

Analisis Sentimen Berbasis Aspek (ABSA) Ulasan Sepatu Ventela di TikTok Shop Menggunakan IndoBERT Fine-Tuned

Aspect-Based Sentiment Analysis (ABSA) of Ventela Shoe Reviews on TikTok Shop Using Fine-Tuned IndoBERT

Fitrawansyah Butas
Program Studi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
luydian17@gmail.com

Amiruddin*
Program Studi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
amier.76@gmail.com

Maryam Hasan
Program Studi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
maryamhasan0709@gmail.com

Rezqiwati Ishak
Program Studi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
rezqi.uig@gmail.com

Rofiq Harun
Program Studi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
rofiqharun18@gmail.com

Andi Kamaruddin
Program Studi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
andi.kamaruddin.ak@gmail.com

Diterima : Juni 2026
Disetujui : Juli 2026
Dipublikasi : Juli 2026

Abstrak—Volume ulasan konsumen yang masif pada platform *social commerce* TikTok Shop menyulitkan merek sepatu lokal seperti Ventela memahami persepsi konsumen secara terstruktur, sementara kajian *Aspect-Based Sentiment Analysis* (ABSA) berbahasa Indonesia pada platform ini masih sangat terbatas. Penelitian ini bertujuan menerapkan IndoBERT *fine-tuned* untuk klasifikasi sentimen berbasis aspek dan mengukur persepsi konsumen terhadap empat aspek produk, yaitu Kenyamanan, Desain, Ketahanan, dan Harga. Melalui pendekatan eksperimen komputasional, 1.000 ulasan dikumpulkan, dianotasi secara otomatis berbasis leksikon dengan penanganan negasi, direstrukturisasi menjadi 706 pasangan ulasan-aspek dan dibagi menggunakan stratified split 80:20, lalu digunakan untuk melatih dan membandingkan tiga model, yakni TF-IDF dengan *Logistic Regression*, TF-IDF dengan Linear SVM, dan IndoBERT *fine-tuned*. Hasil pengujian pada 142 sampel uji menunjukkan IndoBERT *fine-tuned* unggul dengan *Accuracy* 0,8521 dan F1-Macro 0,7813, melampaui kedua *baseline* pada empat dari lima metrik utama. Analisis terhadap 706 pasangan ulasan-aspek mengidentifikasi Desain (75,6% positif) dan Harga (71,8% positif) sebagai kekuatan utama, sedangkan Kenyamanan (32,7% negatif) dan Ketahanan (30,8% negatif) menjadi area perbaikan yang berkaitan dengan penentuan ukuran serta kualitas lem dan jahitan. Penelitian ini berkontribusi memperkaya literatur ABSA Bahasa Indonesia pada domain *social commerce* sekaligus menghasilkan simulator berbasis web siap pakai yang dikembangkan dengan Gradio untuk pemantauan persepsi konsumen secara berkala guna mendukung pengambilan keputusan berbasis data.

Kata Kunci—*Aspect-Based Sentiment Analysis; IndoBERT; Natural Language Processing; klasifikasi teks; social commerce*

Abstract—The massive volume of consumer reviews on the *social commerce* platform TikTok Shop makes it difficult for local shoe brands such as Ventela to understand consumer perception in a structured manner, while Indonesian-language *Aspect-Based*

Sentiment Analysis (ABSA) studies on this platform remain very limited. This study aims to apply *fine-tuned IndoBERT* for aspect-based sentiment classification and to measure consumer perception of four product aspects, namely Comfort, Design, Durability, and Price. Using a computational experiment approach, 1,000 reviews were collected, automatically annotated using a lexicon-based method with negation handling, restructured into 706 review-aspect pairs and divided using an 80:20 stratified split, and used to train and compare three models: TF-IDF with *Logistic Regression*, TF-IDF with Linear SVM, and *fine-tuned IndoBERT*. Testing on 142 test samples shows that *fine-tuned IndoBERT* is superior, achieving an *Accuracy* of 0.8521 and an F1-Macro of 0.7813 and surpassing both baselines on four of five primary metrics. Analysis of 706 review-aspect pairs identifies Design (75.6% positive) and Price (71.8% positive) as the main strengths, while Comfort (32.7% negative) and Durability (30.8% negative) emerge as improvement areas related to sizing and the quality of adhesive and stitching. This study enriches Indonesian ABSA literature in the *social commerce* domain and delivers a ready-to-use web-based simulator built with Gradio to facilitate periodic consumer-perception monitoring for data-driven decision-making processes.

Keywords—*Aspect-Based Sentiment Analysis; IndoBERT; Natural Language Processing; text classification; social commerce*

I. PENDAHULUAN

Perkembangan teknologi digital di Indonesia mengubah paradigma transaksi jual beli dari mode konvensional menuju *social commerce*, yaitu integrasi media sosial dengan perdagangan elektronik. TikTok Shop muncul sebagai platform *social commerce* dominan dengan basis pengguna luas, terutama Generasi Z dan milenial yang memiliki kecenderungan tinggi terhadap *shoppertainment* [1]. Fitur *live streaming* terbukti berpengaruh signifikan terhadap

keputusan pembelian melalui mekanisme *parasocial relationship* dan *utilitarian shopping value* [2], [3]. Pertumbuhan transaksi ini diikuti meningkatnya volume ulasan konsumen yang menjadi bentuk *electronic word-of-mouth* dan mediasi persepsi konsumen dengan minat beli [4], [5]. Namun, volume ulasan yang besar pada setiap halaman produk menyulitkan analisis manual karena satu ulasan dapat memuat beberapa aspek dengan polaritas berbeda, sehingga otomatisasi analisis menjadi kebutuhan mendesak.

Industri sepatu lokal Indonesia tumbuh signifikan seiring *tren ethnocentrism* dan dukungan produk dalam negeri di kalangan Generasi Z [6], [7]. Ventela merupakan salah satu *brand* sepatu lokal yang aktif memasarkan produknya melalui TikTok Shop dengan basis penggemar loyal, sehingga ulasan penggunaannya menjadi sumber data berharga untuk memahami persepsi konsumen terhadap kualitas produk. Pemahaman yang akurat atas persepsi konsumen pada setiap aspek fungsionalitas produk diperlukan untuk mendukung pengambilan keputusan strategis pihak manajemen Ventela.

Keempat aspek tersebut dipilih karena merepresentasikan dimensi kualitas produk alas kaki yang paling sering dievaluasi konsumen pada ulasan daring: Kenyamanan berkaitan dengan pengalaman pemakaian dan kesesuaian ukuran, Desain dengan daya tarik visual dan estetika, Ketahanan dengan kualitas material serta konstruksi seperti jahitan dan perekatan, dan Harga dengan persepsi nilai terhadap kualitas yang diterima. Pemetaan sentimen pada tingkat aspek memberikan informasi yang jauh lebih actionable bagi manajemen dibandingkan sentimen agregat pada tingkat ulasan, karena memungkinkan identifikasi secara spesifik dimensi mana yang menjadi keunggulan dan mana yang perlu diperbaiki.

Sentiment analysis dengan pendekatan klasik berbasis fitur, yang memadukan representasi TF-IDF dan algoritma seperti *Naive Bayes*, *Logistic Regression*, serta *Support Vector Machine* terbukti efektif pada ulasan dengan sentimen tunggal, dengan akurasi pada kisaran 75%-92% [8], [9]. Pendekatan serupa banyak diterapkan pada ulasan produk dan layanan di Indonesia [10], [11]. Keterbatasan utama muncul ketika ulasan memuat beberapa aspek dengan polaritas berbeda dalam satu kalimat, seperti pada ulasan TikTok Shop yang umumnya pendek namun padat informasi. *Aspect-Based Sentiment Analysis* (ABSA) hadir sebagai solusi *fine-grained* yang mampu mengekstrak aspek-aspek spesifik beserta polaritasnya secara terpisah [12]. Tinjauan sistematis menunjukkan domain ulasan produk komersial mendominasi studi ABSA dengan tren model berbasis *transformer* yang meningkat [13], dan penerapannya pada *e-commerce* mengungguli analisis sentimen klasik [14].

Bidirectional Encoder Representations from Transformers (BERT) menjadi tonggak penting model bahasa pra-latih yang menangkap konteks *bidirectional* dan dapat diadaptasi melalui *fine-tuning* [15]. Untuk Bahasa Indonesia, IndoBERT dilatih pada korpus berbahasa Indonesia dan mencapai performa *state-of-the-art* pada beragam task NLP [16], [17]. IndoBERT *fine-tuned* terbukti mengungguli model multilingual pada klasifikasi teks Bahasa Indonesia [18], [19], [20], [21], dan optimasi representasi embedding IndoBERT juga efektif untuk teks akademik pendek berbahasa Indonesia [22]. Meskipun demikian, sebagian

besar penelitian ABSA Indonesia masih berfokus pada *marketplace* tradisional seperti Shopee dan Tokopedia [23], [24], [25], [11], sedangkan kajian khusus pada platform *social commerce* TikTok Shop dengan karakteristik ulasan yang berbeda masih sangat terbatas.

Keunggulan arsitektur *transformer* bersumber pada mekanisme *self-attention* yang menimbang relasi antar seluruh token dalam kalimat secara simultan sehingga menghasilkan *contextual embeddings*, yaitu representasi kata yang maknanya menyesuaikan konteks kalimat. Kemampuan inilah yang secara teoretis membuat model *transformer* seperti IndoBERT lebih andal menangani ulasan pendek multi-aspek, ragam slang, serta struktur *adversatif* berkonjungsi "tapi" yang kerap membalik polaritas, dibandingkan representasi *n-gram* pada pendekatan berbasis fitur yang mengabaikan urutan dan konteks kata.

Secara sistematis, penelitian terdahulu dapat dipetakan sebagai berikut. Studi analisis sentimen pada *marketplace* tradisional, seperti klasifikasi ulasan aplikasi Shopee menggunakan *Support Vector Machine* [11] dan analisis produk lokal dengan *Naive Bayes* serta SVM [10], umumnya masih berhenti pada level sentimen dokumen dengan fitur klasik TF-IDF sehingga belum mampu memisahkan polaritas beberapa aspek dalam satu kalimat. Penelitian yang telah mengadopsi kerangka ABSA, misalnya pada ulasan Shopee dengan SVM [23] maupun dengan kombinasi teknik *embedding* dan *classifier* [24], tetap terbatas pada platform *marketplace* tradisional dan sebagian besar belum memanfaatkan arsitektur *transformer*. Di sisi lain, penerapan IndoBERT *fine-tuned* untuk teks berbahasa Indonesia [18], [19] terbukti unggul, namun masih berfokus pada klasifikasi sentimen dokumen dan belum berbasis aspek. Dengan demikian, celah yang belum terjawab adalah penerapan ABSA berbahasa Indonesia pada platform *social commerce* TikTok Shop yang berkarakteristik ulasan pendek dan informal, sekaligus mengintegrasikannya dengan produk luaran berupa simulator siap pakai; posisi inilah yang menjadi kontribusi penelitian ini.

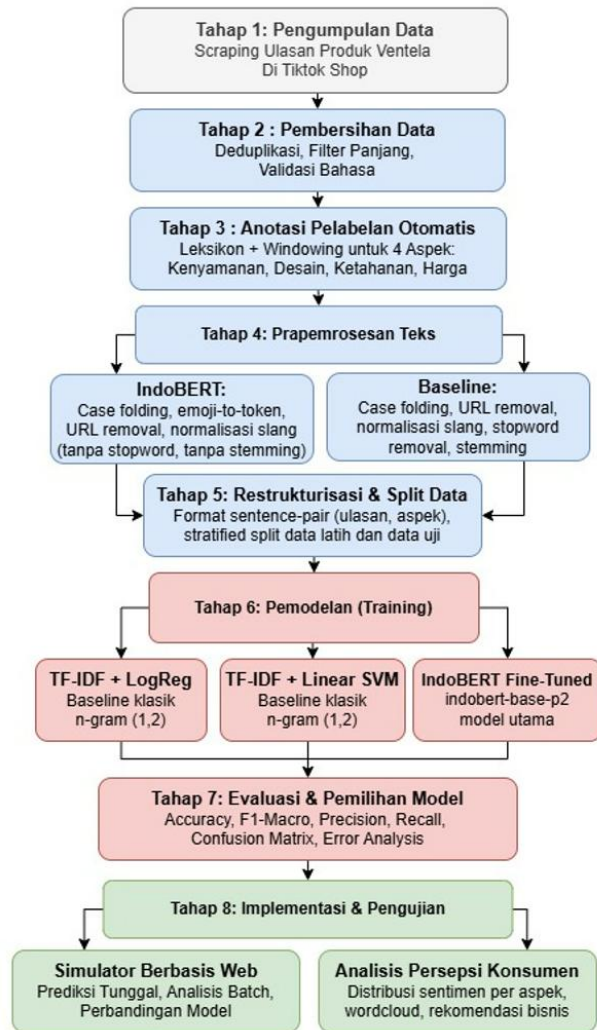
Berdasarkan pemetaan tersebut, terdapat tiga celah penelitian. Pertama, ulasan TikTok Shop cenderung sangat pendek (rata-rata 7,12 kata), banyak *slang*, dan memuat beberapa aspek dalam satu kalimat. Kedua, fokus pada produk sepatu dengan empat aspek (Kenyamanan, Desain, Ketahanan, dan Harga) belum banyak diteliti dalam konteks ABSA Bahasa Indonesia. Ketiga, integrasi pengukuran persepsi konsumen dengan simulator berbasis web yang dapat dipakai langsung oleh pihak Ventela menjadi kontribusi praktis yang relevan. Penelitian ini bertujuan menerapkan IndoBERT *fine-tuned* untuk klasifikasi sentimen berbasis aspek pada ulasan TikTok Shop produk sepatu Ventela serta mengukur persepsi konsumen terhadap keempat aspek tersebut, dengan kontribusi teoretis berupa pengayaan literatur ABSA Bahasa Indonesia dan kontribusi praktis berupa simulator pemantauan persepsi konsumen.

Kontribusi penelitian ini diringkas dalam tiga hal. Pertama, secara metodologis, penelitian menerapkan dan membandingkan IndoBERT *fine-tuned* dengan dua baseline klasik (TF-IDF dengan *Logistic Regression* dan *Linear SVM*) untuk ABSA berformat *sentence-pair* pada ulasan berbahasa Indonesia yang pendek dan informal. Kedua, secara empiris, penelitian memetakan persepsi konsumen terhadap empat aspek fungsional sepatu yang jarang dikaji pada *social*

commerce berbahasa Indonesia. Ketiga, secara praktis, penelitian menghasilkan simulator berbasis web siap pakai untuk memantau persepsi konsumen secara berkelanjutan dan berbasis data.

II. METODE

Penelitian ini menggunakan pendekatan eksperimen komputasional yang terdiri atas delapan tahap berurutan, mulai dari pengumpulan data hingga implementasi dan pengujian simulator. Tahapan penelitian secara keseluruhan diuraikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

A. Pengumpulan Data

Data dikumpulkan dari halaman produk Ventela di TikTok Shop melalui pendekatan semi-otomatis pada periode Juli 2022 sampai Desember 2025. Total 1.000 ulasan berhasil dihimpun dengan tujuh atribut, yaitu nomor, id ulasan, tanggal, teks ulasan, produk, sentimen, dan aspek. Ringkasan transformasi data dari tahap pengumpulan hingga siap uji disajikan pada Tabel 1.

TABEL 1. REKAPITULASI TAHAPAN PENGOLAHAN DATA

Tahap	Operasi	Jumlah Data
Pengumpulan	Scraping ulasan	1.000 ulasan
Pembersihan	Deduplikasi, filter	874 ulasan
Anotasi	Auto-labeling per aspek	874 x 4 aspek

Tahap	Operasi	Jumlah Data
Restrukturisasi	Format <i>sentence-pair</i>	706 pasangan
Split	Stratified 80:20	564 + 142

B. Pembersihan dan Anotasi Data

Empat operasi pembersihan dilakukan berurutan, yaitu deduplikasi teks, filter ulasan kurang dari tiga kata, ekstraksi varian produk, dan pengecekan bahasa, sehingga diperoleh 874 ulasan bersih. Anotasi label awal dilakukan dengan pelabelan otomatis berbasis leksikon dan jendela kata (*windowing*) untuk empat aspek. Untuk setiap kata kunci aspek, diambil jendela tiga kata ke kiri dan ke kanan, lalu skor sentimen dihitung dengan memperhatikan negasi. Tanda total skor menentukan label, yaitu Positif jika skor lebih besar dari nol, Negatif jika kurang dari nol, dan Netral jika sama dengan nol, sedangkan aspek yang tidak muncul diberi label Tidak Disebut.

Leksikon sentimen disusun dari kata-kata opini berbahasa Indonesia yang lazim pada ulasan produk dan diperkaya dengan istilah domain sepatu, sedangkan penanganan negasi diterapkan untuk membalik polaritas pada rentang jendela. Pendekatan pelabelan otomatis ini dipilih karena efisien untuk memberi label awal pada ribuan pasangan ulasan-aspek tanpa biaya anotasi manual berskala besar. Perlu dicatat bahwa label yang dihasilkan bersifat silver standard sehingga masih berpotensi mengandung bias aturan leksikon; oleh karena itu verifikasi anotasi oleh manusia sebagai gold standard beserta pengukuran tingkat kesepakatan antar-anotator (*inter-annotator agreement*) diposisikan sebagai agenda penyempurnaan pada penelitian lanjutan.

C. Prapemrosesan Teks

Prapemrosesan dilakukan berbeda untuk model *baseline* dan IndoBERT sesuai karakteristik masing-masing, sebagaimana disajikan pada Tabel 2. Jalur *baseline* menerapkan operasi yang lebih agresif (*stopword removal* dan *stemming* Sastrawi), sedangkan jalur IndoBERT mempertahankan struktur kalimat asli dan memetakan emoji menjadi token sentimen agar konteks bagi model *transformer* tetap utuh.

TABEL 2. PERBANDINGAN OPERASI PRAPEMROSESAN PER MODEL

Operasi	Baseline	IndoBERT
Case folding	Ya	Ya
Penghapusan URL	Ya	Ya
Penanganan emoji	Hapus	Token
Karakter non-alfabet	Hapus	Hapus
Normalisasi kata tidak baku	Ya	Ya
Stopword removal	Ya	Tidak
Stemming Sastrawi	Ya	Tidak

D. Restrukturisasi dan Pembagian Data

Data diubah dari format wide menjadi format *sentence-pair*, yaitu satu baris untuk setiap pasangan ulasan-aspek. Pasangan berlabel Tidak Disebut tidak diikutkan ke dalam pelatihan, sehingga diperoleh 706 pasangan berlabel dari 874 ulasan. Data dibagi dengan rasio 80:20 menggunakan *stratified split* berdasarkan kombinasi aspek dan label,

menghasilkan 564 pasangan data latih dan 142 pasangan data uji.

E. Pemodelan

Dua model *baseline* dibangun menggunakan vektorisasi TF-IDF dengan n-gram (1,2) berdimensi 1.027 fitur, yaitu TF-IDF dengan *Logistic Regression* dan TF-IDF dengan Linear SVM. Bobot fitur dihitung menggunakan *Term Frequency-Inverse Document Frequency*, dengan komponen IDF mengikuti Persamaan (1).

$$IDF(t) = \ln\left(\frac{1 + N}{1 + df(t)}\right) + 1 \quad (1)$$

dengan N adalah jumlah total dokumen dan $df(t)$ adalah jumlah dokumen yang memuat term t . Persamaan ini mengikuti implementasi default *TfidfVectorizer* pada pustaka *scikit-learn* (`smooth_idf=True`), dengan vektor TF-IDF akhir dinormalisasi L2. Model utama adalah IndoBERT-base-p2 yang di-fine-tune untuk klasifikasi *sentence-pair* dengan tiga kelas keluaran, menggunakan *learning rate* $2e-5$, *batch size* 16, panjang token maksimum 128, *weight decay* 0,01, *warmup ratio* 0,1, lima *epoch*, dan *early stopping* berdasarkan F1-Macro. Probabilitas tiap kelas diperoleh melalui fungsi *softmax* pada Persamaan (2), dengan nilai tertinggi digunakan sebagai label prediksi sekaligus skor *confidence*.

$$P(y_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \quad (2)$$

dengan z_i adalah logit (skor mentah) untuk kelas ke- i , K adalah jumlah kelas (tiga, yaitu Positif, Netral, dan Negatif) sehingga penjumlahan pada penyebut dilakukan atas seluruh K kelas, dan $P(y_i)$ menyatakan probabilitas prediksi kelas ke- i .

Di atas keluaran encoder IndoBERT ditambahkan lapisan klasifikasi (*classification head*) standar untuk klasifikasi sekuens. Representasi token [CLS] terlebih dahulu melewati lapisan *pooler* berupa transformasi linear yang diikuti fungsi aktivasi *tanh*, kemudian sebuah lapisan *dropout* dengan probabilitas 0,1 untuk mengurangi *overfitting*, dan akhirnya sebuah lapisan linear penuh berukuran 768×3 yang memetakan representasi tersembunyi ke tiga kelas keluaran.

Seluruh eksperimen dijalankan pada lingkungan Google Colaboratory dengan akselerator GPU NVIDIA Tesla T4 (16 GB). Model IndoBERT-base-p2 memiliki sekitar 124 juta parameter (berukuran lebih kurang 475 MB dalam presisi 32-bit); proses fine-tuning selama lima *epoch* berlangsung sekitar 8-15 menit, sedangkan pelatihan kedua *baseline* TF-IDF hanya membutuhkan hitungan detik. Pada tahap inferensi, ketiga model mampu memproses ulasan dengan latensi rendah mendekati real-time sehingga secara komputasi layak diintegrasikan ke dalam simulator interaktif.

F. Evaluasi dan Implementasi Simulator

Evaluasi menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan F1-Score per kelas, serta F1-Macro sebagai metrik utama yang tidak sensitif terhadap ketidakseimbangan kelas. Nilai F1-Score dihitung sebagai *harmonic mean* dari *Precision* dan *Recall* sebagaimana Persamaan (3), dilengkapi *confusion matrix* dan analisis kesalahan kualitatif. Model terbaik kemudian diimplementasikan ke dalam simulator berbasis Gradio dengan tiga modul, yaitu Prediksi Tunggal, Analisis Batch, dan Perbandingan Model.

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

III. HASIL DAN PEMBAHASAN

A. Karakteristik dan Distribusi Data

Ulasan TikTok Shop tergolong sangat pendek dengan rata-rata 7,12 kata dan median 7 kata per ulasan. Setelah restrukturisasi, distribusi label final pada 706 pasangan menunjukkan dominasi kelas Positif (61,5%) dan keterbatasan kelas Negatif, terutama pada aspek Desain dan Harga. Distribusi label per aspek beserta rasio Positif terhadap Negatif disajikan pada Tabel 3. Ketidakseimbangan yang tinggi pada Desain (9,6:1) dan Harga (7,6:1) berimplikasi pada performa model untuk kelas Negatif kedua aspek tersebut.

TABEL 3. DISTRIBUSI LABEL FINAL PER ASPEK

Aspek	Positif	Netral	Negatif	Total	Pos:Neg
Kenyamanan	155	24	78	257	2,0 : 1
Desain	96	25	10	131	9,6 : 1
Ketahanan	130	38	72	240	1,8 : 1
Harga	53	18	7	78	7,6 : 1
Total	434	105	167	706	2,6 : 1

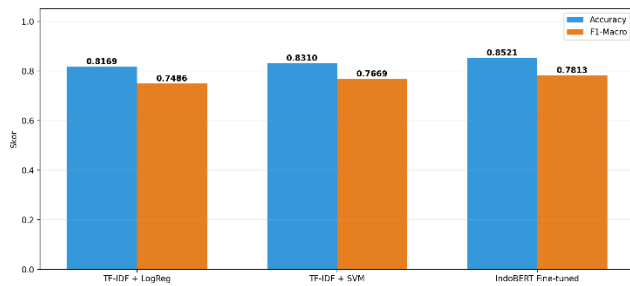
B. Perbandingan Performa Model

Perbandingan langsung ketiga model pada 142 sampel uji disajikan pada Tabel 4. IndoBERT *fine-tuned* mencapai *Accuracy* 0,8521 dan F1-Macro 0,7813, mengungguli kedua *baseline* pada empat dari lima metrik, yaitu *Accuracy*, F1-Macro, F1-Negatif, dan F1-Positif. Pada F1-Netral, Linear SVM justru unggul tipis (0,6316 berbanding 0,5714), mengindikasikan keterbatasan IndoBERT dalam mendeteksi sentimen Netral yang ambigu. Perbandingan visual *Accuracy* dan F1-Macro ketiga model ditampilkan pada Gambar 2.

Selisih *Accuracy* IndoBERT terhadap Linear SVM sebesar 2,11 poin persentase dan terhadap Logistic Regression sebesar 3,52 poin persentase bersifat konsisten pada seluruh aspek maupun pada empat dari lima metrik utama, sehingga keunggulan tersebut tidak bergantung pada satu metrik semata. Pengujian signifikansi statistik secara formal (misalnya uji McNemar atau paired t-test) serta validasi silang k-fold untuk mengukur stabilitas performa diposisikan sebagai langkah penguatan pada penelitian lanjutan.

TABEL 4. PERBANDINGAN PERFORMA TIGA MODEL PADA DATA UJI

Model	Accuracy	F1-Macro	F1-Neg	F1-Net	F1-Pos
TF-IDF + LogReg	0,8169	0,7486	0,8116	0,5455	0,8889
TF-IDF + SVM	0,8310	0,7669	0,7714	0,6316	0,8977
IndoBERT Fine-tuned	0,8521	0,7813	0,8529	0,5714	0,9195

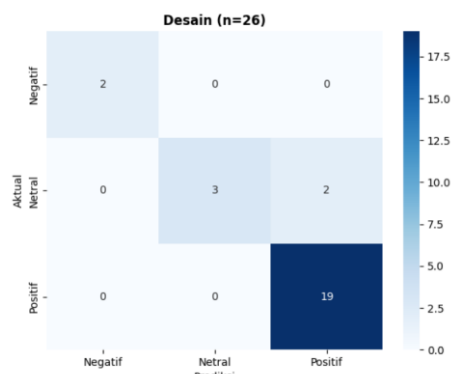


Gambar 2. Perbandingan Performa *Accuracy* dan F1-Macro Tiga Model

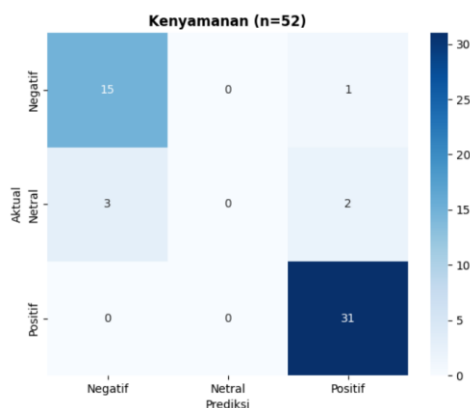
Performa IndoBERT juga bervariasi antar aspek sebagaimana disajikan pada Tabel 5. Aspek Desain memperoleh F1-Macro tertinggi (0,954) karena dominasi data positif yang konsisten, sedangkan aspek Harga memperoleh F1-Macro terendah (0,441) akibat jumlah sampel yang sangat sedikit. Pola kesalahan model secara keseluruhan ditampilkan melalui *confusion matrix* pada Gambar 3.

TABEL 5. PERFORMA INDOBERT PER ASPEK PADA DATA UJI

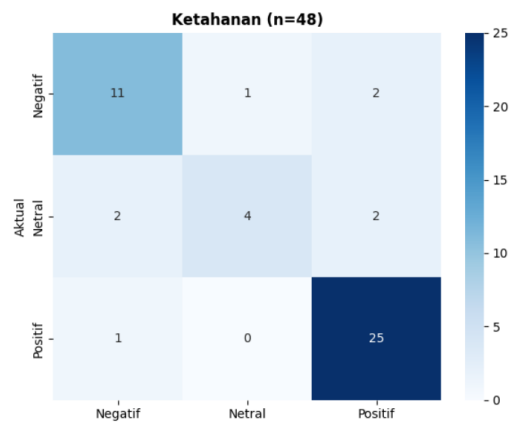
Aspek	n_uji	Accuracy	F1-Macro	Catatan
Desain	26	0,962	0,954	Tertinggi, data positif dominan
Kenyamanan	52	0,904	0,792	Tinggi, sampel cukup besar
Ketahanan	48	0,812	0,721	Sedang, semantik kompleks
Harga	16	0,625	0,441	Terendah, sampel sangat sedikit



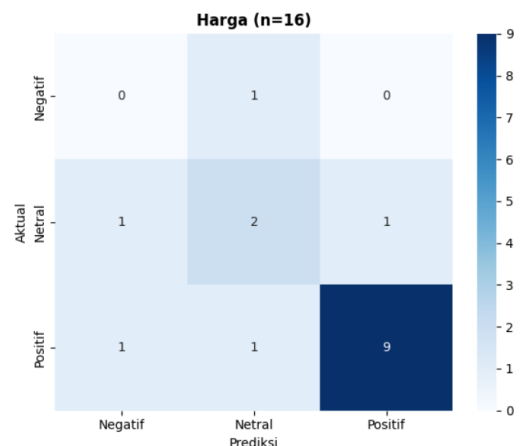
(a). desain



(b). kenyamanan



(c). ketahanan



(d). harga

Gambar 3. *Confusion Matrix* IndoBERT *Fine-Tuned*

C. Analisis Kesalahan (Error Analysis)

Dari 142 sampel uji, terdapat 21 prediksi salah (14,8%). Sebagian contoh kesalahan disajikan pada Tabel 6. Pola kesalahan dominan adalah kesulitan model mendeteksi kelas Netral yang kerap bergeser menjadi Positif atau Negatif. Pola kedua adalah belum konsistennya model menangani struktur adversatif dengan konjungsi tapi yang membalik polaritas, sedangkan pola ketiga adalah kesulitan pada sarkasme dan ironi tersirat. Temuan ini sejalan dengan tantangan ABSA yang sensitif terhadap konteks linguistik kompleks [12], [13].

TABEL 6. CONTOH PREDIKSI SALAH UNTUK ANALISIS KESALAHAN

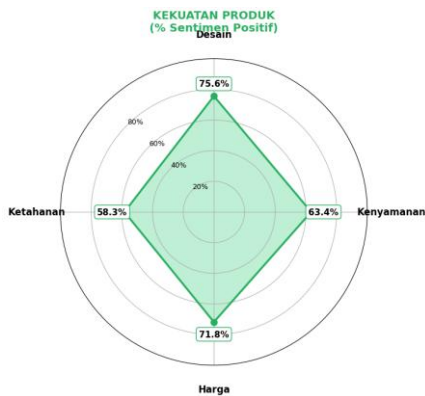
Teks Ulasan	Aspek	Aktual	Prediksi
Model sesuai pesanan hanya ukuran agak longgar	Kenyamanan	Netral	Positif
untuk harga segitu kukira empuk ternyata keras	Harga	Positif	Netral
baru 1 x pakai lem langsung lepas-lepas	Ketahanan	Netral	Negatif
Jahitan tidak rapi.. Tapi ya harga segini mau gmna lagi	Harga	Positif	Negatif
Sepatu biasa saja tapi mahal dengan kualitas biasa	Harga	Negatif	Netral
worth it banget dah beli produk di sini	Harga	Positif	Netral
bakal ke pake lama si ini, nyaman dan pas	Ketahanan	Netral	Positif

D. Pengukuran Persepsi Konsumen

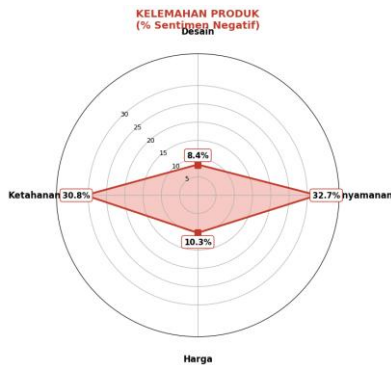
Prediksi IndoBERT pada keseluruhan 706 pasangan menjadi dasar pengukuran persepsi konsumen terhadap empat aspek fungsionalitas, sebagaimana dirangkum pada Tabel 7 dan divisualisasikan pada Gambar 4. Aspek Desain merupakan kekuatan utama produk Ventela dengan 75,6% ulasan positif, diikuti aspek Harga dengan 71,8% positif yang menunjukkan persepsi nilai harga yang baik, sehingga keduanya layak dipertahankan sebagai nilai jual utama. Sebaliknya, Kenyamanan (32,7% negatif) dan Ketahanan (30,8% negatif) menjadi area perbaikan prioritas. Analisis *wordcloud* kata kunci per aspek dan sentimen pada Gambar 5 memperjelas akar keluhan tersebut: panel Kenyamanan negatif didominasi istilah kegedean, kekecilan, kebesaran, dan ukuran yang mengindikasikan masalah penentuan ukuran, sedangkan panel Ketahanan negatif didominasi istilah lem, jebol, kurang, dan lama yang mengindikasikan masalah kontrol kualitas pada perekatan dan jahitan. Temuan ini menghasilkan rekomendasi praktis berupa revisi panduan ukuran serta peningkatan kontrol kualitas lem dan jahitan.

TABEL 7. REKAPITULASI PERSEPSI KONSUMEN PER ASPEK

Aspek	Pos (%)	Net (%)	Neg (%)	Kategori	Implikasi Bisnis
Desain	75,6	16,0	8,4	Kekuatan Utama	Pertahankan sebagai nilai jual utama
Harga	71,8	17,9	10,3	Kekuatan	Sesuai nilai harga, dipertahankan
Kenyamanan	63,4	3,9	32,7	Area Perbaikan	Revisi panduan ukuran
Ketahanan	58,3	10,9	30,8	Area Perbaikan	Fokus QC lem dan jahitan



(a). kekuatan produk



(b). kelemahan produk

Gambar 4. Radar Kekuatan dan Kelemahan Produk Ventela per Aspek



(a). wordcloud aspek kenyamanan



(b). wordcloud aspek desain



(c). wordcloud aspek ketahanan



(d). wordcloud aspek harga

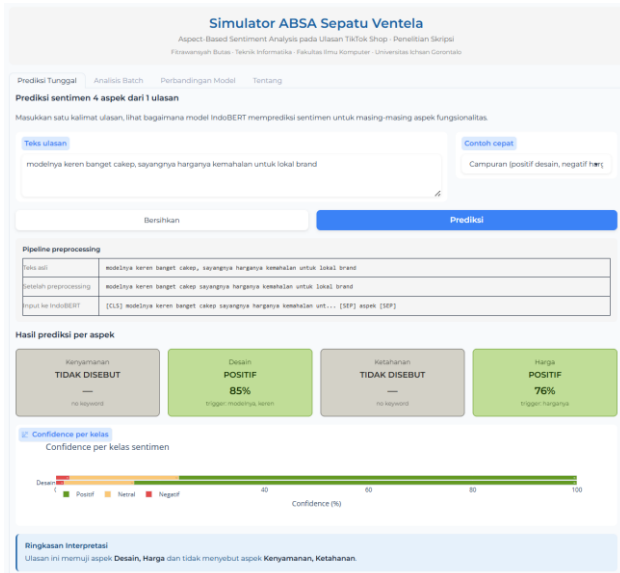
Gambar 5. Wordcloud Kata Kunci per Aspek dan Sentimen

E. Pengujian Simulator

Simulator ABSA Ventela diuji pada tiga modul menggunakan ulasan dari *dataset* asli, dengan rekapitulasi pada Tabel 8. Modul Prediksi Tunggal akurat pada ulasan pujian maupun kritik dengan satu sentimen dominan, namun hanya sebagian tepat pada ulasan campuran yang mengandung kata penghalus (*mitigator*) seperti sayangnya dan agak, konsisten dengan temuan analisis kesalahan. Modul Analisis Batch berhasil memproses 40 ulasan sekaligus dan menghasilkan dasbor otomatis beserta rekomendasi, sedangkan Modul Perbandingan Model mendemonstrasikan keunggulan IndoBERT atas *baseline* pada teks adversatif. Antarmuka simulator ditampilkan pada Gambar 6.

TABEL 8. REKAPITULASI HASIL PENGUJIAN MODUL SIMULATOR

Modul	Skenario	Hasil Utama	Status
Prediksi Tunggal	Pujian	Dua aspek positif terdeteksi tepat	Berhasil
Prediksi Tunggal	Kritik	Dua aspek negatif terdeteksi tepat	Berhasil
Prediksi Tunggal	Campuran	Desain tepat, Harga salah (<i>mitigator</i>)	Sebagian
Analisis Batch	40 ulasan	Skor persepsi 4,6/10, Ketahanan dominan	Berhasil
Perbandingan Model	Teks kompleks	IndoBERT unggul 2 aspek atas <i>baseline</i>	Berhasil



Gambar 6. Antarmuka Simulator ABSA Ventela

F. Pembahasan dan Implikasi

Keunggulan IndoBERT atas *baseline* TF-IDF menegaskan bahwa pemahaman konteks dua arah (*bidirectional*) mampu menangkap kata pemicu kontekstual yang tidak tertangkap oleh representasi *n*-gram, terutama pada ulasan multi-aspek yang pendek dan informal. Temuan ini konsisten dengan penelitian terdahulu yang melaporkan keunggulan model berbasis *transformer* pada ABSA *e-commerce* [14], [23], [24] serta efektivitas optimasi representasi embedding IndoBERT pada teks berbahasa Indonesia [22]. Dibandingkan studi analisis sentimen pada *marketplace* tradisional [10], [11], penelitian ini memperluas kajian ke platform *social commerce* TikTok Shop dengan karakteristik ulasan yang jauh lebih ringkas. Perkembangan model bahasa mutakhir seperti XLM-RoBERTa [26] dan ModernBERT [27] membuka peluang perbandingan lanjutan untuk menakar sejauh mana arsitektur yang lebih baru dapat meningkatkan performa ABSA Bahasa Indonesia. Secara praktis, integrasi model dengan simulator memungkinkan pihak Ventela melakukan pemantauan persepsi konsumen secara berkala dan berbasis data, sehingga keputusan perbaikan produk, penyesuaian strategi pemasaran, dan peningkatan kepuasan konsumen dapat dilakukan lebih terukur.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa IndoBERT *fine-tuned* efektif diterapkan untuk klasifikasi sentimen berbasis aspek pada ulasan TikTok Shop produk sepatu Ventela, dengan *Accuracy* 0,8521 dan *F1-Macro* 0,7813 yang mengungguli *baseline* TF-IDF dengan *Logistic Regression* dan Linear SVM pada empat dari lima metrik. Pengukuran persepsi mengidentifikasi Desain sebagai kekuatan utama (75,6% positif), sedangkan Kenyamanan (32,7% negatif) dan Ketahanan (30,8% negatif) menjadi area perbaikan prioritas yang terkait masalah penentuan ukuran serta kontrol kualitas lem dan jahitan. Simulator berbasis web yang dihasilkan siap dimanfaatkan untuk pemantauan persepsi konsumen secara berkala. Penelitian ini memiliki keterbatasan berupa evaluasi pada satu *brand*, ketidakseimbangan data pada aspek Harga, serta kesulitan model menangani struktur *adversatif* dan sarkasme. Berdasarkan keterbatasan tersebut, penelitian

lanjutan disarankan untuk menggunakan anotasi terverifikasi manusia sebagai *gold standard*, memperbesar dan menyeimbangkan *dataset* melalui augmentasi data kelas minoritas, menguji model bahasa yang lebih mutakhir seperti IndoBERT-large, E5, atau BGE-M3, memperluas pengujian secara lintas merek dan lintas *platform* seperti Shopee dan Tokopedia, serta mengintegrasikan model ke dalam sistem pemantauan persepsi konsumen institusional sebagai bagian dari transformasi digital di lingkungan bisnis. Pengujian fungsional pada tiga modul simulator (Tabel 8) menunjukkan seluruh modul berjalan sesuai rancangan secara *real-time*, sehingga kontribusi praktis penelitian ini terbukti tervalidasi secara fungsional. Secara ilmiah, penelitian ini memperkaya pengembangan metode *Aspect-Based Sentiment Analysis* untuk Bahasa Indonesia, khususnya pada domain *social commerce* dengan karakteristik ulasan pendek dan informal yang selama ini belum banyak dieksplorasi. Selain arah lanjutan yang telah disebutkan, penyempurnaan ke depan mencakup verifikasi anotasi oleh manusia sebagai *gold standard* beserta pengukuran *inter-annotator agreement*, pelaksanaan uji signifikansi statistik dan *k-fold cross-validation*, serta perbandingan dengan model *transformer* mutakhir seperti XLM-RoBERTa dan ModernBERT.

REFERENSI

- [1] H. Fitriani, A. N. Hidayanto, and Q. Munajat, "TikTok Shop: How trust and privacy influence generation Z's purchasing behaviors," *Cogent Bus. & Manag.*, vol. 10, no. 3, p. 2292759, 2023, doi: 10.1080/23311886.2023.2292759.
- [2] D. P. Pradnyamitha and A. F. Maradona, "Actual Purchase on TikTok Live Streaming Commerce: An Analysis of Utilitarian Shopping Value and Attitude Toward Electronic Live Streaming," *Int. J. Sci. Soc.*, vol. 6, no. 4, pp. 268–284, 2024, doi: 10.54783/ijssoc.v6i4.1329.
- [3] M. Chandra, D. W. Sukmaningsih, and E. Sriwardiningsih, "The Impact of Live Streaming On Purchase Intention In Social Commerce In Indonesia," *Procedia Comput. Sci.*, vol. 234, pp. 987–995, 2024, doi: 10.1016/j.procs.2024.03.088.
- [4] Y. Umarni, M. Abdul Aziz, and Alhidayatullah, "Analysis of The Influence of Social Commerce TikTok Shop on Purchase Intention," *Int. J. Soc. Sci. Educ. Commun. Econ.*, vol. 3, no. 5, pp. 1461–1472, 2024, doi: 10.54443/sj.v3i5.435.
- [5] A. D. K. Silalahi, F. M. E. Kurniasari, B. Ndruru, and I. F. Tjeng, "Socio-technical systems and trust transfer in live streaming e-commerce: analyzing stickiness and purchase intentions with SEM-fsQCA," *Front. Commun.*, vol. 9, p. 1305409, 2024, doi: 10.3389/fcomm.2024.1305409.
- [6] P. Heriyati, A. Bismo, and M. Erwinta, "Jakarta's Generation Z and Local Fashion Industry: Unveiling the Impact of Brand Image, Perceived Quality, and Country of Origin," *Binus Bus. Rev.*, vol. 15, no. 1, pp. 69–77, 2024, doi: 10.21512/bbr.v15i1.10162.
- [7] T. S. Dhewi and R. Oktaviani, "Does perceived quality mediate the effect of generation Z's consumer ethnocentrism on local sneakers purchase intention?," *BISMA (Bisnis dan Manajemen)*, vol. 15, no. 2, pp. 139–157, 2023, doi: 10.26740/bisma.v15n2.p139-157.
- [8] R. K. Das, M. Islam, M. M. Hasan, S. Razia, M. Hassan, and S. A. Khushbu, "Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models," *Heliyon*, vol. 9, no. 9, p. e20281, 2023, doi: 10.1016/j.heliyon.2023.e20281.
- [9] C. Y. Chen and J. Y. Lin, "A Comparative Study of Sentiment Analysis on Customer Reviews Using Machine Learning and Deep Learning," *Computers*, vol. 13, no. 12, p. 340, 2024, doi: 10.3390/computers13120340.
- [10] W. Clarisha, A. A. M. Fani, D. F. Surianto, and N. Fadilah, "Sentiment Analysis of Local Sunscreen Skintific, Somethinc, and Avoskin with Naive Bayes and SVM," *Jambura J. Electr. Electron. Eng.*, vol. 7, no. 2, 2025, doi: 10.37905/jjee.v7i2.30257.
- [11] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.

- [12] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 11, pp. 11019–11038, 2023, doi: 10.1109/TKDE.2022.3230975.
- [13] Y. C. Hua, P. Denny, K. Taskova, and J. Wicker, "A systematic review of aspect-based sentiment analysis: domains, methods, and trends," *Artif. Intell. Rev.*, vol. 57, no. 11, p. 296, 2024, doi: 10.1007/s10462-024-10906-z.
- [14] M. M. Yenikar, T. Babu, B. Karunamoorthy, and K. T. Vijayaraghavan, "Aspect-based sentiment classification of user reviews to understand customer satisfaction of e-commerce platforms," *Electron. Commer. Res.*, 2025, doi: 10.1007/s10660-025-09948-4.
- [15] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.
- [16] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP," in *Proceedings of the 28th International Conference on Computational Linguistics*, 2020, pp. 757–770, doi: 10.18653/v1/2020.coling-main.66.
- [17] B. Willie *et al.*, "IndoNLU: benchmark and resources for evaluating Indonesian natural language understanding," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 2020, pp. 843–857, doi: 10.18653/v1/2020.acl-main.85.
- [18] L. Geni, E. Yulianti, and D. I. Sensuse, "Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBERT Language Models," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 3, pp. 746–757, 2023, doi: 10.26555/jiteki.v9i3.26490.
- [19] 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," *Ilk. J. Ilm.*, vol. 14, no. 3, pp. 348–354, 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.
- [20] C. H. Lin and U. Nuha, "Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy," *J. Big Data*, vol. 10, no. 1, p. 88, 2023, doi: 10.1186/s40537-023-00782-9.
- [21] M. F. Kono, I. N. Fajri, and Y. Pristyanto, "Public Sentiment Analysis on Corruption Issues in Indonesia Using IndoBERT Fine-Tuning, Logistic Regression, and Linear SVM," *J. Appl. Informatics Comput.*, vol. 9, no. 5, pp. 2616–2628, 2025, doi: 10.30871/jaic.v9i5.10537.
- [22] R. Ishak and Amiruddin, "Optimizing of IndoBERT Embedding with Ditto Whitening for Measuring Research Title Similarity," *Jambura J. Electr. Electron. Eng.*, vol. 8, no. 1, pp. 46–54, 2026, doi: 10.37905/jjee.v8i1.35554.
- [23] N. P. C. M. D. Permatasari, M. A. Aulia, and I. K. G. D. Putra, "Aspect Based Sentiment Analysis on Shopee Application Reviews Using Support Vector Machine," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 15, no. 2, pp. 100–110, 2024, doi: 10.24843/LKJITI.2024.v15.i02.p03.
- [24] J. Gondohanindijo, "A Comparative Study of Embedding Techniques and Classifiers for Aspect-Based Sentiment Analysis of Shopee Reviews," *Techno.COM*, vol. 24, no. 4, 2025, doi: 10.62411/tc.v24i4.14976.
- [25] M. R. F. Aristya and B. Pamungkas, "Sentiment Analysis of Marketplace Review with Islamic Perspective using Fine-Tuning DistilBERT," *Khazanah J. Relig. Technol.*, vol. 1, no. 2, 2024, doi: 10.15575/kjrt.v1i2.11118.
- [26] A. Conneau *et al.*, "Unsupervised cross-lingual representation learning at scale," in *Proceedings of the 58th annual meeting of the association for computational linguistics*, 2020, pp. 8440–8451, doi: 10.18653/v1/2020.acl-main.747.
- [27] B. Warner *et al.*, "Smarter, better, faster, longer: A modern bidirectional encoder for fast, memory efficient, and long context finetuning and inference," in *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2025, pp. 2526–2547, doi: 10.18653/v1/2025.acl-long.127.