

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi informasi dan komunikasi dalam beberapa dekade terakhir telah membawa perubahan yang sangat signifikan dalam berbagai aspek kehidupan masyarakat, termasuk dalam cara manusia berinteraksi, memperoleh informasi, serta melakukan aktivitas ekonomi (Shan, 2023). Kemajuan teknologi digital yang didukung oleh peningkatan akses internet serta penggunaan perangkat *mobile* telah mendorong terciptanya ekosistem digital yang semakin luas dan kompleks. Transformasi digital ini tidak hanya mengubah pola komunikasi masyarakat, tetapi juga mempengaruhi pola konsumsi, perilaku belanja, serta strategi pemasaran yang digunakan oleh pelaku usaha (Liao et al., 2021).

Indonesia merupakan salah satu negara dengan tingkat pertumbuhan pengguna internet yang sangat pesat. Berdasarkan laporan Asosiasi Penyelenggara Jasa Internet Indonesia Survei APJII (2026), jumlah pengguna internet di Indonesia terus mengalami peningkatan setiap tahunnya dan telah mencapai lebih dari 210 juta pengguna. Tingginya penetrasi internet tersebut menunjukkan bahwa sebagian besar masyarakat Indonesia telah terhubung dengan dunia digital dan memanfaatkan internet dalam berbagai aktivitas sehari-hari, mulai dari komunikasi, hiburan, hingga kegiatan ekonomi seperti transaksi jual beli secara *online*.

Pertumbuhan berbagai platform media sosial dan *e-commerce* di Indonesia yang mengalami perkembangan signifikan salah satunya adalah platform TikTok, yang awalnya dikenal sebagai media berbagi video pendek, namun kini berkembang menjadi platform perdagangan digital melalui fitur TikTok Shop. TikTok Shop memungkinkan pengguna untuk melakukan transaksi jual beli secara langsung di dalam aplikasi melalui fitur *live streaming*, video promosi, dan kolom komentar (Putri et al., 2023).

Tingginya popularitas TikTok di Indonesia turut memperkuat perannya sebagai platform digital yang potensial dalam aktivitas pemasaran dan perdagangan. Berdasarkan laporan *We Are Social* (2025), TikTok termasuk dalam lima besar platform yang paling banyak digunakan di Indonesia dengan tingkat penggunaan mencapai 78,4% dari pengguna internet usia 16 tahun ke atas. Selain itu, tingkat keterlibatan pengguna TikTok juga tergolong tinggi, dengan rata-rata waktu penggunaan harian mencapai sekitar 1 jam 53 menit. Hal ini menunjukkan bahwa TikTok tidak hanya berfungsi sebagai media hiburan, tetapi juga sebagai platform yang efektif dalam menarik perhatian pengguna. Kondisi ini memberikan peluang besar bagi perkembangan TikTok Shop sebagai fitur *e-commerce* yang memanfaatkan tingginya interaksi dan aktivitas pengguna dalam mendorong keputusan pembelian.

Fenomena produk *viral* di TikTok Shop menjadi salah satu faktor yang mempengaruhi keputusan pembelian konsumen (Khikam Ali Majid et al., 2024). Produk yang *viral* biasanya mendapatkan perhatian yang sangat

besar dari pengguna, sehingga menghasilkan ribuan hingga jutaan komentar dari pengguna TikTok. Menurut Nur Intan Salsa Nabila dan Putra (2024) menyatakan bahwa *viralitas* sebuah merek dipengaruhi oleh sentimen yang berkembang di kalangan konsumen. Komentar-komentar tersebut berisi berbagai opini, pengalaman, maupun penilaian konsumen terhadap produk yang dipromosikan. Komentar ini dapat berupa sentimen positif, negatif, maupun netral yang mencerminkan persepsi konsumen terhadap kualitas produk, pelayanan, maupun pengalaman pembelian.

Banyaknya komentar yang muncul pada video *review* produk di TikTok menjadikan analisis manual terhadap opini pengguna menjadi sulit dilakukan karena jumlah data yang sangat besar. Oleh karena itu, diperlukan suatu metode otomatis untuk menganalisis sentimen dari komentar-komentar tersebut. Analisis sentimen merupakan salah satu teknik dalam bidang *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau emosi yang terkandung dalam suatu teks (Julianti et al., 2024).

Dalam penelitian analisis sentimen, terdapat berbagai metode yang dapat digunakan, selain itu *Machine Learning* dan *Deep Learning* telah menjadi solusi yang semakin populer dalam mengatasi tugas-tugas kompleks seperti klasifikasi sentimen mulai dari metode *machine learning* tradisional hingga *deep learning* dan model berbasis *transformer* (Tyas et al., 2024). *Support Vector Machine* (SVM) merupakan *algoritma machine learning* yang digunakan untuk klasifikasi data dengan cara menentukan garis pemisah

optimal (*hyperplane*) antar kelas (Fauzi, 2020) dalam jurnal (Damayanti et al., 2024). Menurut Kamal dan Astri (2025) dalam penggabungan kode, kaya aferesis, afiksasi dan serapan teknik *machine learning* klasik masih kurang andal dalam bahasa Indonesia Teknik berbasis lexicon atau *machine learning* klasik kurang andal pada bahasa Indonesia sehingga membuka peluang perkembangan metode seperti *deep learning*. *Long Short-Term Memory* (LSTM) merupakan salah satu metode *deep learning* untuk klasifikasi teks (Naquitasia et al., 2022). Sebaliknya, *Bidirectional Encoder Representations from Transformers* (BERT) yang dalam dari teks yang tidak berlabel dengan mengondisikan konteks kiri dan kanan secara bersama-sama di semua lapisan (Mandhasiya et al., 2024). Model BERT yang dilatih khusus menggunakan kamus bahasa Indonesia yang terbukti efektif, membagi topik dan menyatukan hubungan antar kata yaitu IndoBERT (Yulianti & Nissa, 2024).

Support Vector Machine merupakan salah satu algoritma klasifikasi dalam *supervised learning* yang banyak digunakan karena kemampuannya dalam memisahkan data secara optimal menggunakan *hyperplane* (Husada & Paramita, 2021). Di sisi lain, *Long Short-Term Memory* (LSTM) merupakan salah satu jenis jaringan saraf tiruan dalam *deep learning* yang dirancang untuk memproses data berurutan seperti teks. LSTM mampu menangkap hubungan antar kata dalam sebuah kalimat sehingga dapat memahami konteks yang lebih baik dibandingkan metode *machine learning* tradisional.

Perkembangan terbaru dalam bidang pemrosesan bahasa alami ditandai dengan munculnya model berbasis transformer, salah satunya adalah BERT (*Bidirectional Encoder Representations from Transformers*). Untuk bahasa Indonesia, dikembangkan model IndoBERT yang merupakan model BERT khusus untuk Bahasa Indonesia dengan menggunakan data teks dalam jumlah besar yang berasal dari berbagai sumber, seperti koran *online* dan web berbahasa Indonesia. Model ini dilatih menggunakan teknik *transfer learning* sehingga mampu memahami pola dan struktur bahasa Indonesia dengan baik. Agar dapat digunakan untuk tugas tertentu, dilakukan proses *fine-tuning*, yaitu penyesuaian model yang sudah dilatih sebelumnya agar dapat memberikan hasil yang lebih optimal, misalnya dalam analisis sentimen (Devlin et al., 2019).

Namun demikian, sebagian besar penelitian terdahulu masih menggunakan data dari platform seperti Twitter, Instagram, YouTube, maupun *marketplace* konvensional seperti Shopee dan Tokopedia. Penelitian yang secara khusus mengkaji komentar pada video review produk TikTok masih sangat terbatas, padahal platform ini memiliki karakteristik bahasa yang berbeda. Komentar pada video review produk TikTok umumnya menggunakan bahasa yang tidak baku, singkatan, slang, emoji, serta campuran bahasa yang sering kali muncul dalam satu kalimat. Variasi bahasa yang dinamis dan tidak terstruktur ini dapat menjadi tantangan tersendiri bagi model analisis sentimen dalam memahami konteks secara akurat.

Oleh karena itu, diperlukan penelitian yang secara khusus membandingkan kinerja metode *Support Vector Machine* (SVM), *Long Short-Term Memory* (LSTM), dan *IndoBERT* pada data komentar pada video review produk di TikTok. Penelitian ini dilakukan untuk melihat kemampuan masing-masing metode dalam mengklasifikasikan sentimen, serta mengetahui metode yang paling efektif dalam menangani karakteristik bahasa tidak formal yang umum digunakan pada platform tersebut.

Meskipun berbagai metode telah banyak digunakan dalam analisis sentimen, masih terdapat kesenjangan penelitian (*research gap*), terutama dalam hal perbandingan kinerja antara metode *machine learning* seperti SVM, metode *deep learning* seperti LSTM, serta model berbasis transformer seperti *IndoBERT* pada konteks data tertentu. Beberapa penelitian sebelumnya menunjukkan bahwa setiap metode memiliki keunggulan masing-masing. Salah satunya, penelitian oleh Alya Sahrani dan Rizki Hikmawan (2025) menunjukkan bahwa metode *IndoBERT* mampu menghasilkan performa yang tinggi dengan akurasi mencapai 95%, bahkan dalam kondisi tertentu dapat lebih unggul dibandingkan LSTM. Di sisi lain, penelitian oleh Apriyadi & Styawati (2025) menunjukkan bahwa *IndoBERT* mampu mengungguli SVM dengan akurasi mencapai 85%, sehingga dinilai lebih efektif dalam memahami konteks bahasa.

Penelitian lain oleh Damayanti et al. (2024), justru menemukan bahwa SVM lebih unggul dibandingkan LSTM dalam analisis ulasan aplikasi dengan akurasi sebesar 83,5%. Hasil-hasil tersebut menunjukkan bahwa

belum terdapat kesimpulan yang konsisten mengenai metode terbaik, karena performa model sangat dipengaruhi oleh jenis data dan karakteristik bahasa yang digunakan.

Oleh karena itu, diperlukan penelitian lanjutan untuk membandingkan ketiga metode tersebut pada data yang berbeda agar diperoleh hasil yang lebih menyeluruh.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana gambaran sentimen pengguna terhadap komentar pada video review produk di TikTok/ berdasarkan komentar yang diberikan?
2. Bagaimana perbandingan kinerja metode *Support Vector Machine* (SVM), *Long Short-Term Memory* (LSTM), dan *IndoBERT* dalam melakukan klasifikasi sentimen komentar video review produk di TikTok?
3. Metode manakah yang memiliki kinerja terbaik berdasarkan hasil evaluasi *accuracy*, *precision*, *recall*, dan *F1-score*, serta apa faktor yang mempengaruhi perbedaan performa antar metode dalam klasifikasi sentimen?

1.3 Batasan Masalah

Agar penelitian lebih terarah dan tidak menyimpang dari tujuan penelitian, maka ditetapkan beberapa batasan masalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini berupa komentar video review produk di TikTok.
2. Analisis sentimen dilakukan dengan mengklasifikasikan komentar ke dalam sentimen positif, negatif, dan netral.
3. Metode yang digunakan dalam penelitian ini hanya terbatas pada *Support Vector Machine* (SVM), *Long Short-Term Memory* (LSTM), dan *IndoBERT*.
4. Video yang dipakai berjumlah 10 video.
5. Kriteria pemilihan video yaitu kategori *trending* TikTok dalam rentang waktu 7 hingga 30 hari terakhir, memiliki jumlah penonton minimal 3 juta (*views*), serta jumlah *likes* minimal 50 ribu dengan target 10 video dan maksimal 2000 komentar.
6. Proses analisis teks dilakukan melalui tahapan *preprocessing* yang meliputi *cleaning*, *case folding*, *normalisasi*, *tokenizing*, *stopword*, dan *stemming*.
7. Evaluasi kinerja model dilakukan menggunakan *metrik accuracy*, *precision*, *recall*, dan *F1-score*.
8. Data penelitian berupa komentar pada video review produk di TikTok yang tidak memiliki keranjang kuning, sehingga objek penelitian berfokus pada komentar pada video review produk di TikTok.
9. Komentar yang diambil dari TikTok minimal 1000 dan maksimal 2000.

10. Pelatihan dan pengujian ketiga model dilakukan menggunakan konfigurasi parameter yang telah ditentukan dengan mempertimbangkan sumber daya komputasi dan waktu pelatihan yang tersedia.
11. Jumlah *epoch* dan nilai parameter C dibatasi untuk menyesuaikan dengan keterbatasan *hardware* yang digunakan dalam penelitian, yaitu nilai C pada SVM diuji pada rentang 1–10, *epoch* pada LSTM diuji pada rentang 1–10, dan *epoch* pada IndoBERT diuji pada rentang 1–5.
12. Data yang digunakan merupakan komentar berbahasa Indonesia.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Menganalisis distribusi sentimen (positif, negatif, dan netral) pada komentar review video produk di TikTok.
2. Membandingkan kinerja metode *Support Vector Machine* (SVM), *Long Short-Term Memory* (LSTM), dan IndoBERT untuk menentukan metode yang paling efektif berdasarkan metrik evaluasi yang digunakan.
3. Mengetahui metode dengan akurasi terbaik setelah dibandingkan serta menganalisis faktor-faktor yang mempengaruhi perbedaan kinerja antar metode dalam klasifikasi sentimen.

1.5 Manfaat Penelitian

Berdasarkan tujuan penelitian yang telah ditetapkan, maka manfaat penelitian ini adalah sebagai berikut:

1.5.1 Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu pengetahuan khususnya di bidang *Natural Language Processing* (NLP) dan analisis sentimen, serta menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan analisis sentimen pada media sosial dan platform *e-commerce*.

1.5.2 Manfaat Praktis

1. Bagi Peneliti : Penelitian ini dapat menambah wawasan dan pengalaman dalam menerapkan metode *machine learning* dan *deep learning* dalam analisis sentimen.
2. Bagi Pelaku : Bisnis *online* hasil penelitian ini dapat membantu pelaku bisnis untuk memahami persepsi konsumen terhadap produk yang dipasarkan melalui TikTok berdasarkan analisis komentar pengguna.
3. Bagi Peneliti : Selanjutnya penelitian ini dapat dijadikan referensi untuk pengembangan penelitian lebih lanjut terkait perbandingan metode analisis sentimen pada data media sosial.