaat Akademis: Memberikan kontribusi pada pengembangan ilmu pengetahuan di bidang pengolahan bahasa alami dan analisis sentimen, khususnya untuk bahasa Indonesia dan konteks e-commerce.
- Manfaat Praktis: Menjadi referensi bagi pengelola platform e-commerce, seperti Shopee, dalam memilih metode analisis sentimen yang efektif untuk meningkatkan pengelolaan ulasan dan layanan pelanggan.
- Manfaat Sosial: Mendukung peningkatan kepercayaan dan kenyamanan konsumen dalam bertransaksi online, serta mendorong pertumbuhan ekonomi digital yang inklusif.

1.8 Sistematika Penulisan

Penulisan skripsi ini disusun dengan sistematika sebagai berikut:

Bab 1 Pendahuluan: Berisi latar belakang, identifikasi masalah, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

Bab 2 Tinjauan Pustaka: Menguraikan teori-teori dan kajian literatur yang