

Sentiment Analysis Of E-Commerce Reviews Using Fine-Tuned Indobert With Class Weights Strategy

Abdan Syakura^{*1}, Dewi Soyusiawaty²

^{1,2}Informatics, Faculty of Industrial Technology, Ahmad Dahlan University, Indonesia

Email: ¹abdansbgr0@gmail.com

Received : Jan 28, 2026; Revised : Feb 17, 2026; Accepted : Feb 21, 2026; Published : Jun 15, 2026

Abstract

MSMEs in the e-commerce sector face difficulties in converting large volumes of unstructured customer review data into actionable business insights. This challenge is exacerbated by the ambiguity of star ratings, which often do not align with the content of the reviews, making automated sentiment analysis of the text essential. This study implements a systematic sentiment analysis workflow on a case study of 15,278 customer reviews of Toko Pasar Stan Jogja. The method used is fine-tuning a pre-trained Transformer model, namely IndoBERT, which is optimized with class weighting techniques to handle unbalanced datasets. The model's performance was comprehensively evaluated using Accuracy, Precision, Recall, F1-Score, Confusion Matrix, and word cloud visualization metrics. The test results showed that the developed model had very high performance, achieving an overall accuracy of 96.99% and an average F1-Score of 0.97 on the test data. Qualitative analysis also successfully identified that product quality ("fresh") and logistics efficiency ("fast") were the main drivers of satisfaction, while the main complaints centered on the condition of the product upon arrival ("damaged," "rotten"). This research proves that the optimized Transformer model is not only effective for sentiment classification, but also serves as a strategic tool for extracting concrete business insights.

Keywords : *Deep Learning, IndoBERT, Sentiment Analysis, Tokopedia.*

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License



1. PENDAHULUAN

Perkembangan internet saat ini sudah sangat besar, mempengaruhi sektor jual beli, mengubah gaya hidup dan perilaku, seperti munculnya perdagangan elektronik. Dulu, untuk mendapatkan produk yang diinginkan, Anda harus pergi ke tempat penjualan produk tersebut. Masyarakat kini dapat dengan mudah memesan makanan, minuman, transportasi, dan hotel dari *smartphone* pribadinya [1]. Tokopedia adalah salah satu pasar C2C (konsumen-ke-konsumen) yang paling terkenal di era industri 4.0. Website ini merupakan tempat bertemunya penjual dan pembeli untuk memperdagangkan barang tanpa komisi [2]. Manfaat marketplace bagi pelanggan adalah memberikan peluang interaksi antar pelanggan untuk memberikan tambahan informasi pembelian melalui fitur review [3]. Ulasan memungkinkan pembeli untuk memberikan tanggapan produk dalam bentuk ulasan dan penilaian bintang [4].

Toko Pasar Stan Jogja merupakan salah satu toko online di Tokopedia yang menawarkan produk kebutuhan rumah tangga seperti Sembako, Daging Ayam dan Ikan, Sayuran, dan Buah - Buah yang terletak di Kab. Sleman, Yogyakarta. Berdasarkan hasil wawancara dengan pemilik toko Pasar Stan Jogja. Toko ini sudah berdiri 5 Tahun dengan penjualan offline, lalu yang 1 membuat pemilik toko beralih ke sistem penjualan online yaitu karena adanya wabah Covid-19. Untuk itu pemilik toko melakukan observasi terkait Sistem penjualan online, untuk konsumen yang tidak dapat keluar rumah karena wabah virus Covid. Ada 15 ribu banyaknya ulasan yang diterima pada penjualan online di tokopedia dengan contoh "Barang sesuai deskripsi, respon dan pengiriman cepat. Terimakasih" dan "Tidak sesuai di gambar/merk, kemasan

amburadul”, menganalisis sentimen secara manual membutuhkan waktu dan tenaga yang besar. Ketidakkampuan untuk mengidentifikasi dan mengatasi masalah dapat mengakibatkan penurunan kepuasan konsumen dan berakibat pada hilangnya konsumen, dan juga ulasan negatif dapat merusak reputasi toko online dan menghambat penjualan.

Analisis sentimen biasanya dilakukan dengan memberi label sentimen, seperti positif, negatif, atau netral, pada teks berdasarkan kontennya [5]. Pada tingkat dokumen, tujuannya adalah untuk menentukan sentimen keseluruhan dari satu teks utuh, seperti satu ulasan produk. Pada tingkat kalimat, setiap kalimat dianalisis secara terpisah untuk menangkap sentimen yang mungkin beragam dalam satu dokumen [6]. Analisis sentimen telah berevolusi dari metode berbasis leksikon dan *machine learning* klasik yang memiliki keterbatasan dalam memahami konteks, menuju era *deep learning* dengan arsitektur Transformer sebagai standar terkini (*state-of-the-art*) [7]. Model berbasis *Transformer*, seperti BERT (dan variannya IndoBERT) serta GPT, unggul dalam menangkap nuansa bahasa yang kompleks dan hubungan antar kata secara mendalam melalui mekanisme *attention* [8], mengatasi kelemahan model sebelumnya seperti RNN atau CNN [9]. Menjawab kebutuhan akan model bahasa yang andal untuk pemrosesan teks Bahasa Indonesia, tim peneliti yang terdiri dari akademisi dan praktisi industri (termasuk dari Universitas Indonesia, HKUST, dan Gojek) mengembangkan IndoBERT. Model ini merupakan implementasi arsitektur BERT yang dilatih dari awal (*pre-trained from scratch*) menggunakan korpus masif bernama Indo4B, yang terdiri dari sekitar 4 miliar kata dari sumber-sumber yang representatif dalam Bahasa Indonesia. Selain pemilihan model yang tepat, tantangan data tidak seimbang (*imbalanced data*) pada ulasan produk juga menjadi perhatian krusial yang dapat membiarkan prediksi model [10]. Penanganan *imbalanced data* pada level data bertujuan menyeimbangkan distribusi kelas secara fisik untuk mencegah bias model. Pendekatan ini meliputi teknik *oversampling* (seperti SMOTE) yang menambah sampel minoritas namun berisiko menyebabkan *overfitting*, serta teknik *undersampling* yang mengurangi sampel mayoritas tetapi memiliki kelemahan berupa potensi hilangnya informasi berharga. Oleh karena itu, penelitian ini mengadopsi pendekatan modern pada level algoritma menggunakan teknik *class weights*, yang memodifikasi proses pembelajaran agar model lebih sensitif terhadap kelas minoritas tanpa mengubah distribusi data asli, guna menghasilkan klasifikasi yang akurat, adil, dan mendalam [11].

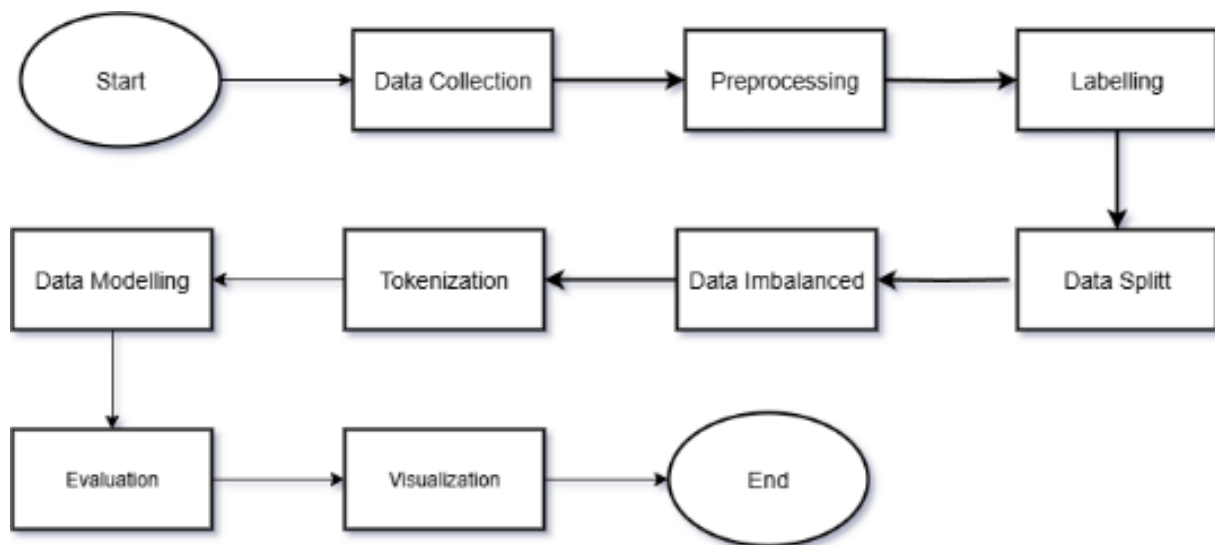
Penelitian terdahulu secara konsisten telah membuktikan kapabilitas model berbasis Transformer, khususnya IndoBERT, sebagai metode *state-of-the-art* untuk berbagai tugas klasifikasi teks berbahasa Indonesia. Meskipun keunggulan model IndoBERT dalam menangani teks berbahasa Indonesia telah banyak divalidasi, optimasi performa pada dataset tidak seimbang masih menjadi tantangan yang terbuka. Penelitian terbaru oleh Ihalauw dkk [12], mengevaluasi penggunaan teknik *oversampling* SMOTE pada IndoBERT untuk klasifikasi tiga kelas sentimen politik, namun hanya mencapai akurasi sebesar 87%. Di sisi lain, Qolbu dkk [13], membandingkan performa model tradisional dengan IndoBERT menggunakan strategi pembobotan kelas (*class weighting*), akan tetapi penelitian tersebut membatasi analisis pada dua kelas sentimen (positif dan negatif) dan mengecualikan kelas netral karena dianggap ambigu. Hal ini meninggalkan celah penelitian mengenai efektivitas strategi *class weights* pada model Transformer untuk klasifikasi multi-kelas (tiga kelas) yang mencakup ulasan netral, khususnya pada domain spesifik ulasan produk UMKM lokal. Penelitian ini hadir untuk mengisi celah tersebut dengan mengintegrasikan strategi *class weights* pada model IndoBERT untuk menghasilkan klasifikasi sentimen tiga kelas yang lebih akurat, adil, dan mampu menangkap nuansa evaluatif pelanggan secara menyeluruh.

Kebaruan (*novelty*) penelitian ini terletak pada integrasi model IndoBERT yang di-*fine-tune* dengan optimasi strategi *class weights* yang diterapkan secara spesifik pada karakteristik ulasan pasar lokal yang sangat tidak seimbang. Berbeda dengan pendekatan *oversampling* atau *undersampling* yang berisiko mengubah distribusi informasi asli, penggunaan *class weights* memodifikasi *loss function* agar model lebih sensitif terhadap ulasan negatif tanpa mengorbankan integritas data mentah.

Penelitian ini bertujuan untuk mengimplementasikan alur kerja analisis sentimen yang akurat dan adil menggunakan IndoBERT dan *class weights*. Evaluasi performa dilakukan secara komprehensif menggunakan metrik akurasi, *confusion matrix*, presisi, *recall*, dan *F1-Score* untuk memastikan model memberikan landasan kuat bagi pengambilan keputusan strategis UMKM dalam meningkatkan kualitas produk dan layanan.

2. METODE

Penelitian ini mengusulkan sebuah alur kerja yang sistematis untuk analisis sentimen pada ulasan produk. Alur kerja ini sepenuhnya memanfaatkan arsitektur *Deep Learning Transformer*, yang diimplementasikan melalui model IndoBERT, untuk menganalisis teks secara kontekstual. Tujuan dari pendekatan ini adalah untuk menghasilkan klasifikasi sentimen yang akurat serta wawasan kualitatif yang mendalam mengenai persepsi pelanggan.



Gambar 1. Alur Penelitian

2.1. Data Collection

Tahapan pengumpulan sumber data primer disini dilakukan dengan observasi langsung pada aplikasi tokopedia, lalu data disimpan menjadi dataset pada aplikasi *Excel* sebelum dilakukan pemrosesan. Data scraping merupakan suatu teknik pengumpulan data yang bertujuan untuk mengekstraksi informasi dari situs web atau sistem tertentu secara sistematis. Secara umum, data extraction digunakan dalam berbagai aktivitas yang berhubungan dengan pemasaran digital, salah satunya adalah riset konten [14]. Data yang dikumpulkan disini yaitu berupa ulasan pada produk Pasar Stan Jogja. Analisis sentimen ini berfokus pada akuisisi data primer. Sebanyak 15.278 ulasan pelanggan berhasil dikumpulkan dari halaman produk Toko Pasar Stan Jogja di platform *e-commerce* Tokopedia. Proses pengumpulan data dilakukan secara intensif dalam rentang waktu satu minggu, dimulai dari tanggal 5 Mei 2025. Dari periode tersebut, berhasil dikumpulkan dataset mentah sebanyak 15.278 ulasan pelanggan. Data ini menjadi dataset mentah yang akan diproses pada tahap selanjutnya.

2.2. Preprocessing

Data mentah yang tidak terstruktur kemudian melalui serangkaian proses pembersihan untuk meningkatkan kualitasnya. Tahapan ini meliputi case folding (mengubah teks menjadi huruf kecil), pembersihan dari noise seperti URL dan tanda baca, serta normalisasi teks untuk menyeragamkan kata-kata tidak baku menjadi bentuk formalnya [15].

2.3. Labelling

Karena data ulasan tidak memiliki label sentimen, dilakukan proses pelabelan otomatis menggunakan metode berbasis *leksikon*. Setiap ulasan yang telah dibersihkan diklasifikasikan ke dalam label 'positif', 'negatif', atau 'netral' berdasarkan dominasi kata-kata positif dan negatif di dalamnya. Pendekatan ini bekerja dengan memanfaatkan kamus sentimen, yaitu sebuah daftar kata yang telah diberi skor polaritas positif atau negatif sebelumnya. Meskipun metode ini cepat dan tidak memerlukan data latih berlabel, kelemahannya adalah ketidakmampuannya dalam memahami konteks atau kata-kata baru yang tidak ada dalam kamus. Meskipun demikian, pendekatan berbasis leksikon tetap menjadi teknik yang sangat berguna untuk menghasilkan baseline atau label awal pada dataset besar yang tidak berlabel [16]. Setelah data ulasan dibersihkan, dataset mentah yang terdiri dari 15.278 ulasan tersebut masih belum memiliki label sentimen. Untuk melatih model supervised learning seperti IndoBERT, diperlukan data berlabel yang berfungsi sebagai ground truth atau "kunci jawaban".

2.4. Data Splitting

Dalam metodologi *Supervised Machine Learning*, pembagian dataset menjadi beberapa bagian independen merupakan langkah fundamental untuk memastikan validitas dan objektivitas evaluasi model. Dataset yang telah bersih dan berlabel dibagi menjadi tiga bagian dengan proporsi 70:15:15, menghasilkan 10.695 data latih, 2.292 data validasi, dan 2.291 data uji untuk memastikan evaluasi yang objektif [17].

2.5. Data Imbalanced

Pada tahap ini, dilakukan analisis terhadap distribusi kelas sentimen pada data latih dan ditemukan bahwa jumlah data tidak seimbang. Untuk mengatasi hal ini, diterapkan teknik *Class Weights* dengan bobot tensor ([0.5888, 2.4233, 1.1249]) untuk kelas 'positif', 'negatif', dan 'netral', guna memastikan model belajar secara adil [18].

2.6. Tokenization

Sebelum dimasukkan ke dalam model, setiap ulasan diubah menjadi format numerik melalui proses Tokenisasi. Proses ini menggunakan *tokenizer* dari model IndoBERT untuk memecah kalimat menjadi token dan mengonversinya menjadi ID numerik yang dapat dipahami oleh model [19].

2.7. Data Modelling

Tahap inti di mana model IndoBERT dilatih kembali (*fine-tuning*) menggunakan data latih yang telah dioptimalkan dengan *class weights*. Proses pelatihan ini dikonfigurasi dengan hiperparameter spesifik: *batch size* sebesar 16, *optimizer AdamW*, dan *learning rate* $5e-5$. Berdasarkan hasil pemantauan pada data validasi, ditemukan bahwa model mencapai performa generalisasi terbaik (ditandai dengan *validation loss* terendah) pada epoch ke-2, sehingga model dari epoch inilah yang dipilih untuk evaluasi akhir [20].

2.8. Evaluation

Setelah proses pelatihan selesai, model dengan performa terbaik pada data validasi dipilih untuk menjalani Evaluasi Akhir. Evaluasi ini dilakukan pada data uji untuk mendapatkan penilaian yang objektif terhadap kemampuan generalisasi model. Kinerja model diukur secara kuantitatif menggunakan serangkaian metrik standar *confusion matrix* Untuk memvisualisasikan performa dan pola kesalahan model [21]. Model terbaik yang telah dilatih kemudian diuji kinerjanya pada 2.292 data uji. Hasil evaluasi menunjukkan performa yang sangat tinggi, dengan pencapaian akurasi keseluruhan sebesar 96.99% dan *F1-Score* rata-rata 0.97.

2.9. Visualization

Dengan melengkapi analisis kuantitatif dan menyajikan wawasan yang lebih intuitif, hasil penelitian kemudian divisualisasikan. Tahap Visualisasi ini mencakup pembuatan *Word Cloud* dari ulasan yang diklasifikasikan sebagai positif dan negatif. Visualisasi ini bertujuan untuk mengidentifikasi kata-kata kunci dan tema utama yang menjadi pendorong kepuasan maupun sumber keluhan pelanggan. Untuk mengekstrak wawasan kualitatif, hasil klasifikasi pada data uji kemudian divisualisasikan menggunakan *word cloud* [22]. Hasil visualisasi berhasil mengidentifikasi faktor kepuasan utama seperti 'segar' dan 'cepat' pada ulasan positif, serta sumber keluhan utama seperti 'rusak' dan 'lama' pada ulasan negatif.

3. HASIL

Bab ini menyajikan hasil implementasi dan analisis mendalam dari penelitian yang telah dilakukan. Pembahasan diawali dengan gambaran umum dataset hasil pra-pemrosesan untuk memberikan konteks pada data yang digunakan. Selanjutnya, akan dipaparkan hasil dari eksperimen pelatihan model IndoBERT yang telah dioptimalkan menggunakan teknik *class weights* untuk menangani ketidakseimbangan data. Fokus utama bab ini adalah evaluasi performa model secara kuantitatif menggunakan metrik standar (akurasi, presisi, *recall*, *F1-Score*) serta analisis kualitatif melalui visualisasi *word cloud* untuk mengekstrak wawasan bisnis yang konkret bagi Toko Pasar Stan Jogja.

3.1. Data Collection

Pada pengumpulan data yang diambil dari platform Tokopedia dengan jumlah data sebanyak 15278 ulasan yang berhasil dikumpulkan, data ini berfungsi sebagai jembatan untuk proses Prapemrosesan.

Tabel 1. Hasil Pengumpulan Data

No	Ulasan
1	tiba minggu pagi. fresh segar kayak blonjoan pagi. Thx
2	pengiriman on time, pesanan sudah sesuai.
3	maaf baru kasih ulasan sekarang, berasnya saya terima dengan selamat. makasih seller. semoga laris terus yaa
4	TIDAK AMANAH ! YG DIKIRIM PRODUK BEDA DG GAMBAR DAN DESKRIPSI
5	sesuai pesananseller gercep...mantab
6	mantap seller sangat komunikatif pengiriman cepat
7	banyak yg tidak bagus , pdahal sebelum nya aman aman saja , banyak layu
8	sesuai deskripsi, seller gercep dan ramah. moga2 cepet jadi kombuchanya
9	barang busuk dikirim, cara bungkus asal. mana kotor masih ada tanah lgsg dikirim. minimal dicuci bos
10	Koq packingannya gak seperti kemrin kmarin tp after all barang nya msh seger

Tabel 1 dapat diobservasi bahwa data ulasan mentah sangat bervariasi dan mengandung banyak noise, seperti singkatan, kesalahan ketik, dan penggunaan tanda baca yang tidak standar. Untuk memastikan model dapat mengolah data secara efektif, diperlukan tahapan pra-pemrosesan. Proses pembersihan teks ini akan dibahas pada bagian selanjutnya.

3.2. Preprocessing

Dari Tabel 2 menyajikan hasil dari preprocessing untuk meningkatkan kualitas data input, serangkaian teknik pra-pemrosesan teks diterapkan. Tahapan ini meliputi pembersihan teks dasar (cleaning), normalisasi kata-kata tidak baku, hingga proses stemming untuk mengubah kata menjadi bentuk dasarnya.

Tabel 2. Hasil Preprocessing Data

No	Ulasan	Preprocessing		
		Clean Text	Normalization Text	Tokenizer Text
1	tiba minggu pagi. fresh segar kayak blonjoan pagi. Thx	tiba minggu pagi. fresh segar kayak blonjoan pagi thx.	Tiba minggu pagi fresh segar kayak belanjaan pagi terima kasih	“minggu” “pagi” “fresh” “segar” “kayak” “belanjaan” “pagi” “terima” “kasih”
2	pengiriman on time, pesanan sudah sesuai.	pengiriman on time, pesanan sudah sesuai.	kirim on time, pesanan sudah sesuai.	“kirim” ”on” ”time” “pesanan” “sudah” “sesuai”
3	maaf baru kasih ulasan sekarang, berasnya saya terima dengan selamat. makasih seller. semoga laris terus yaa	maaf baru kasih ulasan sekarang, beras saya terima dengan selamat. makasih seller. semoga laris terus yaa	maaf baru kasih ulasan sekarang, beras saya terima dengan selamat. terima kasih seller. semoga laris terus ya	“maaf” “baru” “kasih” “ulasan” “ulasan” “sekarang” “beras” “saya” “terima” “dengan” “selamat” “makasih” “seller” “semoga” “laris” “terus” “yaa”
4	TIDAK AMANAH ! YG DIKIRIM PRODUK BEDA DG GAMBAR DAN DESKRIPSI !! SENGAJA GAK SY RETUR TAPI SY KASIH B.1 BIAR YG LAIN BS BACA YG SY PESAN TEH POCI VANILA 25 KANTONG SESUA...	tidak amanah ! yang dikirim produk beda dengan gambar dan deskripsi !! sengaja tidak saya retur tapi saya kasih b.1 biar yang lain bs baca yang saya pesan teh poci vanila 25 kantong sesua...	tidak amanah ! yang dikirim produk beda dengan gambar dan deskripsi !! sengaja tidak saya retur tapi saya kasih bintang 1 biar yang lain bisa baca yang saya pesan teh poci vanila 25 kantong sesuai...	“tidak” “amanah” “yang” “dikirim” “produk” “beda” “dengan” “gambar” “dan” “deskripsi” “sengaja” “tidak” “saya” “retur” “tapi” “saya” “kasih” “bintang 1” “biar” “yang” “lain” “bisa” “baca” “yang” “saya” “pesan” “teh” “poci” “vanilla 25” “kantong” “sesuai”
5	sesuai pesananseller gercep...mantab	sesuai pesananseller gercep...mantab	sesuai pesan seller... gerak cepat... mantap	“sesuai” “pesan” “seller” “gerak cepat” “mantap”

3.3. Labelling

Penelitian ini menggunakan metode pelabelan berbasis leksikon (lexicon-based). Pendekatan ini bekerja dengan cara mencocokkan kata-kata dalam ulasan dengan dua daftar kata yang telah didefinisikan sebelumnya: sebuah leksikon kata-kata positif dan sebuah leksikon kata-kata negatif.

Tabel 3. Hasil Labelling Data

Sentimen	Jumlah Ulasan	Persentase
Positive	8.647	56,6%
Neutral	4.522	29,6%
Negative	2.108	13,8%
Total	15.278	100%

Terlihat jelas bahwa data hasil pelabelan Tabel 3 bersifat tidak seimbang (imbalanced). Kelas 'positive' merupakan kelas mayoritas yang mendominasi hampir 60% dari seluruh dataset, sementara kelas 'negative' adalah kelas minoritas.

3.4. Data Splitting

Untuk membangun dan mengevaluasi model secara objektif, keseluruhan dataset yang berjumlah 15.278 ulasan dibagi menjadi tiga set data independen dengan proporsi 70% untuk pelatihan, 15% untuk validasi, dan 15% untuk pengujian. Langkah ini krusial untuk memastikan model dapat melakukan generalisasi pada data baru. Distribusi data setelah pembagian adalah sebagai berikut:

- Data latih terdiri dari 10654 ulasan. Set data ini merupakan bagian terbesar dan digunakan secara langsung untuk "mengajarkan" model transformer IndoBERT dalam mengenali pola-pola yang berkaitan dengan sentimen positif, negatif, dan netral.
- Data validasi terdiri dari 2292 ulasan. Set data ini digunakan selama proses pelatihan untuk memantau performa model di setiap epoch. Hasil evaluasi pada data validasi menjadi acuan untuk mencegah *overfitting* dan membantu dalam proses penyetelan hiperparameter.
- Data uji terdiri dari 2292 ulasan. Set data ini "disembunyikan" dan tidak pernah digunakan selama proses pelatihan maupun validasi. Data ini hanya digunakan satu kali pada tahap akhir untuk memberikan penilaian final yang paling objektif terhadap performa model yang telah selesai dilatih.

3.5. Tokenization

Proses ini merupakan langkah fundamental untuk mengubah data teks kualitatif menjadi representasi numerik yang dapat diproses oleh model Transformer. Dalam penelitian ini, digunakan AutoTokenizer yang spesifik untuk model pra-terlatih yang dipilih, yaitu indobenchmark/indobert-base-p1. Penggunaan tokenizer yang bersesuaian dengan modelnya sangat krusial karena keduanya berbagi kamus kosakata (*vocabulary*) yang sama, sehingga pemetaan kata ke ID token menjadi konsisten.

Tabel 4. Hasil Tokenisasi Model

Proses	Sebelum	Sesudah
Tokenisasi	mantaap.....bawang baguss.... tolong kualitasnya. langgaanan...	['[CLS]', 'mant', '###aaa', '##p', '!', '!', '!', '!', '!', 'bawang', 'bagus', '##s', '!', '!', '!', '!', 'tolong', 'dijaga', 'kualitasnya', '!', 'untuk', 'langg', '##aan', '##an', '!', '!', '!', '[SEP]', '[PAD]']
Encode	['[CLS]', 'mant', '###aaa', '##p', '!', '!', '!', '!', '!', 'bawang', 'bagus', '##s', '!', '!', '!', '!', 'tolong', 'dijaga', 'kualitasnya', '!', 'untuk', 'langg', '##aan', '##an', '!', '!', '!', '[SEP]', '[PAD]']	[2, 4142, 10097, 30367, 30470, 30470, 30470, 30470, 30470, 5374, 1305, 30362, 30470, 30470, 30470, 30470, 3854, 10286, 11128, 30470, 90, 19052, 160, 5, 30470, 30470, 30470, 3, 0]
Attention Mask	[2, 4142, 10097, 30367, 30470, 30470, 30470, 30470, 30470, 5374, 1305, 30362, 30470, 30470, 30470, 30470, 3854, 10286, 11128, 30470, 90, 19052, 160, 5, 30470, 30470, 30470, 3, 0]	[1, 0]

3.6. Data Modelling

Pada tahap pemodelan, arsitektur yang digunakan adalah IndoBERT yang diambil dari model pra-terlatih (pre-trained model) dari Hugging Face. Model ini kemudian diadaptasi untuk tugas klasifikasi tiga sentimen (positif, negatif, netral) dengan mengonfigurasi ulang lapisan klasifikasi akhirnya untuk memiliki tiga output (num_labels=3).

Tabel 5. Eksperimen Hyperparameter

Model	Parameter	Nilai yang Dicoba	Hasil Terbaik
BERT	Epochs	3, 5, 10	5
	Learning Rate	4e-5, 5e-5	5e-5
	Batch Size	16, 32, 64	32
	Max Sequence Length		128
	Length	128, 256, 512	
	Data Split	80:10:10, 70:15:15	70:15:15

Berdasarkan hasil dari serangkaian eksperimen hyperparameter tuning yang telah dipaparkan, konfigurasi final untuk model penelitian ini telah ditetapkan, seperti yang diringkas pada Tabel 5. Model dilatih selama 5 epoch dengan learning rate 5e-5 dan batch size 32, karena kombinasi ini terbukti menghasilkan akurasi validasi tertinggi. Selain itu, panjang sekuens maksimum (Max Sequence Length) ditetapkan pada 128 token karena sudah mencukupi untuk sebagian besar ulasan, dan rasio pembagian data (Data Split) yang digunakan untuk laporan akhir adalah 70:15:15. Seluruh hasil akhir penelitian yang disajikan dalam bab ini didasarkan pada model yang dilatih menggunakan konfigurasi optimal tersebut.

Hasil perhitungan class weights menghasilkan sebuah tensor [0.5888, 2.4233, 1.1249] yang berfungsi sebagai faktor penalti selama proses pelatihan model. Setiap nilai dalam tensor ini secara berurutan merepresentasikan bobot untuk kelas 'positif' (indeks 0), 'negatif' (indeks 1), dan 'netral' (indeks 2). Nilai bobot yang paling rendah (0.5888) pada kelas 'positif' menandakan bahwa ini adalah kelas mayoritas, sementara bobot tertinggi (2.4233) pada kelas 'negatif' menunjukkan bahwa ini adalah kelas minoritas yang jumlah datanya paling sedikit. Secara praktis, saat model membuat kesalahan pada ulasan yang seharusnya 'negatif', nilai kesalahannya (loss) akan diperbesar 2.4 kali lipat. Pemberian "hukuman" yang lebih berat ini secara efektif memaksa model untuk lebih memperhatikan dan belajar lebih keras dalam mengenali pola dari kelas minoritas, dengan tujuan akhir untuk menghasilkan model yang lebih adil dan seimbang.

Beberapa hiperparameter kunci ditetapkan untuk proses ini: pelatihan dilakukan selama 5 epoch dengan ukuran batch sebesar 32. *Optimizer* yang dipilih adalah AdamW, yang merupakan standar *de-facto* untuk model Transformer, dengan *learning rate* awal sebesar 5e-5. Untuk membantu konvergensi model, diterapkan pula penjadwal *learning rate linear* yang akan menyesuaikan laju pembelajaran secara dinamis selama proses pelatihan berlangsung.

	train_loss	train_accuracy	val_loss	val_accuracy
1	0.394547	0.852347	0.255483	0.906632
2	0.163559	0.946325	0.167151	0.945899
3	0.098181	0.967552	0.158198	0.953316
4	0.05724	0.979241	0.152882	0.954625
5	0.030968	0.988592	0.161938	0.958551

Gambar 2. Hasil Pelatihan Data

Gambar 2 hasil dari proses pelatihan model, sebagaimana dirangkum dalam tabel, menunjukkan sebuah kurva pembelajaran yang sangat baik dan efektif selama lima epoch. Akurasi pada data latih

(*train_accuracy*) menunjukkan peningkatan yang konsisten dan signifikan pada setiap epoch, dimulai dari 85.2% dan mencapai 98.8% pada epoch kelima. Hal ini mengindikasikan bahwa model berhasil mempelajari pola-pola yang relevan dari data yang diberikan kepadanya dengan sangat baik. Lebih penting lagi, akurasi pada data validasi (*val_accuracy*), yang merupakan indikator utama kemampuan generalisasi model pada data baru, juga menunjukkan performa yang sangat kuat. Akurasi validasi meningkat tajam pada dua epoch pertama dan kemudian stabil di tingkat yang sangat tinggi, mencapai puncaknya sebesar 95.8% pada epoch kelima. Meskipun terdapat sedikit celah (*gap*) antara akurasi latihan dan validasi pada epoch akhir yang mengindikasikan adanya *overfitting* ringan, tingginya angka akurasi validasi yang stabil membuktikan bahwa model yang dilatih memiliki kemampuan yang sangat baik untuk melakukan prediksi pada data baru yang belum pernah dilihat sebelumnya.

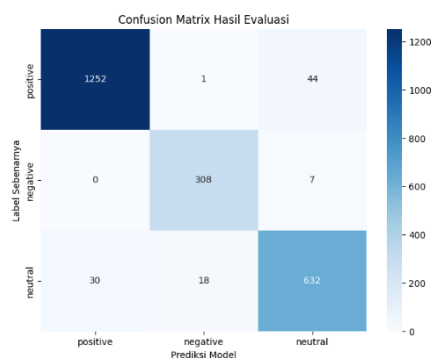
3.7. Evaluation

Penelitian ini memasuki tahap evaluasi akhir, tahap ini bertujuan untuk mengukur performa sejati dan kemampuan generalisasi model pada data yang sepenuhnya baru. Oleh karena itu, model terbaik tersebut akan diuji menggunakan data uji, yang telah disimpan terpisah dan tidak pernah terlibat dalam proses pelatihan maupun validasi. Performa model akan diukur secara kuantitatif menggunakan serangkaian metrik standar, termasuk *confusion matrix*, akurasi, presisi, *recall*, dan *F1-Score*, yang akan dipaparkan pada bagian berikut.

Tabel 6. Hasil Klasifikasi

	Precision	Recall	F1-Score	Support
Negative	0.96	0.97	0.97	315
Neutral	0.95	0.95	0.95	680
Positive	0.98	0.98	0.98	1297
Accuracy			0.97	2292
Macro Avg	0.97	0.97	0.97	2292
Weighted Avg	0.97	0.97	0.97	2292

Tabel 6 di atas menyajikan laporan klasifikasi akhir yang merangkum performa model pada 2.292 data uji. Secara keseluruhan, model berhasil mencapai akurasi sebesar 97%, yang mengindikasikan kemampuan generalisasi yang sangat kuat pada data baru. Analisis lebih mendalam pada laporan klasifikasi menunjukkan bahwa performa model tidak hanya tinggi, tetapi juga sangat seimbang di ketiga kelas sentimen. Kelas Positif menunjukkan kinerja terbaik dengan *F1-Score* 0.98, didukung oleh nilai *precision* dan *recall* yang sama tingginya. Ini membuktikan model sangat andal dalam mengidentifikasi ulasan positif. Kinerja pada kelas negatif dan netral juga sangat kuat, dengan *F1-Score* masing-masing 0.97 dan 0.95. Nilai *macro avg* dan *weighted avg* yang identik (0.97) semakin mengonfirmasi bahwa performa tinggi model bersifat konsisten di semua kelas, terlepas dari perbedaan jumlah data (*support*).



Gambar 3. Hasil Confusion Matrix

terbatas. Hal ini menghasilkan model yang adil (*fair*) karena memberikan performa seimbang di seluruh kelas sentimen.

Jika dibandingkan dengan penelitian terdahulu, pendekatan ini menawarkan stabilitas performa yang lebih baik pada data tidak seimbang. Penelitian oleh Merdiansah dkk [23], yang menggunakan IndoBERT standar tanpa optimasi pembobotan kelas menunjukkan bahwa meskipun akurasi global tinggi, deteksi pada kelas minoritas tetap menjadi tantangan. Sementara itu, dibandingkan dengan metode konvensional seperti SVM yang digunakan oleh Fadilah dkk [24], yang memperoleh akurasi sebesar 92% pada kernel linier, penggunaan IndoBERT dalam studi ini memberikan keunggulan signifikan dengan selisih akurasi mencapai 4,99%. Hal ini membuktikan keunggulan arsitektur Transformer dalam ekstraksi fitur otomatis tanpa perlu *feature engineering* yang rumit. Penggunaan *class weights* dalam penelitian ini juga sejalan dengan temuan Mujahid dkk [25], yang menyatakan bahwa penanganan *imbalanced data* pada tingkat algoritma secara signifikan meningkatkan nilai *Recall* tanpa mengorbankan akurasi keseluruhan. Hal ini dibuktikan dalam penelitian ini dengan perolehan nilai *Recall* yang seimbang di angka 0,96 - 0,97 untuk semua kelas (Positif, Netral, dan Negatif). Integrasi ini melengkapi keterbatasan model Transformer standar dalam menangani distribusi kelas yang sangat jomplang pada dataset ulasan produk lokal.

Urgensi penelitian ini terletak pada perlunya digitalisasi sistem evaluasi bagi UMKM lokal seperti Toko Pasar Stan Jogja. Di tengah persaingan *e-commerce* yang ketat, reputasi digital ditentukan oleh ulasan pelanggan. Dampak nyata dari penelitian ini adalah tersedianya model otomatis yang mampu mengekstrak wawasan bisnis secara instan. Hasil analisis *word cloud* mengungkapkan bahwa kepuasan pelanggan sangat bergantung pada aspek kesegaran ('segar', 'bagus'), sementara keluhan berfokus pada masalah fisik produk ('pecah', 'kemasan'). Dengan model ini, pemilik UMKM dapat melakukan deteksi dini terhadap penurunan kualitas layanan secara *real-time* tanpa harus membaca ribuan ulasan secara manual, yang pada akhirnya berdampak pada efisiensi operasional dan peningkatan retensi pelanggan.

Terlepas dari hasil yang dicapai, penelitian ini memiliki beberapa keterbatasan. Pertama, dataset yang digunakan bersifat spesifik pada satu jenis toko (*fresh products*), sehingga model mungkin memerlukan *re-training* atau penyesuaian parameter jika diterapkan pada domain produk yang berbeda (misalnya elektronik atau pakaian) yang memiliki karakteristik bahasa berbeda. Kedua, penelitian ini belum mengeksplorasi penggunaan teknik augmentasi data seperti *back-translation* untuk memperkaya kelas minoritas, yang berpotensi meningkatkan performa lebih jauh. Terakhir, model ini sangat bergantung pada sumber daya komputasi tinggi untuk proses *fine-tuning*, yang mungkin menjadi kendala bagi implementasi langsung pada perangkat dengan spesifikasi rendah.

5. KESIMPULAN

Bab ini menyajikan kesimpulan yang ditarik dari hasil analisis dan pembahasan pada bab-bab sebelumnya. Selain itu, akan dipaparkan pula keterbatasan yang ada dalam penelitian ini serta saran yang dapat menjadi acuan, baik untuk pengembangan penelitian sejenis di masa depan maupun untuk perbaikan pada objek studi.

1. Penelitian ini telah berhasil mengembangkan sebuah alur kerja untuk analisis sentimen pada ulasan produk Toko Pasar Stan Jogja menggunakan arsitektur *Transformers* IndoBERT. Proses pengembangan meliputi beberapa tahapan kunci, yaitu pengumpulan dan pembersihan data mentah, pra-pemrosesan teks yang mencakup normalisasi dan *stemming*, pelabelan sentimen awal menggunakan metode berbasis *leksikon*, hingga implementasi *fine-tuning* pada model *pretrained* IndoBERT.
2. Model IndoBERT yang dioptimalkan menunjukkan kinerja yang sangat tinggi dengan akurasi keseluruhan sebesar 96,99% dan *weighted F1-Score* sebesar 0,97. Hal ini mengonfirmasi bahwa pendekatan *fine-tuning* Transformer yang dikombinasikan dengan penanganan data tidak seimbang

- sangat efektif, akurat, dan andal untuk mengklasifikasikan sentimen ulasan produk dalam bahasa Indonesia yang bersifat informal.
3. Urgensi penelitian ini terletak pada penyediaan solusi digital otomatis bagi UMKM untuk memahami persepsi pelanggan secara cepat di tengah pertumbuhan *e-commerce*. Dampak signifikannya adalah pemilik usaha dapat mengidentifikasi masalah kritis secara *real-time*—seperti isu kesegaran produk dan kendala logistik yang ditemukan dalam penelitian ini—tanpa perlu melakukan survei manual yang memakan waktu. Implementasi model ini memberikan basis data strategis bagi pelaku UMKM dalam meningkatkan kualitas layanan dan daya saing di pasar digital.
 4. Keterbatasan penelitian ini membuka peluang untuk pengembangan lebih lanjut. Saran untuk penelitian mendatang meliputi. Membandingkan performa IndoBERT dengan model-model arsitektur Transformer yang lebih baru atau model yang spesifik untuk domain *e-commerce* dapat menjadi area penelitian yang menarik. Pengembangan aplikasi atau *dashboard* monitoring sentimen berbasis *web* atau *mobile* yang mengintegrasikan model ini agar dapat digunakan secara langsung oleh pelaku bisnis.

DAFTAR PUSTAKA

- [1] R. Anadra, H. Wijayanto, and K. Sadik, “Sentiment Analysis of Tokopedia Customer Reviews Using BiLSTM and IndoBERT with Comparative Analysis of Preprocessing and Labeling Methods,” *International Journal of Advances in Data and Information Systems*, vol. 6, no. 3, pp. 773–788, Dec. 2025, doi: 10.59395/ijadis.v6i3.1458.
- [2] Wahyu Widyananda, Maskur, and Ahmad Fauzi, “Machine Learning and Transformer-based Model for Sentiment Analysis of Indonesian E-Commerce Reviews,” *The Indonesian Journal of Computer Science*, vol. 14, no. 4, Aug. 2025, doi: 10.33022/ijcs.v14i4.4980.
- [3] D. Purnamasari, A. B. Aji, S. Madenda, I. M. Wiryana, and S. Harmanto, “SENTIMENT ANALYSIS METHODS FOR CUSTOMER REVIEW OF INDONESIA E-COMMERCE,” *International Journal of Innovative Computing, Information and Control*, vol. 20, no. 1, pp. 47–60, Feb. 2024, doi: 10.24507/ijicic.20.01.47.
- [4] K. Ilmiyah and I. Krishernawan, “Pengaruh Ulasan Produk, Kemudahan, Kepercayaan, Dan Harga Terhadap Keputusan Pembelian Pada Marketplace Shopee Di Mojokerto,” *Maker: Jurnal Manajemen*, vol. 6, no. 1, pp. 31–42, 2020, doi: 10.37403/mjm.v6i1.143.
- [5] C. Raffel *et al.*, “Exploring the limits of transfer learning with a unified text-to-text transformer,” *Journal of Machine Learning Research*, vol. 21, pp. 1–67, 2020.
- [6] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, “Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN,” *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 348–354, Dec. 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.
- [7] M. Naz, H., Ahuja, S., & Kumar, “Recent Trends in Deep Learning Based Sentiment Analysis,” *SN Comput. Sci.*, vol. 4, 2023.
- [8] N. Patwardhan, S. Marrone, and C. Sansone, “Transformers in the Real World: A Survey on NLP Applications,” *Information (Switzerland)*, vol. 14, no. 4, 2023, doi: 10.3390/info14040242.
- [9] Wahyu Widyananda, Maskur, and Ahmad Fauzi, “Machine Learning and Transformer-based Model for Sentiment Analysis of Indonesian E-Commerce Reviews,” *The Indonesian Journal of Computer Science*, vol. 14, no. 4, Aug. 2025, doi: 10.33022/ijcs.v14i4.4980.
- [10] and P. I. D. A. Indrawati, L. Ilmu, P. Indonesia, “Penerapan Teknik Kombinasi Oversampling Dan Undersampling Hybrid Oversampling and Undersampling Techniques To Handling Imbalanced Dataset,” vol. 4, no. no. doi: 10.33387/jiko., pp. 38–43, 2021.
- [11] B. BAKIRARAR and A. H. ELHAN, “Class Weighting Technique to Deal with Imbalanced Class Problem in Machine Learning: Methodological Research,” *Turkiye Klinikleri Journal of Biostatistics*, vol. 15, no. 1, pp. 19–29, 2023, doi: 10.5336/biostatic.2022-93961.
- [12] S. A. Ihalauw, N. Trezandy Lapatta, D. Wiria Nugraha, Wirdayanti, and C. Ar Lamasitudju, “Analisis Sentimen Terhadap Kinerja Awal Pemerintahan Menggunakan IndoBERT Dan SMOTE

- Pada Media Sosial X,” *Jurnal Algoritma*, vol. 22, no. 2, Nov. 2025, doi: 10.33364/algoritma/v.22-2.2957.
- [13] A. Anas Qolbu and N. Fitriyati, “Performa Naïve Bayes, SVM, dan IndoBERT pada Analisis Sentimen Twitter IndiHome dengan Strategi Penanganan Data Tidak Seimbang,” vol. 814, no. 1, pp. 29–44, 2025, doi: 10.14421/fourier.2025.141.29-44.
- [14] B. Setiawan, “A Review of Sentiment Analysis Applications in Indonesia Between 2023-2024.”
- [15] R. N. R. Suropto, “Teknik pre-processing dan classification dalam data science.,” *Mie.Binus.Ac.Id.*
- [16] M. Nabil *et al.*, “Jurnal Teknologi dan Manajemen Informatika Sentiment Analysis of NTB Syariah Bank Application Services using The Naïve Bayes and Support Vector Machine Methods Article Info ABSTRACT,” vol. 11, no. 2, pp. 177–188, 2025, [Online]. Available: <http://http://jurnal.unmer.ac.id/index.php/jtmi>
- [17] H. Babaei, M. Zamani, and S. Mohammadi, “The impact of data splitting methods on machine learning models: A case study for predicting concrete workability,” *Machine Learning for Computational Science and Engineering*, vol. 1, no. 1, p. 21, 2025, doi: 10.1007/s44379-025-00021-3.
- [18] and J. G. A. Gupta, N. Tatbul, R. Marcus, S. Zhou, I. Lee, “CLASS-WEIGHTED EVALUATION METRICS FOR IMBALANCED DATA CLASSIFICATION,” vol. 55, no. 1, pp. 1–14, 2020.
- [19] V. Chandradev, I. M. A. D. Suarjana, and I. P. A. Bayupati, “Chandradev, Analisis Sentimen Review Hotel menggunakan Metode Deep Learning BERT 107 Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT,” *Jurnal Buana Informatika*, vol. 14, pp. 107–115, 2023.
- [20] Maulidya Prastita Syah, Ajeng Puspa Wardani, Mohammad Idhom, and Trimono, “Perbandingan Representasi Teks Tf-Idf Dan Bert Terhadap Akurasi Cosine Similarity Dalam Penilaian Otomatis Jawaban Berbasis Teks,” *Data Sciences Indonesia (DSI)*, vol. 5, no. 1, pp. 47–59, 2025, doi: 10.47709/dsi.v5i1.6021.
- [21] Y. Liu, J. Yu, Y. Xu, Z. Li, and Q. Zhu, “A Survey on Transformer Context Extension: Approaches and Evaluation,” 2025, [Online]. Available: <https://arxiv.org/abs/2503.13299v1>
- [22] M. Galih Pradana, “Penggunaan Fitur Wordcloud Dan Document Term Matrix Dalam Text Mining,” *Jurnal Ilmiah Infromatika (JIF)*, vol. 08, no. 01, pp. 38–43, 2020.
- [23] R. Merdiansah, S. Siska, and A. Ali Ridha, “Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT,” *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 7, no. 1, pp. 221–228, 2024, doi: 10.55338/jikoms.v7i1.2895.
- [24] G. T. Fadilah, L. Muflikhah, and R. S. Perdana, “Analisis Sentimen Produk Hijab Pada E-Commerce Tokopedia Menggunakan Algoritma Support Vector Machine Dan Indobert Embedding,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 2, pp. 1–9, 2025, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [25] M. Mujahid *et al.*, “Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering,” *J. Big Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-024-00943-4.