

BAB I

PENDAHULUAN

1.1 Latar Belakang

Secara kuantitatif, skala permasalahan ulasan pada platform e-commerce dapat dilihat dari besarnya volume aktivitas e-commerce di Indonesia. Gross Merchandise Value (GMV) sektor e-commerce Indonesia tercatat tumbuh signifikan dari sekitar US\$25 miliar pada tahun 2019 menjadi sekitar US\$62 miliar pada tahun 2023, menjadikan Indonesia sebagai negara dengan GMV e-commerce tertinggi di Asia Tenggara, jauh di atas Thailand yang berada di posisi kedua dengan sekitar US\$22 miliar [47]. Sejalan dengan itu, trafik kunjungan ke platform Tokopedia tercatat mencapai sekitar 77,6 juta kunjungan pada Juli 2024, menempatkannya sebagai platform e-commerce dengan trafik kunjungan terbesar kedua di Indonesia setelah Shopee [47], yang berarti volume ulasan baru yang dihasilkan konsumen juga bertambah dalam skala besar setiap harinya

Permasalahan tersebut tercermin secara langsung pada objek penelitian ini. Proses pengumpulan data pada penelitian ini menghasilkan 18.742 ulasan mentah dari lima kategori produk (elektronik, fashion, olahraga, handphone, dan pertukangan) pada platform Tokopedia hanya dalam waktu sekitar 5 jam proses scraping menggunakan satu instance Selenium, dan setelah tahap pembersihan data masih tersisa 15.000 ulasan yang layak dianalisis lebih lanjut. Skala ini sesungguhnya hanya potret kecil dari kondisi riil di lapangan: penelitian terhadap ulasan produk Eiger Adventure pada satu toko resmi (Official Store) di Tokopedia saja melaporkan 3.670 ulasan yang terkumpul dalam periode kurang dari enam bulan [43], sementara pada skala platform secara keseluruhan, jutaan ulasan dihasilkan setiap harinya oleh konsumen di berbagai marketplace [43].

Selain berdampak pada beban kerja penjual, ketidaktepatan dalam memahami sentimen ulasan turut berdampak pada pengalaman calon pembeli. Studi mengenai perilaku konsumen e-commerce menunjukkan bahwa social proof yang mencakup

ulasan pelanggan, testimoni, dan rating produk turut memengaruhi keputusan pembelian konsumen secara signifikan [48], sehingga ketidaktepatan dalam memahami sentimen ulasan secara tepat dapat berdampak langsung pada reputasi toko dan keputusan pembelian di platform e-commerce. Dengan kata lain, keterlambatan atau kesalahan dalam menangkap sentimen negatif tidak hanya merugikan penjual dari sisi operasional, tetapi juga berisiko menurunkan kepercayaan pembeli terhadap toko maupun platform secara keseluruhan.

Selain berdampak pada beban kerja penjual, ketidaktepatan dalam memahami sentimen ulasan turut berdampak pada pengalaman calon pembeli. Studi mengenai perilaku konsumen e-commerce menunjukkan bahwa social proof yang mencakup ulasan pelanggan, testimoni, dan rating produk turut memengaruhi keputusan pembelian konsumen secara signifikan [48], sehingga ketidaktepatan dalam memahami sentimen ulasan secara tepat dapat berdampak langsung pada reputasi toko dan keputusan pembelian di platform e-commerce. Dengan kata lain, keterlambatan atau kesalahan dalam menangkap sentimen negatif tidak hanya merugikan penjual dari sisi operasional, tetapi juga berisiko menurunkan kepercayaan pembeli terhadap toko maupun platform secara keseluruhan.

Kondisi tersebut menunjukkan bahwa pembacaan manual tidak lagi memadai untuk memahami persepsi konsumen dari ulasan dalam jumlah besar, sehingga dibutuhkan pendekatan otomatis yang mampu mengekstraksi opini dan emosi pengguna dari data teks secara sistematis. Analisis sentimen, sebagai salah satu pendekatan penting dalam Natural Language Processing (NLP) [1], [2], bertujuan untuk mengklasifikasikan polaritas opini ke dalam kategori tertentu, seperti positif, negatif, atau netral. Melalui pendekatan ini, ribuan ulasan yang sebelumnya tidak mungkin dibaca satu per satu dapat diringkas menjadi indikator sentimen yang mudah dipantau, baik oleh penjual maupun pengelola platform.

Tantangan utama dalam analisis sentimen tidak hanya terletak pada volume data yang besar, tetapi juga pada kompleksitas bahasa alami itu sendiri, termasuk ambiguitas makna, penggunaan bahasa informal, variasi gaya penulisan, serta konteks semantik yang bergantung pada situasi tertentu [3], [4]. Tantangan ini

menjadi semakin signifikan ketika analisis dilakukan pada bahasa dengan sumber daya terbatas, seperti Bahasa Indonesia, yang memiliki karakteristik morfologi, afiksasi, dan struktur kalimat yang berbeda dari bahasa Inggris [5]. Kondisi ini diperparah oleh karakteristik ulasan e-commerce berbahasa Indonesia yang menggunakan bahasa informal, singkatan, dan campuran sentimen dalam satu kalimat, sebagaimana juga ditemukan pada studi ulasan aplikasi belanja online lain seperti Sayurbox, yang tetap memerlukan pengecekan manual pada sebagian data sekalipun proses pelabelan awal telah diotomatisasi menggunakan pendekatan lexicon [45].

Dalam praktik penelitian dan pengembangan sistem analisis sentimen, pendekatan Machine Learning konvensional telah lama digunakan sebagai solusi utama. Metode seperti Naïve Bayes dan Support Vector Machine dikenal memiliki keunggulan dari sisi efisiensi komputasi, kemudahan implementasi, serta stabilitas performa pada dataset berukuran kecil hingga menengah [6], [7]. Naïve Bayes memiliki beberapa keunggulan seperti metode yang sederhana, proses komputasi yang cepat, serta mampu menghasilkan tingkat akurasi yang cukup tinggi dalam berbagai penelitian analisis sentimen, meskipun metode ini memiliki keterbatasan karena mengasumsikan bahwa setiap atribut atau fitur dalam data bersifat independen satu sama lain, sehingga asumsi tersebut tidak selalu terpenuhi dalam data teks nyata dan dapat mempengaruhi tingkat akurasi klasifikasi [8]. Di sisi lain, Support Vector Machine juga memiliki kelemahan terutama dalam proses pemilihan parameter yang tepat agar model dapat bekerja secara optimal. Pendekatan ini umumnya mengandalkan representasi fitur berbasis frekuensi kata, seperti bag-of-words atau Term Frequency–Inverse Document Frequency (TF-IDF), yang mengasumsikan independensi antar kata dan mengabaikan urutan serta konteks semantik dalam teks [9], [10].

Meskipun pendekatan Machine Learning konvensional cukup efektif dalam berbagai skenario, keterbatasannya mulai terlihat ketika dihadapkan pada teks yang memiliki makna kontekstual kompleks, ironi, atau hubungan semantik antar kata yang tidak dapat direpresentasikan secara linear [11]. Perkembangan model

berbasis transformer menunjukkan peningkatan yang signifikan dalam performa klasifikasi [12], mendorong berkembangnya pendekatan Deep Learning yang mampu mempelajari representasi fitur secara otomatis dan hierarkis dari data mentah tanpa bergantung pada rekayasa fitur manual [13]. Model-model ini mampu menangkap ketergantungan jangka panjang dan konteks dua arah dalam teks, sehingga menghasilkan representasi kata yang lebih kaya dan kontekstual [14], [15]. Dalam konteks Bahasa Indonesia, munculnya model pra-latih seperti IndoBERT memberikan peluang besar untuk meningkatkan performa analisis sentimen dengan memanfaatkan pengetahuan linguistik yang telah dipelajari dari korpus besar berbahasa Indonesia.

Namun, penerapan model Deep Learning tidak terlepas dari tantangan, terutama terkait kebutuhan sumber daya komputasi yang tinggi, kompleksitas proses pelatihan, serta risiko overfitting pada dataset dengan ukuran terbatas [16]. Selain itu, penggunaan model Deep Learning tidak selalu menjamin peningkatan performa yang signifikan dibandingkan metode Machine Learning konvensional, terutama pada domain atau karakteristik data tertentu [17]. Oleh karena itu, pemilihan metode yang tepat harus didasarkan pada evaluasi empiris yang komprehensif, bukan semata-mata pada tren teknologi. Hingga saat ini, masih terdapat kesenjangan penelitian yang membahas perbandingan menyeluruh antara pendekatan Machine Learning klasik dan Deep Learning dalam analisis sentimen Bahasa Indonesia dengan skenario eksperimen yang konsisten dan terkontrol [18].

Permasalahan volume ulasan yang besar ini diperparah oleh satu tantangan tambahan yang jarang diperhatikan pada penelitian analisis sentimen berbasis e-commerce, yaitu potensi inkonsistensi antara rating bintang yang diberikan konsumen dengan sentimen yang sebenarnya tertulis pada teks ulasan. Studi analisis sentimen ulasan Shopee menggunakan BERT mengidentifikasi bahasa informal, ambiguitas semantik, dan inkonsistensi antara sentimen tekstual dan peringkat bintang sebagai tantangan yang memengaruhi akurasi klasifikasi pada ulasan e-commerce berbahasa Indonesia [49]. Hal ini berarti volume ulasan yang besar tidak hanya menyulitkan pembacaan manual dari sisi jumlah, tetapi juga

menyulitkan proses pelabelan otomatis berbasis rating itu sendiri, karena rating semata tidak selalu mencerminkan sentimen tekstual yang aktual, sehingga validitas label yang digunakan untuk melatih model perlu diuji, bukan diasumsikan benar begitu saja. Pendekatan pelabelan otomatis berbasis rating (rating-based labeling) yang digunakan pada penelitian ini juga merupakan strategi yang umum diterapkan pada studi analisis sentimen marketplace berskala besar justru karena keterbatasan waktu anotasi manual pada dataset seukuran ini [44].

Urgensi penelitian ini didasari oleh kenyataan bahwa pelaku usaha e-commerce harus memproses ribuan ulasan setiap hari untuk memahami persepsi konsumen, namun proses pembacaan manual tidak lagi memungkinkan seiring meningkatnya volume ulasan. Tanpa sistem klasifikasi sentimen yang akurat dan efisien, informasi penting mengenai kepuasan maupun keluhan konsumen berisiko terlewat, sehingga dapat menghambat pengambilan keputusan bisnis yang tepat waktu. Permasalahan ini menjadi semakin kompleks pada ulasan berbahasa Indonesia karena gaya penulisan yang informal, penggunaan singkatan, dan campuran sentimen dalam satu kalimat, sehingga diperlukan metode klasifikasi yang benar-benar teruji pada karakteristik bahasa tersebut, bukan hanya hasil adaptasi dari penelitian berbahasa Inggris. Penelitian ini berbeda dari penelitian-penelitian terdahulu yang umumnya hanya membandingkan dua metode secara terpisah misalnya IndoBERT dengan SVM saja, atau Naïve Bayes dengan SVM saja dan hanya berfokus pada satu kategori produk. Penelitian ini menghadirkan perbandingan tiga metode Naïve Bayes, SVM, dan IndoBERT secara simultan dalam satu skenario eksperimen yang konsisten, menggunakan dataset multi-kategori dari platform Tokopedia, sehingga hasil perbandingan lebih representatif dan dapat menjadi acuan praktis bagi pelaku e-commerce dalam memilih metode analisis sentimen yang sesuai dengan kebutuhan dan keterbatasan sumber daya yang dimiliki. Kesenjangan penelitian yang membandingkan tiga metode klasifikasi sekaligus dalam satu skenario eksperimen yang konsisten juga memiliki relevansi praktis: pelaku usaha e-commerce dengan sumber daya komputasi terbatas perlu mengetahui apakah selisih performa antara metode Deep Learning dan Machine Learning konvensional cukup signifikan untuk membenarkan biaya

komputasi tambahan, ataukah metode konvensional yang lebih ringan sudah memadai untuk kebutuhan operasional sehari-hari.

Hasil penelitian ini secara khusus bermanfaat bagi tiga kelompok pemangku kepentingan. Pertama, bagi pelaku usaha dan penjual di platform e-commerce, sistem ini dapat digunakan sebagai alat bantu otomatis untuk memantau persepsi konsumen terhadap produk secara cepat tanpa harus membaca seluruh ulasan satu per satu, sehingga keluhan yang bersentimen negatif dapat diprioritaskan untuk ditindaklanjuti lebih awal. Kedua, bagi pengelola platform e-commerce, hasil analisis sentimen agregat dapat menjadi salah satu indikator kualitas layanan maupun produk yang dijual di platform tersebut, mendukung pengambilan keputusan terkait moderasi konten maupun rekomendasi produk. Ketiga, bagi akademisi dan peneliti di bidang NLP, penelitian ini memberikan acuan empiris mengenai perbandingan performa pendekatan Machine Learning konvensional dan Deep Learning berbasis transformer pada konteks Bahasa Indonesia, yang dapat menjadi dasar pengembangan penelitian lanjutan, khususnya pada domain ulasan e-commerce yang bersifat informal dan multikelas.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, terdapat beberapa permasalahan yang dapat dirumuskan dalam penelitian ini, yaitu:

1. Bagaimana kinerja metode *Machine Learning* konvensional (*Naïve Bayes* dan *SVM*) dalam analisis sentimen teks Bahasa Indonesia?
2. Sejauh mana metode *Deep Learning* berbasis IndoBERT dapat meningkatkan akurasi dan efektivitas klasifikasi sentimen dibandingkan metode *Machine Learning* konvensional?
3. Apa saja tantangan utama yang ditemukan ketika menerapkan kedua pendekatan tersebut pada dataset Bahasa Indonesia?

1.3 Batasan Masalah

Penelitian ini memiliki beberapa batasan yang bertujuan untuk memperjelas lingkup penelitian dan meningkatkan fokus pada permasalahan yang diidentifikasi. Batasan-batasan tersebut adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah ulasan produk berbahasa Indonesia yang bersumber dari *platform e-commerce* Tokopedia, yang dibatasi pada lima kategori produk, yaitu elektronik, fashion, olahraga, handphone, dan pertukangan.
2. Kategori sentimen yang dianalisis dibatasi pada tiga kelas: positif, negatif, dan netral.
3. Metode Machine Learning konvensional yang dibandingkan hanya mencakup *Naïve Bayes* dan *SVM* dengan fitur representasi berbasis TF-IDF.
4. Model *Deep Learning* yang digunakan adalah IndoBERT yang telah dipratrein untuk Bahasa Indonesia, tanpa *fine-tuning* arsitektur tambahan di luar parameter default *transformer*.

1.4 Tujuan dan Manfaat Penelitian

Penelitian ini dirancang untuk memberikan kontribusi ilmiah dalam bidang analisis sentimen Bahasa Indonesia melalui pendekatan komparatif antara metode *Machine Learning* konvensional dan *Deep Learning*. Dengan semakin berkembangnya model berbasis *transformer* dan meningkatnya kompleksitas data teks, diperlukan kajian yang tidak hanya berfokus pada peningkatan akurasi semata, tetapi juga pada pemahaman karakteristik, kelebihan, serta keterbatasan masing-masing pendekatan secara sistematis dan objektif.

Melalui evaluasi empiris yang terstruktur, penelitian ini berupaya menjawab kebutuhan akan panduan metodologis dalam pemilihan algoritma analisis sentimen yang sesuai dengan karakteristik Bahasa Indonesia. Selain itu, penelitian ini juga bertujuan untuk menjembatani kesenjangan antara pendekatan konvensional yang lebih sederhana dan pendekatan modern yang lebih kompleks, sehingga hasil penelitian dapat dimanfaatkan secara luas oleh peneliti maupun praktisi di bidang *Natural Language Processing*.

1.4.1 Tujuan Penelitian

Tujuan utama dari penelitian ini adalah untuk melakukan analisis dan perbandingan kinerja metode Machine Learning dan Deep Learning dalam tugas analisis sentimen teks berbahasa Indonesia. Secara lebih rinci, tujuan penelitian ini meliputi:

1. Mengkaji kemampuan metode *Machine Learning* konvensional, khususnya *Naïve Bayes* dan *Support Vector Machine*, dalam mengklasifikasikan sentimen teks Bahasa Indonesia berdasarkan representasi fitur berbasis statistik.
2. Mengevaluasi efektivitas model *Deep Learning* berbasis IndoBERT dalam menangkap konteks semantik dan hubungan antar kata dalam teks Bahasa Indonesia untuk meningkatkan akurasi klasifikasi sentimen.
3. Membandingkan performa kedua pendekatan tersebut secara kuantitatif menggunakan metrik evaluasi, sehingga diperoleh gambaran objektif mengenai keunggulan dan keterbatasan masing-masing metode.

1.4.2 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan manfaat yang signifikan baik dari sisi teoretis maupun praktis. Adapun manfaat penelitian ini dijelaskan sebagai berikut:

1. Penelitian ini diharapkan dapat memperkaya kajian akademik dalam bidang analisis sentimen dan *Natural Language Processing*, khususnya untuk Bahasa Indonesia. Dengan menyajikan perbandingan empiris antara metode *Machine Learning* konvensional dan *Deep Learning*, penelitian ini dapat menjadi referensi bagi studi lanjutan yang berfokus pada pengembangan model, evaluasi algoritma, atau eksplorasi pendekatan hibrida dalam analisis sentimen.
2. Penelitian ini diharapkan dapat memberikan panduan bagi pengembang sistem dan praktisi data dalam memilih metode analisis sentimen yang sesuai dengan kebutuhan dan keterbatasan yang dihadapi, seperti ukuran dataset, kompleksitas bahasa, dan ketersediaan sumber daya komputasi.

Hasil penelitian ini dapat dimanfaatkan dalam berbagai aplikasi nyata, termasuk analisis opini publik, pemantauan kepuasan pelanggan, dan pengambilan keputusan berbasis data.

1.5 Sistematika Penulisan

Sistematika penulisan artikel ini disusun untuk memudahkan alur pemahaman dan keterbacaan sebagai berikut:

BAB I PENDAHULUAN

Berisi latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan.

BAB II LANDASAN TEORI

Mengulas teori-teori kunci tentang analisis sentimen, representasi teks, algoritma Naïve Bayes, SVM, dan model IndoBERT, serta studi komparatif sebelumnya.

BAB III METODOLOGI PENELITIAN

Menjelaskan desain penelitian, dataset yang digunakan, proses preprocessing, konfigurasi eksperimen, dan metrik evaluasi.

BAB IV ANALISIS DAN HASIL PENELITIAN

Menjabarkan hasil penerapan CRISP-DM secara menyeluruh, mulai dari business understanding, data understanding, data preparation, modeling (Naïve Bayes, SVM, dan IndoBERT), deployment sistem analisis sentimen berbasis Streamlit, hingga analisis topik ulasan konsumen menggunakan WordCloud.

BAB V SIMPULAN DAN SARAN

Berisi simpulan atas hasil perbandingan ketiga model serta saran bagi penelitian dan implementasi selanjutnya, dilengkapi dengan Daftar Pustaka dan Lampiran sebagai pendukung penelitian.