

BAB I

PENDAHULUAN

Bab pendahuluan ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika penulisan yang akan digunakan pada penelitian Perbandingan TF-IDF dan SBERT dengan *Logistic Regression* untuk *Aspect Level Sentiment Analysis* pada Ulasan *E-commerce* Studi Kasus Aplikasi Shopee.

1.1 Latar Belakang

Perkembangan ekonomi digital di Indonesia berlangsung pesat dan mendorong perubahan pada berbagai sektor, termasuk Usaha Mikro, Kecil, dan Menengah (UMKM) (Rafiani dkk., 2024). Adopsi platform digital seperti *e-commerce* dan dompet digital meningkat signifikan, terutama selama pandemi *Coronavirus Disease 2019 (COVID-19)* (Kushadiani dkk., 2021). Proyeksi menunjukkan nilai *e-commerce* diperkirakan mencapai 130 miliar *United States Dollar (USD)* pada tahun 2020, meningkat sekitar 1625% dibandingkan tahun 2013 (Pitoyo dkk., 2020).

Pertumbuhan pesat sektor *e-commerce* tidak hanya mengubah cara konsumen berbelanja, tetapi juga menghasilkan volume data interaksi konsumen yang sangat besar. Salah satu sumber data paling kaya dan langsung dari konsumen adalah ulasan pengguna yang memuat opini, pengalaman, dan evaluasi terhadap berbagai aspek layanan (Mandala dkk., 2023). Data ulasan ini memberikan gambaran mendalam tentang kebutuhan dan preferensi konsumen, sehingga membantu perusahaan memahami perilaku belanja konsumen serta mengidentifikasi berbagai aspek yang perlu ditingkatkan.

Analisis sentimen tradisional merupakan pendekatan yang berfokus pada klasifikasi polaritas sentimen pada tingkat ulasan secara keseluruhan, tanpa mempertimbangkan aspek-aspek spesifik yang dibahas di dalamnya. Dalam konteks *e-commerce*, pendekatan ini dinilai kurang memadai karena satu ulasan sering kali memuat opini yang berbeda terhadap berbagai aspek, seperti produk, harga, maupun layanan (Paramita & Jusak, 2025). Kendati memiliki keterbatasan tersebut, analisis sentimen tradisional tetap berperan penting dalam penelitian awal dan terbukti mampu memberikan hasil yang cukup akurat. Model seperti *Naive Bayes (NB)*, *Support Vector Machine (SVM)*, dan *Logistic Regression (LR)* misalnya,

telah menunjukkan efektivitas pada tugas analisis sentimen umum dengan tingkat akurasi mencapai sekitar 88-89% pada dataset tertentu (Soni dkk., 2024).

Untuk mengatasi keterbatasan analisis sentimen tradisional, dikembangkan metode *Aspect-Based Sentiment Analysis* (ABSA) yang mengekstraksi aspek spesifik dalam ulasan dan mengklasifikasikan polaritas sentimen pada masing-masing aspek (Kocon dkk., 2021). Teknik ABSA bervariasi dari model *supervised* klasik seperti LR NB, dan SVM, model *deep learning* seperti *Recurrent Neural Network* (RNN), *Long Short-Term Memory* (LSTM), dan *Convolutional Neural Network* (CNN), hingga model *transformer pretrained* seperti *Bidirectional Encoder Representations from Transformers* (BERT) dan turunannya (Keivandarian & Carvalho, 2025).

Salah satu aspek penting dalam ABSA adalah model representasi teks yang digunakan. Dua paradigma fundamental untuk representasi teks telah mendominasi klasifikasi teks, yaitu paradigma *sparse* dan *dense*. Paradigma *sparse* menggunakan pendekatan *bag-of-words* yang berasumsi bahwa kehadiran dan frekuensi token tertentu menjadi indikator kuat kelas dokumen melalui distribusi statistik kata, sedangkan paradigma *dense* menggunakan pendekatan *contextual embedding* yang berasumsi bahwa makna muncul dari relasi kontekstual antar-token, sehingga mampu menangkap parafrase, sinonimi, dan nuansa semantik yang tidak dapat ditangkap hanya dari kehadiran token individual (Westin, 2024). Perbandingan kedua paradigma ini (*sparse* dan *dense*) penting untuk mengetahui apakah klasifikasi teks lebih dipengaruhi frekuensi token atau konteks kalimat, dan berdampak langsung di pemilihan model, efisiensi komputasi, dan kemudahan analisis.

Representasi *sparse* dibangun melalui berbagai metode, seperti *Count Vectorizer*, *Term Frequency-Inverse Document Frequency* (TF-IDF), dan *Best Matching 25* (BM25). *Count Vectorizer* menghitung frekuensi kemunculan setiap kata dalam dokumen namun memberikan bobot yang sama untuk semua kata tanpa mempertimbangkan kepentingan relatifnya (Bhansali dkk., 2022). BM25 dirancang khusus untuk *information retrieval* (IR) dengan mekanisme *term frequency saturation* dan *document length normalization* yang efektif untuk dokumen dengan panjang bervariasi (Kadhim, 2019). TF-IDF menggunakan skema pembobotan statistik yang menghasilkan representasi vektor untuk unit teks berdasarkan frekuensi kata relatif terhadap keseluruhan korpus (Aydemir dkk., 2025).

Di antara metode *sparse* yang tersedia, TF-IDF memiliki landasan teoretis yang kuat dalam teori informasi dan status praktis sebagai standar emas (*gold standard*) untuk dasar perbandingan klasifikasi teks, serta bersifat mudah direproduksi dan efisien (Sheridan & Onsjö, 2024). Dibandingkan *Count Vectorizer* yang memberikan bobot seragam, TF-IDF mampu mengidentifikasi *term* yang benar-benar relevan dengan memberikan bobot lebih tinggi pada kata domain-spesifik yang penting untuk ABSA, dengan pencapaian akurasi 83% dibandingkan 75% pada klasifikasi berita Myanmar (Minn dkk., 2025). Dibandingkan BM25 yang dirancang untuk dokumen dengan panjang bervariasi, TF-IDF lebih sesuai untuk ulasan *e-commerce* yang cenderung pendek dan homogen, dengan pencapaian F1 score 89,77% dibandingkan 89,16% pada ekstraksi fitur ulasan di Twitter (Kadhim, 2019). Karakteristik TF-IDF yang mudah direproduksi, efisien, dan jelas menjadikannya pilihan yang tepat dibandingkan alternatif *sparse* lainnya.

Representasi *dense* memiliki berbagai metode seperti *Word to Vector* (Word2Vec), *Global Vectors for Word Representation* (GloVe), *Bidirectional Encoder Representations from Transformers* (BERT), dan *Sentence-BERT* (SBERT). Word2Vec dan GloVe menghasilkan representasi vektor statis untuk setiap kata berdasarkan ko-okurensi kata dalam korpus, namun tidak dapat menangkap perubahan makna kata berdasarkan konteks kalimat (Leonid dkk., 2025). BERT menggunakan mekanisme *attention bidirectional* untuk merepresentasikan kata, frasa, maupun kalimat penuh sebagai vektor padat yang mampu menangkap makna semantik dan hubungan kontekstual (Devlin dkk., 2019). SBERT dirancang khusus untuk mengatasi keterbatasan komputasi BERT dengan menghasilkan *sentence embedding* yang bermakna secara semantik dan efisien untuk perbandingan pada tingkat kalimat (Taneja dkk., 2023).

Di antara metode *dense* yang tersedia, SBERT mampu mengatasi keterbatasan *embedding* kata statis milik Word2Vec dan GloVe dengan menangkap konteks kalimat secara menyeluruh. Hal ini tercermin dari pencapaian akurasi 90,9% pada tugas klasifikasi teks deteksi *cyberstalking*, lebih tinggi dibandingkan GloVe (85%) dan Word2Vec (81,2%) (Gautam & Bansal, 2022). Pemahaman semantik tingkat kalimat, efisiensi komputasi, dan fleksibilitas menjadikan SBERT pilihan yang tepat dibandingkan alternatif *dense* lainnya.

Pemilihan algoritma klasifikasi yang tepat sama pentingnya dengan pemilihan metode representasi teks dalam ABSA. Berbagai *classifier* dapat menjadi kandidat untuk klasifikasi

sentimen, seperti NB, SVM, dan LR. NB bekerja dengan asumsi adanya independensi antar fitur, yang dalam praktiknya sering tidak terpenuhi pada teks alami, meskipun tetap unggul dari sisi efisiensi komputasi (Zhang dkk., 2021). SVM efektif untuk data berdimensi tinggi karena mampu mencari pemisah optimal, namun memiliki keterbatasan pada kompleksitas komputasi ketika diterapkan pada dataset besar (Porcello, 2019). LR bekerja dengan memodelkan probabilitas kelas menggunakan fungsi logistik yang memetakan kombinasi linear dari fitur *input* ke rentang probabilitas antara 0 dan 1 (Glahn, 2014).

Di antara kandidat *classifier* yang tersedia, LR menawarkan keseimbangan optimal antara performa, efisiensi komputasi, dan keterjelasan hasil (Geisler dkk., 2023). Sebagai linear *classifier*, LR bekerja dengan baik pada fitur *sparse* maupun *dense*, sehingga memungkinkan perbandingan adil antar metode representasi teks tanpa bias dari arsitektur *classifier* (Silva dkk., 2025). Konsistensi performa LR pada berbagai jenis representasi menjadikannya pilihan yang tepat untuk mengisolasi pengaruh kualitas representasi dari pengaruh arsitektur *classifier*.

Kombinasi TF-IDF dengan LR dan kombinasi SBERT dengan LR masing-masing memiliki karakteristik yang perlu dipertimbangkan dalam menyeimbangkan akurasi, kemudahan pemahaman, dan efisiensi. Penelitian sebelumnya menunjukkan bahwa kombinasi TF-IDF dengan LR cenderung lebih stabil pada dataset kecil dengan jumlah sampel kurang dari 10.000. Kestabilan ini disebabkan oleh kemampuan TF-IDF dalam menangkap fitur diskriminatif tanpa beban komputasi besar, dengan pencapaian akurasi hingga 93,8% melampaui kombinasi BERT dengan LR yang mencapai 90,9% pada konteks yang sama (Gautam & Bansal, 2022). Sebaliknya, kombinasi BERT dengan LR lebih sesuai untuk dataset besar dan kompleks dengan jumlah sampel lebih dari atau sama dengan 10.000. Keunggulan ini bersumber dari kemampuannya memodelkan konteks secara lebih bernuansa dengan akurasi 84,73% dibandingkan TF-IDF dengan LR yang mencapai 83,73%, meskipun membutuhkan sumber daya komputasi lebih tinggi untuk proses *embedding* dan inferensi (Llaca-Sánchez dkk., 2025).

Penelitian ini bertujuan untuk membandingkan kombinasi TF-IDF dengan LR dan kombinasi SBERT dengan LR pada tugas ABSA, khususnya pada *Aspect Sentiment Classification* (ASC) yaitu menentukan polaritas sentimen terhadap aspek tertentu (Gomes dkk., 2022). Pembatasan pada ASC memungkinkan perbandingan yang lebih jelas antara

representasi *sparse* (TF-IDF) dan representasi *dense* (SBERT) dengan linear *classifier* yang sama (LR), sehingga pengaruh representasi terhadap performa dapat diisolasi tanpa variabel tambahan dari modul ekstraksi aspek. Untuk mendapatkan data ulasan yang representatif dan memandai, penelitian ini mengambil data ulasan dari platform *e-commerce* Shopee. Shopee merupakan platform *Cross-Border E-commerce* (CBEC) terbesar dengan rata-rata kunjungan bulanan sekitar 200 juta (Hermawan, 2020). Selain itu, survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) menunjukkan Shopee sebagai *e-commerce* paling sering diakses masyarakat Indonesia tahun 2025 pada proporsi 53,22% (APJII, 2025).

Perbandingan yang dilakukan memberi bukti mengenai karakteristik ulasan Shopee, apakah lebih bersifat *keyword-driven* yang menekankan kata kunci eksplisit, atau *context-driven* yang menekankan pemahaman konteks kalimat secara menyeluruh. Bukti tersebut menjadi dasar sistematis bagi eksperimen lanjutan untuk mengeksplorasi representasi *sparse* maupun *dense* lainnya sesuai kebutuhan analisis. Selain itu, hasil perbandingan model dan penyajian dataset ulasan Shopee berannotasi dapat menjadi acuan bagi penelitian selanjutnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, makalah ini dirumuskan untuk membahas perbandingan kinerja antara representasi teks *sparse* (TF-IDF) dan *dense* (SBERT) dalam mengklasifikasikan sentimen ulasan konsumen *e-commerce* Shopee. Analisis dilakukan menggunakan *classifier Logistic Regression* pada sub-tugas ABSA, yaitu ASC dengan aspek yang diperoleh secara *data-driven* dari ulasan konsumen.

1.3 Tujuan dan Manfaat Penelitian

Tujuan penelitian ini adalah membandingkan kinerja antara TF-IDF dan SBERT dengan *classifier Logistic Regression* dalam ASC pada ulasan *e-commerce* untuk aplikasi Shopee. Manfaat dari penelitian ini adalah memberi hasil kajian serta rekomendasi metodologis dalam menentukan representasi teks dan prosedur klasifikasi yang tepat untuk pada tugas ASC ulasan *e-commerce* untuk aplikasi Shopee.

1.4 Ruang Lingkup

Ruang Lingkup dalam penelitian ini adalah sebagai berikut:

1. Penelitian dilakukan menggunakan bahasa pemrograman Python dan perkakas Kaggle.

2. Data yang digunakan berupa ulasan teks pengguna aplikasi Shopee di *platform* Google Play Store dengan bahasa Indonesia yang diambil pada tanggal 17-21 Oktober 2025.
3. Aspek yang dianalisis didapatkan melalui proses identifikasi aspek dari 1000 sampel ulasan terbaru (17-18 Oktober 2025) dan bersifat *data-driven*.
4. Polaritas sentimen setiap aspek bersifat biner (P untuk positif atau N untuk negatif).
5. Lingkup ulasan yang dianalisis adalah ulasan teks.
6. Pembatasan tugas ABSA yang dikerjakan yaitu ASC, yang menerima masukan berupa ulasan dan aspek untuk menghasilkan prediksi sentimen pada aspek yang dimasukkan.

1.5 Sistematika Penulisan

Sistematika penulisan yang digunakan untuk penyusunan laporan tugas akhir, yaitu:

BAB I PENDAHULUAN

Bab ini menjelaskan latar belakang penelitian, rumusan masalah, batasan masalah, ruang lingkup, tujuan, manfaat, dan sistematika penulisan dalam penyusunan skripsi Perbandingan TF-IDF dan SBERT dengan *Logistic Regression* untuk ABSA pada ulasan *e-commerce* aplikasi Shopee.

BAB II TINJAUAN PUSTAKA

Bab ini menjelaskan studi pustaka mengenai teori yang berhubungan dengan pelaksanaan dan penyusunan skripsi. Studi pustaka yang dibahas pada bab ini adalah konsep dasar analisis sentimen, ABSA, pemilihan ulasan *e-commerce* Shopee, identifikasi aspek melalui *manual review* dan *hierarchical clustering*, pembentukan *final dataset* dengan anotasi manual dan kontrol kualitas dataset menggunakan *Kappa Score*, representasi teks *sparse* dan *dense*, TF-IDF, SBERT, *Logistic Regression* sebagai *classifier*, *grid search* dengan *k-fold Cross-Validation*, metrik evaluasi, dan *McNemar test*.

BAB III METODOLOGI PENELITIAN

Bab ini membahas langkah-langkah yang dilakukan dalam penelitian yang bertujuan untuk membandingkan TF-IDF dan SBERT dengan *classifier Logistic Regression* untuk ABSA pada ulasan *e-commerce* aplikasi Shopee yang meliputi pengumpulan data eksploratori, *preprocessing* data

eksploratori, identifikasi aspek melalui *hierarchical clustering* dan *manual review*, pengumpulan data terfokus dengan *seed keywords* dari hasil identifikasi aspek, anotasi manual dan kontrol kualitas menggunakan *Fleiss Kappa* dan *Cohen's Kappa* untuk pembentukan *final dataset*, *preprocessing final dataset* guna pemodelan, pengembangan model dengan TF-IDF dan SBERT dengan *classifier Logistic Regression*, pemilihan *hyperparameter* dengan *grid search hyperparameter tuning* menggunakan *k-fold Cross-Validation*, evaluasi model, dan *McNemar Test* untuk analisis komparatif.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan implementasi serta analisis hasil perbandingan dua pendekatan representasi teks, yaitu representasi *sparse* dan representasi *dense*, dalam tugas ABSA pada ulasan *e-commerce* Shopee. Pertama, bab ini diawali dengan pemaparan hasil fase identifikasi aspek, yang menjelaskan proses pembentukan aspek melalui *hierarchical clustering* dan *manual review*. Selanjutnya, disajikan hasil pembentukan *final dataset* melalui anotasi manual, yang kemudian dikontrol kualitasnya menggunakan *Fleiss Kappa* dan *Cohen's Kappa*. Setelah itu, bab ini memaparkan hasil *hyperparameter* terbaik untuk model TF-IDF dan SBERT melalui *grid search* dengan *k-fold Cross-Validation*, baik pada masing-masing aspek secara terpisah maupun secara lintas aspek dengan *hyperparameter* universal. Bagian berikutnya membahas perbandingan performa TF-IDF dan SBERT yang dikombinasikan dengan *Logistic Regression*, baik ketika menggunakan *hyperparameter* terbaik per aspek maupun ketika menggunakan *hyperparameter* universal. Hasil evaluasi komparatif diperkuat dengan analisis menggunakan *McNemar Test*. Bab ini ditutup dengan pembahasan komprehensif mengenai perbedaan karakteristik dan perilaku kedua model, serta penyajian ringkasan temuan utama.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari penjelasan hasil penelitian dan saran pengembangan penelitian ABSA untuk ulasan *e-commerce* bahasa Indonesia.