



Classification of Public Opinion on Affiliate Marketing on TikTok Based on Specific Aspects Using Fine-Tuned IndoBERT

Klasifikasi Opini Publik Berbasis Aspek terhadap Affiliate Marketing di TikTok Menggunakan *Fine-Tuned* IndoBERT

Vina Oktafiyana¹, Aditia Yudhistira^{2*}

^{1,2}Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer,
Universitas Teknokrat Indonesia, Indonesia

E-Mail: ¹vina_oktafiyana@teknokrat.ac.id, ²aditiayudhistira@teknokrat.ac.id

Received May 08th 2026; Revised Jun 15th 2026; Accepted Jul 08th 2026; Available Online Jul 19th 2026

Corresponding Author: Aditia Yudhistira

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

Affiliate marketing on TikTok has rapidly grown into a commission-based digital marketing strategy that involves content creators promoting products to users. This activity generates various public opinions on product quality, creator credibility, pricing and promotions, shopping experiences, and the quality of content shared on social media. This study aims to analyze user sentiment toward affiliate marketing practices using an aspect-based sentiment classification (ABSC) approach based on pretrained Indonesian Bidirectional Encoder Representations from Transformers (IndoBERT). A total of 10,000 Indonesian-language TikTok comments were collected and processed through several stages, including preprocessing, deduplication, aspect identification using keyword matching and cosine similarity based on Term Frequency Inverse Document Frequency (TF-IDF), and sentiment labeling using IndoBERT. The novelty of this study lies in the integration of an aspect identification approach combining keyword matching and TF-IDF cosine similarity with a fine-tuned IndoBERT model for aspect-based sentiment classification of Indonesian-language TikTok comments on affiliate marketing. This approach provides a more contextual opinion analysis in the affiliate marketing domain, which has received limited attention in previous studies. The model was then fine-tuned using weighted cross-entropy loss to address class imbalance in the dataset. The experimental results achieved an accuracy of 89.79%, a weighted F1-score of 89.74%, and a macro F1-score of 82.29%. Furthermore, evaluation using stratified 5-fold cross-validation yielded an average accuracy of 91.03%, indicating that the proposed model demonstrates stable and effective performance in aspect-based sentiment classification of Indonesian-language comments.

Keywords: *Affiliate Marketing, Aspect-Based Sentiment Classification, IndoBERT, Sentiment Analysis, TikTok*

Abstrak

*Affiliate marketing pada TikTok berkembang pesat sebagai strategi pemasaran digital berbasis komisi yang melibatkan kreator konten untuk mempromosikan produk kepada pengguna. Aktivitas ini memunculkan berbagai opini publik terkait kualitas produk, kredibilitas kreator, harga dan promosi, pengalaman belanja, serta kualitas konten yang dibagikan melalui media sosial. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap praktik *affiliate marketing* menggunakan pendekatan *aspect-based sentiment classification* (ABSC) berbasis *Indonesian Bidirectional Encoder Representations from Transformers* (IndoBERT). Sebanyak 10.000 komentar TikTok berbahasa Indonesia dikumpulkan dan diproses melalui tahapan *preprocessing*, deduplikasi, identifikasi aspek menggunakan *keyword matching* dan *Term Frequency — Inverse Document Frequency* (TF-IDF) *cosine similarity*, serta pelabelan sentimen menggunakan *pretrained* IndoBERT. Kebaruan penelitian ini terletak pada penerapan pendekatan terintegrasi antara identifikasi aspek berbasis *keyword matching* dan TF-IDF *cosine similarity* dengan model IndoBERT yang difine-tune untuk klasifikasi sentimen berbasis aspek pada komentar TikTok berbahasa Indonesia terkait *affiliate marketing*. Pendekatan ini mampu memberikan analisis opini yang lebih kontekstual pada domain *affiliate marketing* yang masih jarang dibahas dalam penelitian sebelumnya. Model kemudian di-fine-tune menggunakan *weighted cross-entropy loss* untuk menangani ketidakseimbangan kelas pada data. Hasil pengujian menunjukkan akurasi sebesar 89,79%, *weighted F1-score* sebesar 89,74%, dan *macro F1-score* sebesar 82,29%. Evaluasi menggunakan *stratified 5-fold cross-validation* menghasilkan rata-rata akurasi sebesar 91,03%, sehingga menunjukkan bahwa model memiliki performa yang stabil dan efektif dalam klasifikasi sentimen berbasis aspek pada komentar berbahasa Indonesia.*

Kata Kunci: *Affiliate Marketing, Analisis Sentimen, Aspect-Based Sentiment Classification, IndoBERT, TikTok*



1. PENDAHULUAN

Perkembangan teknologi digital telah mengubah perilaku konsumen dalam mencari, memperoleh, dan mengevaluasi informasi produk secara signifikan. Media sosial tidak lagi hanya digunakan sebagai sarana komunikasi dan hiburan, tetapi juga berkembang menjadi platform pemasaran digital dan transaksi elektronik. Salah satu platform yang mengalami perkembangan pesat adalah TikTok. Platform ini berhasil mengintegrasikan konten hiburan dengan aktivitas perdagangan elektronik melalui fitur *TikTok Shop*, keranjang kuning, dan *live streaming* [1]. Integrasi tersebut mendorong munculnya model pemasaran berbasis *affiliate marketing*, yaitu strategi pemasaran digital yang melibatkan kreator konten dalam mempromosikan produk kepada pengikutnya dengan imbalan komisi berdasarkan performa penjualan.

Model *affiliate marketing* dinilai lebih efektif dibandingkan iklan konvensional karena mampu membangun komunikasi yang lebih personal dan interaktif dengan konsumen. Pertumbuhan *affiliate marketing* di TikTok Indonesia juga diiringi oleh meningkatnya partisipasi pengguna dalam bentuk komentar, ulasan, dan interaksi digital lainnya [2]. Komentar-komentar tersebut mencerminkan berbagai opini publik terhadap aktivitas promosi yang dilakukan kreator, mulai dari kualitas produk, kredibilitas kreator, harga dan promosi, pengalaman belanja, hingga kualitas konten yang disajikan [3]. Informasi tersebut memiliki nilai penting bagi pelaku industri, penjual, maupun peneliti untuk memahami persepsi konsumen secara lebih mendalam dan mengevaluasi efektivitas strategi pemasaran digital.

Namun demikian, volume komentar yang sangat besar menjadikan proses analisis opini secara manual tidak efisien dan rentan terhadap subjektivitas. Dalam penelitian ini, jumlah data komentar yang dikumpulkan mencapai 10.000 komentar berbahasa Indonesia. Kondisi tersebut mendorong kebutuhan akan pendekatan otomatis melalui analisis sentimen. Analisis sentimen merupakan cabang *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi polaritas opini pengguna ke dalam kategori positif, negatif, atau netral [4]. Pendekatan ini telah banyak diterapkan pada media sosial, ulasan produk, dan platform *e-commerce* untuk memahami persepsi konsumen terhadap suatu produk atau layanan.

Beberapa penelitian sebelumnya telah menerapkan *Indonesian Bidirectional Encoder Representations from Transformers* (IndoBERT) untuk analisis sentimen pada ulasan *e-commerce* dan media sosial. Penelitian yang dilakukan oleh Widyananda *et al.* menggunakan IndoBERT pada ulasan *e-commerce* berbahasa Indonesia dan memperoleh akurasi sebesar 94,1%, lebih tinggi dibandingkan Support Vector Machine sebesar 89,5% dan Multinomial Naive Bayes sebesar 84,2% [5]. Penelitian lain oleh Dhendra dan Utomo membandingkan IndoBERT dengan beberapa model *transformers* dan *deep learning* pada ulasan layanan *e-government*. Hasil penelitian menunjukkan bahwa IndoBERT memperoleh *F1-score* sebesar 0,882 dan mampu mengungguli Multilingual Bidirectional Encoder Representations from Transformers (Mbert), Cross-lingual Language Model - RoBERTa (XLM-R), Convolutional Neural Network (CNN), serta Bidirectional Long Short-Term Memory (BiLSTM) dalam klasifikasi sentimen berbahasa Indonesia [6]. Selain itu, penelitian oleh Athfaal *et al.* menerapkan IndoBERT pada analisis sentimen kebijakan pajak *marketplace* menggunakan komentar Instagram berbahasa Indonesia dan menghasilkan akurasi sebesar 96% setelah menerapkan *Random Over Sampling* untuk menangani ketidakseimbangan kelas [7]. Hasil penelitian tersebut menunjukkan bahwa IndoBERT memiliki kemampuan yang baik dalam memahami konteks bahasa Indonesia, termasuk teks informal pada media sosial.

Meskipun demikian, sebagian besar penelitian terdahulu masih berfokus pada klasifikasi sentimen tunggal tanpa mempertimbangkan aspek spesifik yang dibahas dalam komentar pengguna. Keterbatasan ini menjadi signifikan dalam konteks *affiliate marketing*, karena satu komentar dapat mengandung lebih dari satu opini dengan polaritas berbeda pada aspek yang berbeda [8]. Sebagai contoh, seorang pengguna dapat memberikan sentimen positif terhadap kualitas produk, namun pada saat yang sama menyampaikan sentimen negatif terhadap harga atau kualitas promosi yang dilakukan kreator. Pendekatan klasifikasi sentimen tunggal berpotensi menghilangkan detail informasi tersebut sehingga mengurangi kedalaman analisis yang dihasilkan.

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan *aspect-based sentiment classification* (ABSC) yang memungkinkan identifikasi sentimen terhadap aspek tertentu dalam suatu teks [9]. Pendekatan ABSC mampu menghasilkan analisis yang lebih detail karena setiap opini dapat dianalisis berdasarkan aspek yang dibahas. Dalam beberapa tahun terakhir, perkembangan model berbasis *transformers*, khususnya BERT dan variannya, menunjukkan performa unggul pada berbagai tugas NLP [10]. Salah satu pengembangan BERT untuk bahasa Indonesia adalah IndoBERT yang dilatih menggunakan korpus berbahasa Indonesia sehingga lebih mampu memahami struktur bahasa, kosakata informal, singkatan, *code-mixing*, dan pola komunikasi pengguna Indonesia dibandingkan model multibahasa [11]. Karakteristik komentar TikTok yang cenderung informal dan mengandung ekspresi emosional semakin memperkuat relevansi penggunaan IndoBERT pada penelitian ini.

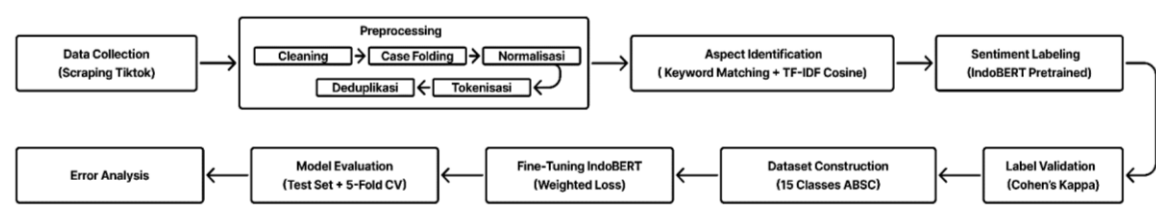
Beberapa penelitian mengenai ABSC menggunakan model BERT dan variannya menunjukkan bahwa pendekatan tersebut mampu memberikan performa yang lebih baik dibandingkan metode klasifikasi sentimen konvensional [12]. Namun, penelitian terkait ABSC pada komentar TikTok berbahasa Indonesia, khususnya dalam konteks *affiliate marketing*, masih sangat terbatas. Sebagian besar penelitian sebelumnya menggunakan data berbahasa Inggris atau berfokus pada ulasan *e-commerce* umum sehingga belum mampu

merepresentasikan karakteristik komentar TikTok berbahasa Indonesia yang cenderung informal, singkat, dan kontekstual. Selain itu, identifikasi aspek pada penelitian terdahulu umumnya masih menggunakan pendekatan berbasis kata kunci sederhana tanpa mempertimbangkan kesamaan konteks antarkomentar.

Berdasarkan kesenjangan penelitian tersebut, penelitian ini bertujuan untuk menganalisis performa IndoBERT yang di-*fine-tune* dalam melakukan klasifikasi sentimen berbasis aspek pada komentar TikTok berbahasa Indonesia. Penelitian ini juga mengevaluasi kestabilan model menggunakan pendekatan *stratified 5-fold cross-validation*. Untuk mencapai tujuan tersebut, penelitian ini mengusulkan kerangka kerja ABSC berbasis IndoBERT yang meliputi tahapan *preprocessing* komentar TikTok, identifikasi aspek menggunakan kombinasi *keyword matching* dan *TF-IDF cosine similarity*, pelabelan sentimen menggunakan *pretrained* IndoBERT, serta proses *fine-tuning* menggunakan *weighted cross-entropy loss* untuk menangani ketidakseimbangan distribusi kelas [13]. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan analisis sentimen berbasis aspek pada data media sosial berbahasa Indonesia, khususnya pada domain *affiliate marketing* di TikTok.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan kerangka kerja Natural Language Processing (NLP) berbasis Aspect-Based Sentiment Classification (ABSC). Alur penelitian meliputi pengumpulan data, *preprocessing* teks, identifikasi aspek, pelabelan sentimen, validasi label, konstruksi dataset berbasis 15 kelas aspek-sentimen, *fine-tuning* model IndoBERT, evaluasi model, serta analisis kesalahan. Pendekatan yang digunakan bersifat hibrida, yaitu mengombinasikan metode berbasis aturan (*keyword matching*) dan pembelajaran berbasis representasi (*TF-IDF cosine similarity*) untuk identifikasi aspek, serta model *deep learning* IndoBERT untuk klasifikasi multi-kelas. Evaluasi model dilakukan menggunakan *train-test split* dan *stratified 5-fold cross-validation* untuk mengukur performa dan *robustness* model [14]. Alur penelitian ditampilkan pada Gambar 1.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Data dikumpulkan melalui proses scraping komentar dari video TikTok terkait affiliate marketing pada periode Januari–Maret 2026. Pengambilan data dilakukan menggunakan layanan Apify dengan API token yang diintegrasikan ke dalam bahasa pemrograman Python. Data diekstraksi secara otomatis menggunakan actor pada platform Apify, dengan metadata meliputi id, *username*, komentar, *like_comment*, waktu, dan *video_id* [15]. Total data yang diperoleh sebanyak 10.000 komentar (10.000 baris, enam kolom) dan digunakan sebagai dataset utama dalam analisis. Proses pengambilan data dilakukan dengan memperhatikan aspek etika penelitian digital. Data yang digunakan hanya berasal dari komentar publik yang tersedia secara terbuka pada platform TikTok tanpa mengakses informasi privat pengguna. Seluruh identitas pengguna dianonimkan dan data digunakan hanya untuk kepentingan akademik serta analisis agregat. Penelitian ini juga tidak menyebarkan informasi personal maupun melakukan interaksi langsung dengan pengguna platform. Sebagai ilustrasi, sebagian data yang diperoleh ditunjukkan pada Tabel 1 berikut.

Tabel 1. Contoh Data Komentar TikTok

id	username	komentar	like_comment	waktu	video_id
1	user123	kak ini beneran ampuh apa cuma endorse doang? 😞	15	2026-01-15 08:21	vid001
2	cantikshop	packaging lucu sih, tapi hasil di aku biasa aja	8	2026-02-03 14:10	vid002
3	reviewbyme	harganya lumayan mahal, ada yang udah coba? worth it ga sih 🤔	5	2026-03-10 19:45	vid003

2.2. Preprocessing Teks

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data teks sebelum proses analisis. Proses ini dirancang untuk menangani karakteristik khas data media sosial, seperti penggunaan bahasa tidak baku, simbol, serta variasi penulisan [16]. Proses meliputi *cleaning* untuk menghapus URL, mention, hashtag, emoji, dan karakter non-alfanumerik, dilanjutkan dengan *case folding*, normalisasi kata menggunakan kamus baku dan kamus domain affiliate TikTok, serta tokenisasi. *Stopword removal* hanya diterapkan pada analisis

eksplorasi dan tidak digunakan pada IndoBERT agar konteks tetap terjaga, sedangkan stemming tidak dilakukan karena IndoBERT mampu memahami variasi morfologis melalui *embedding kontekstual*. Selain itu, dilakukan deduplikasi sehingga diperoleh 6.508 data unik dari 10.000 data awal. Hasil *preprocessing* ditunjukkan pada Tabel 2.

Tabel 2. Contoh Hasil Preprocessing

Tahapan	Hasil
Teks Asli	harganya lumayan mahal, ada yang udah coba? worth it ga sih 🤔
Cleaning	harganya lumayan mahal ada yang udah coba worth it ga sih
Case Folding	harganya lumayan mahal ada yang udah coba worth it ga sih
Normalisasi	harganya lumayan mahal ada yang sudah coba worth it tidak sih
Tokenisasi	[harganya, lumayan, mahal, ada, yang, sudah, coba, worth, it, tidak, sih]
Stopword Removal	[harganya, lumayan, mahal, coba, worth]

2.3. Identifikasi Aspek

Lima aspek utama *affiliate marketing* TikTok diidentifikasi melalui kajian pustaka serta analisis sampel komentar[17]. Masing-masing aspek beserta contoh kata kunci representatif disajikan pada Tabel 3.

Tabel 3. Kategori aspek dan contoh kata kunci representatif

No	Aspek	Contoh Kata Kunci
1	Kualitas Produk	produk bagus, produk rusak, tidak sesuai deskripsi, original, packaging rapi, bahan berkualitas
2	Kepercayaan & Kredibilitas Affiliate	review jujur, kreator bohong, terpercaya, scam, penipuan, paid promote, endorsement
3	Harga & Promosi	harga murah, overpriced, diskon, voucher, free ongkir, <i>worth it, value for money</i>
4	Pengalaman Belanja & Pengiriman	pengiriman cepat, lama sampai, checkout mudah, retur, COD, garansi, TikTok Shop
5	Konten & Pengaruh Kreator	konten informatif, review misleading, FYP, viral, tutorial, unboxing, rekomendasi

Penentuan aspek dilakukan dengan pendekatan dua tahap. Tahap pertama adalah *keyword matching*, yaitu pencocokan komentar terhadap daftar kata kunci pada setiap aspek secara langsung. Metode ini berhasil mengklasifikasikan 3.306 komentar atau sebesar 50,8% dari total 6.508 komentar. Tahap kedua menggunakan pendekatan *TF-IDF cosine similarity* untuk komentar yang belum terklasifikasi pada tahap pertama, yaitu sebanyak 3.202 komentar. Pada tahap ini, setiap komentar direpresentasikan sebagai vektor TF-IDF dan dibandingkan dengan vektor prototipe masing-masing aspek menggunakan persamaan (1).

$$\text{sim}(d, a) = \frac{v_d \cdot v_a}{|v_d| |v_a|} \quad (1)$$

Cosine similarity digunakan untuk mengukur kemiripan antara vektor komentar V_d dan vektor aspek V_a berdasarkan kesamaan pola kata dalam ruang vektor. Semakin mendekati nilai 1, semakin tinggi kesesuaian komentar terhadap aspek tersebut. Pada penelitian ini, ambang batas *cosine similarity* ditetapkan sebesar ($\theta=0.05$) berdasarkan hasil eksplorasi distribusi nilai similaritas untuk memperoleh keseimbangan antara jumlah komentar terklasifikasi dan relevansi aspek yang dihasilkan. *Threshold* yang terlalu rendah menghasilkan banyak komentar yang tidak relevan, sedangkan *threshold* yang terlalu tinggi menyebabkan banyak komentar tidak terklasifikasi. Komentar kemudian diklasifikasikan ke dalam aspek dengan nilai kemiripan tertinggi apabila melebihi ambang batas tersebut. Proses ini menghasilkan 5.183 komentar berlabel aspek (79,6%), sedangkan 1.325 komentar (20,4%) dieliminasi karena tidak memenuhi ambang batas [18].

2.4. Pelabelan Sentiment

Pelabelan sentimen (positif, netral, dan negatif) dilakukan menggunakan model *pretrained mdhugol/indonesia-bert-sentiment-classification* dari Hugging Face melalui *pipeline sentiment-analysis*. Proses inferensi menggunakan *truncation* hingga 512 token untuk menyesuaikan panjang teks dengan batas model. Model ini dipilih karena telah dilatih pada korpus berbahasa Indonesia dan memiliki kemampuan memahami konteks semantik melalui mekanisme *self-attention*, sehingga lebih efektif dalam menangkap sentimen implisit dibandingkan pendekatan berbasis kata kunci [19]. Untuk meningkatkan reliabilitas pelabelan, sebagian sampel data hasil prediksi model *pretrained* diverifikasi secara manual oleh anotator manusia. Validasi ini dilakukan untuk memastikan kesesuaian label sentimen yang dihasilkan model dengan konteks komentar berbahasa Indonesia di TikTok. Dengan demikian, label yang dihasilkan tidak sepenuhnya

bergantung pada *silver standard labeling*, tetapi juga memperoleh verifikasi manusia pada sebagian data sebagai bentuk kontrol kualitas anotasi.

2.5. Validasi Label (Inter-Annotator Agreement)

Konsistensi Validasi konsistensi pelabelan dilakukan menggunakan *Inter-Annotator Agreement* dengan metrik Cohen's Kappa. Hasil model *pretrained* dibandingkan dengan metode berbasis leksikon sebagai pembandingan. Pendekatan leksikon menggunakan 48 kata positif dan 35 kata negatif, dengan mekanisme *lexicon matching* dan agregasi skor sederhana untuk menentukan polaritas sentimen [20]. Nilai kesepakatan dihitung menggunakan Persamaan (2).

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (2)$$

di mana P_o merupakan proporsi kesepakatan observasi antara kedua annotator, sedangkan P_e adalah proporsi kesepakatan yang diharapkan terjadi secara acak. Nilai κ digunakan untuk mengukur tingkat konsistensi antara kedua metode pelabelan, di mana nilai yang lebih tinggi menunjukkan kesesuaian yang lebih baik [21].

2.6. Analisis N-gram dan Ekstraksi Fitur TF-IDF

Analisis n-gram (bigram dan trigram) per kelas sentimen dilakukan menggunakan *CountVectorizer* dari pustaka *scikit-learn* untuk memperoleh frekuensi kemunculan kata berurutan dalam teks. Selain itu, representasi fitur berbasis TF-IDF digunakan untuk memberikan bobot pada setiap term dengan mempertimbangkan tingkat kepentingannya dalam dokumen. Pada tahap ini, parameter *max_features* ditetapkan sebesar 5.000 untuk membatasi ukuran kosakata yang digunakan dalam pemodelan. Hasil ekstraksi fitur menghasilkan matriks TF-IDF dengan dimensi (5183×197) setelah penyesuaian terhadap kosakata yang benar-benar muncul pada korpus penelitian [22].

2.7. Fine-Tuning IndoBERT

Tugas Aspect-Based Sentiment Classification (ABSC) diformulasikan sebagai *sequence classification* dengan 15 label output yang merupakan kombinasi 5 aspek dan 3 kategori sentimen. Model pretrained IndoBERT *indobenchmark/indobert-base-p1* digunakan dengan tambahan *classification head* untuk memetakan representasi teks ke setiap kelas. Pada tahap prapemrosesan digunakan teks hasil normalisasi dalam bentuk lowercase dan kata baku. Stopword tidak dihapus karena kata negasi, seperti “tidak”, berperan penting dalam menentukan polaritas sentimen dan dapat dipahami melalui mekanisme self-attention pada BERT [23]. Untuk mengatasi ketidakseimbangan kelas, digunakan strategi *class weighting* dengan pendekatan *balanced* yang dihitung menggunakan persamaan (3).

$$w_c = \frac{n_{classes} \cdot n_c}{n_{samples}} \quad (3)$$

di mana w_c adalah bobot kelas $ke - c$, $n_{samples}$ adalah total data, $n_{classes}$ adalah jumlah kelas, dan n_c adalah jumlah sampel pada kelas tertentu. Dataset kemudian dibagi secara *stratified* menjadi 80% data latih (4.145 data), 10% validasi (519 data), dan 10% data uji (519 data) untuk menjaga distribusi kelas tetap proporsional pada setiap subset [24]. Konfigurasi *hyperparameter* yang digunakan dalam proses fine-tuning pada Tabel 4.

Tabel 4. Konfigurasi hyperparameter fine-tuning IndoBERT

Hyperparameter	Nilai
Model pra-latih	indobenchmark/indobert-base-p1
Panjang urutan maksimum	128 token
Ukuran batch (train/eval)	16
Learning rate	2×10^{-5}
Epoch	10
Warmup ratio	0,10
Weight decay	0,01
Optimizer	AdamW
Loss function	Weighted Cross-Entropy
Metrik model terbaik	F1-Score (Weighted)
Strategi evaluasi	Setiap epoch
Perangkat komputasi	CPU

2.8. Evaluasi Model

Evaluasi model dilakukan menggunakan dua pendekatan. Pertama, evaluasi pada *hold-out test set* yang mencakup metrik akurasi, presisi, *recall*, serta F1-score dengan rata-rata *weighted* dan *macro* untuk setiap

kelas, disertai analisis *confusion matrix* guna melihat distribusi kesalahan prediksi secara rinci. Kedua, dilakukan Stratified 5-Fold Cross-Validation, di mana model IndoBERT dilatih ulang dari awal pada setiap *fold* dengan konfigurasi yang identik untuk memastikan konsistensi performa. Dari hasil lima *fold* tersebut, dihitung *confidence interval* sebesar 95% menggunakan distribusi *t* sebagai berikut dapat dilihat pada persamaan 4 [25].

$$CI_{95\%} = \bar{x} \pm t_{(0.975,4)} \cdot \frac{s}{\sqrt{5}} \quad (4)$$

dengan $t_{(0.975,4)} = 2.776$, $\bar{x}$ sebagai rata-rata performa, dan s sebagai standar deviasi antar *fold*. Selain itu, analisis kesalahan dilakukan untuk mengidentifikasi jenis misklasifikasi yang terjadi, yang dikategorikan menjadi tiga kelompok, yaitu kesalahan pada aspek saja, kesalahan pada sentimen saja, serta kesalahan pada keduanya. Analisis ini digunakan untuk memahami karakteristik error model secara lebih mendalam

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian dari seluruh tahapan *pipeline*, mulai dari pengumpulan data hingga evaluasi model IndoBERT pada tugas Aspect-Based Sentiment Classification (ABSC). Hasil ditampilkan dalam bentuk statistik dan metrik evaluasi, disertai pembahasan untuk menginterpretasikan temuan utama pada setiap tahap proses. Pembahasan difokuskan pada karakteristik dataset, efektivitas metode identifikasi aspek berbasis *hybrid approach*, kinerja pelabelan sentimen menggunakan model *pretrained*, serta performa IndoBERT pada klasifikasi 15 kelas. Selain itu, analisis kesalahan juga dilakukan untuk mengevaluasi keterbatasan model dalam menangani data komentar media sosial yang bersifat informal.

3.1. Karakteristik Dataset dan Hasil Preprocessing

Dataset awal terdiri dari 10.000 komentar TikTok (10.000 baris dengan enam atribut, yaitu: id, username, komentar, like_comment, waktu, dan video_id). Setelah dilakukan proses deduplikasi berbasis kesamaan teks, sebanyak 4.547 komentar duplikat (45,5%) dihapus, sehingga diperoleh 6.508 komentar unik yang digunakan sebagai data analisis. Statistik lengkap perubahan jumlah data pada setiap tahapan disajikan pada Tabel 5.

Tabel 5. Statistik dataset pada setiap tahapan preprocessing

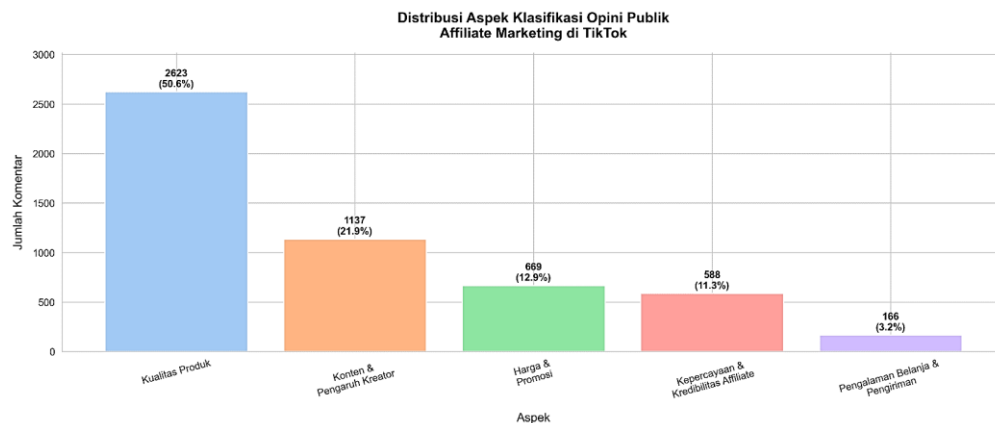
Tahap	Jumlah Data	Persentase dari Awal
Data mentah awal	10.000	100,0%
Setelah deduplikasi	6.508	65,1%
<i>Keyword matching</i> berhasil	3.306	50,8% ^a
TF-IDF similarity berhasil	1.877	28,8% ^a
Dataset akhir (berlabel aspek)	5.183	79,6% ^a
Dibuang (di bawah threshold)	1.325	20,4% ^a
Data latih (Train — 80%)	4.145	-
Data validasi (Validation — 10%)	519	-
Data uji (Test — 10%)	519	-

Persentase dihitung berdasarkan 6.508 komentar setelah deduplikasi. Analisis panjang komentar setelah preprocessing menunjukkan bahwa rata-rata jumlah kata per komentar adalah 4,03 kata dengan standar deviasi 1,28, median 4 kata, serta rentang 1–8 kata. Distribusi panjang komentar cenderung sempit (*low variance*), yang disebabkan oleh proses preprocessing seperti *stopword removal* yang mengurangi jumlah token, sehingga variasi panjang teks menjadi lebih homogen dibandingkan dengan data asli.

3.2. Distribusi Aspek

Distribusi komentar pada setiap aspek setelah proses identifikasi ditunjukkan pada Gambar 2. Aspek Kualitas Produk mendominasi dengan 2.623 komentar (50,6%), yang menunjukkan bahwa pengguna paling banyak menyoroti atribut fisik dan kualitas produk dalam konteks *affiliate marketing* di TikTok. Aspek Konten & Pengaruh Kreator berada pada urutan kedua (21,9%), diikuti oleh Harga & Promosi (12,9%), Kepercayaan & Kredibilitas Affiliate (11,3%), dan Pengalaman Belanja & Pengiriman (3,2%) sebagai aspek dengan proporsi terendah.

Ketimpangan distribusi antar aspek menunjukkan adanya *class imbalance* yang signifikan, dengan rentang proporsi antara 3,2% hingga 50,6%. Kondisi ini menegaskan pentingnya penerapan *weighted cross-entropy loss* pada tahap pelatihan model untuk mengurangi bias terhadap kelas mayoritas.

**Gambar 2.** Distribusi komentar per aspek

3.3. Distribusi Sentimen dan Validasi Label

Pelabelan sentimen menggunakan model *pretrained* IndoBERT menghasilkan distribusi sebagai berikut: positif sebanyak 2.258 komentar (43,6%), negatif sebanyak 1.961 komentar (37,8%), dan netral sebanyak 964 komentar (18,6%). Distribusi silang antara aspek dan sentimen disajikan pada Tabel 6.

Tabel 6. Distribusi komentar berdasarkan aspek dan sentimen (*crosstab*)

Aspek	Positif	Netral	Negatif	Total
Harga & Promosi	436 (65,2%)	123 (18,4%)	110 (16,4%)	669
Kepercayaan & Kredibilitas Affiliate	308 (52,4%)	24 (4,1%)	256 (43,5%)	588
Konten & Pengaruh Kreator	599 (52,7%)	204 (17,9%)	334 (29,4%)	1.137
Kualitas Produk	908 (34,6%)	469 (17,9%)	1.246 (47,5%)	2.623
Pengalaman Belanja & Pengiriman	7 (4,2%)	144 (86,7%)	15 (9,0%)	166
Total	2.258 (43,6%)	964 (18,6%)	1.961 (37,8%)	5.183

Tabel 6 menunjukkan bahwa aspek Kualitas Produk memiliki proporsi sentimen negatif tertinggi (47,5%), yang mengindikasikan adanya kesenjangan antara ekspektasi yang dibangun melalui konten *affiliate* dan pengalaman aktual konsumen. Sebaliknya, aspek Pengalaman Belanja & Pengiriman didominasi oleh sentimen netral (86,7%) yang merefleksikan komentar yang bersifat informasional. Selain itu, tingginya sentimen negatif pada aspek Kepercayaan & Kredibilitas Affiliate (43,5%) menunjukkan adanya isu kepercayaan terhadap otentisitas kreator. Validasi label menggunakan Cohen's Kappa menghasilkan nilai $\kappa = 0,5682$ (*moderate agreement*). Matriks perbandingan ditunjukkan pada Tabel 7.

Tabel 7. Matriks perbandingan label model *pretrained* dan leksikon

Model Pretrained \ Leksikon	Negatif	Netral	Positif
Negatif	906	612	443
Netral	7	803	154
Positif	0	249	2.009

Perbedaan utama terjadi pada kelas negatif, yang sering diklasifikasikan sebagai netral atau positif oleh metode leksikon. Hal ini menunjukkan keterbatasan pendekatan berbasis kamus dalam menangkap sentimen kontekstual pada teks pendek. Oleh karena itu, meskipun nilai Kappa berada pada tingkat moderat, model IndoBERT tetap digunakan sebagai *silver standard* karena kemampuannya dalam memahami konteks semantik. Namun, potensi ambiguitas akibat panjang komentar yang pendek tetap menjadi keterbatasan dalam penelitian ini.

3.4. Distribusi Label dan Ketidakseimbangan Kelas

Berdasarkan distribusi aspek dan sentimen, kombinasi keduanya menghasilkan 15 kelas label dengan tingkat ketidakseimbangan yang sangat ekstrem (Tabel 7). Kelas terbesar, Kualitas Produk–Negatif (1.246 sampel), dibandingkan dengan kelas terkecil, Pengalaman Belanja & Pengiriman–Positif (7 sampel), menunjukkan rasio sekitar 178:1. Dapat dilihat pada Tabel 8.

Tabel 8. Distribusi 15 kelas dan bobot

No	Kelas (Aspek–Sentimen)	Jumlah	Bobot Kelas
1	Kualitas Produk–Negatif	1.246	0,2774

No	Kelas (Aspek–Sentimen)	Jumlah	Bobot Kelas
2	Kualitas Produk–Positif	908	0,3806
3	Konten & Pengaruh Kreator–Positif	599	0,5769
4	Kualitas Produk–Netral	469	0,7369
5	Harga & Promosi–Positif	436	0,7941
6	Konten & Pengaruh Kreator–Negatif	334	1,0350
7	Kepercayaan & Kredibilitas Affiliate–Positif	308	1,1233
8	Kepercayaan & Kredibilitas Affiliate–Negatif	256	1,3546
9	Konten & Pengaruh Kreator–Netral	204	1,6850
10	Pengalaman Belanja & Pengiriman–Netral	144	2,3822
11	Harga & Promosi–Netral	123	2,7912
12	Harga & Promosi–Negatif	110	3,1402
13	Kepercayaan & Kredibilitas Affiliate–Netral	24	13,8167
14	Pengalaman Belanja & Pengiriman–Negatif	15	23,0278
15	Pengalaman Belanja & Pengiriman–Positif	7	55,2667

Tabel 8 menunjukkan bahwa distribusi kelas didominasi oleh aspek Kualitas Produk, terutama pada sentimen positif dan negatif, sedangkan aspek Pengalaman Belanja & Pengiriman memiliki jumlah data paling sedikit. Ketidakseimbangan kelas ini berpotensi menyebabkan model bias terhadap kelas mayoritas. Oleh karena itu, diterapkan class weight dengan memberikan bobot lebih tinggi pada kelas minoritas agar model lebih sensitif terhadap kelas yang jarang muncul dan menghasilkan performa klasifikasi yang lebih seimbang.

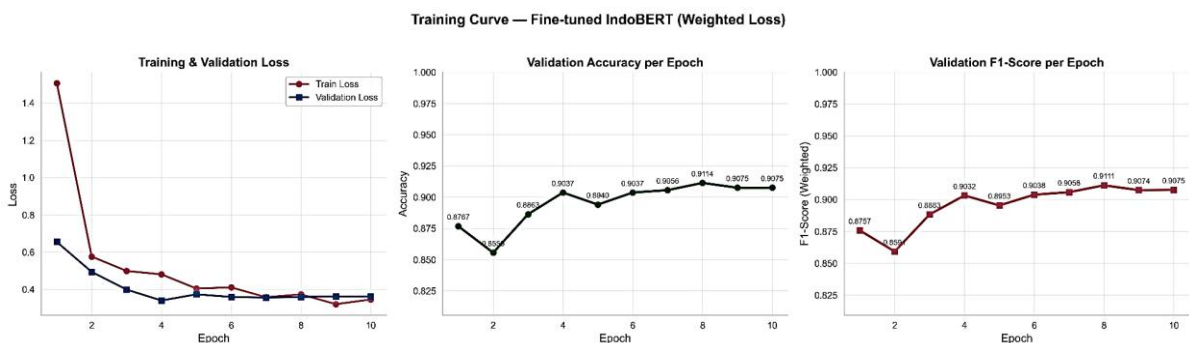
3.5. Pelatihan dan Kurva Learning Fine-tuned IndoBERT

Proses fine-tuning dilakukan selama 10 *epoch* dengan pemilihan model terbaik berdasarkan *weighted F1-Score* pada data validasi. Ringkasan kinerja model disajikan pada Tabel 9.

Tabel 9. Kurva learning *fine-tuning* IndoBERT

Epoch	Training Loss	Validation Loss	Accuracy
1	0,5515	0,6541	0,8767
2	0,4645	0,4932	0,8555
3	0,4374	0,3985	0,8863
4	0,6011	0,3397	0,9037
5	0,4700	0,3729	0,8940
6	0,3704	0,3591	0,9037
7	0,3986	0,3552	0,9056
8	0,2882	0,3587	0,9114
9	0,4593	0,3618	0,9075
10	0,3354	0,3619	0,9075

Model *Training loss* menurun tajam pada awal pelatihan dan kemudian melandai, dengan sedikit fluktuasi pada *epoch* ke-4. *Validation loss* mencapai minimum pada *epoch* ke-4 dan cenderung stabil setelahnya, menandakan model mulai konvergen. Sementara itu, akurasi dan F1-Score terus meningkat hingga *epoch* ke-8 sebelum stabil. Perbedaan tren antara *loss* dan metrik evaluasi menunjukkan bahwa peningkatan performa klasifikasi masih terjadi meskipun *loss* telah mendatar, terutama karena pengaruh distribusi kelas yang tidak seimbang. Tidak ditemukan indikasi *overfitting*, sehingga *epoch* ke-8 dipilih sebagai titik optimal dengan keseimbangan terbaik antara konvergensi dan performa. Visualisasi kurva pelatihan pada Gambar 3 memperkuat temuan pada Tabel 8, yaitu bahwa model mencapai titik optimal pada *epoch* ke-8, yang merepresentasikan keseimbangan terbaik antara konvergensi *loss* dan performa klasifikasi.



Gambar 3. Kurva *learning* IndoBERT selama *fine-tuning*

3.6. Performa Model pada Hold-out Test Set

Evaluasi kinerja model terbaik yang diperoleh pada *epoch* ke-8 dilakukan menggunakan *hold-out test set* ($n = 519$) untuk mengukur kemampuan generalisasi model terhadap data yang tidak terlihat selama proses pelatihan. Hasil evaluasi tersebut disajikan pada Tabel 10.

Tabel 10. Performa keseluruhan *fine-tuned* IndoBERT

Metrik	Weighted Average	Macro Average
Presisi	90,60%	85,72%
<i>Recall</i>	89,79%	81,23%
F1-Score	89,74%	82,29%
Akurasi	89,79%	—

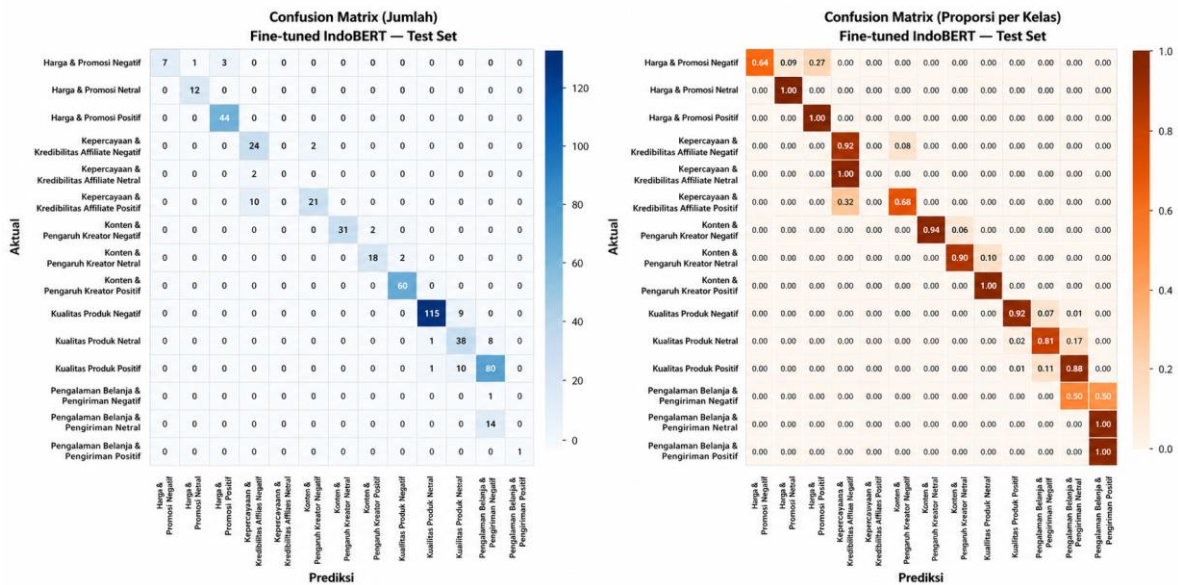
Model mencapai akurasi 89,79% dan weighted F1-Score 89,74%, yang menunjukkan performa sangat baik pada klasifikasi 15 kelas dengan distribusi tidak seimbang. Selisih antara metrik weighted dan macro (~7–8%) mengindikasikan adanya variasi performa antar kelas, terutama pada kelas minoritas. Performa per kelas disajikan pada Tabel 11.

Tabel 11. Performa keseluruhan *fine-tuned* IndoBERT

Kelas (Aspek Sentimen)	Presisi	Recall	F1-Score	Support
Harga & Promosi Negatif	1,00	0,64	0,78	11
Harga & Promosi Netral	0,92	1,00	0,96	12
Harga & Promosi Positif	0,94	1,00	0,97	44
Kepercayaan & Kredibilitas Affiliate Negatif	0,67	0,92	0,77	26
Kepercayaan & Kredibilitas Affiliate Netral	0,00	0,00	0,00	2
Kepercayaan & Kredibilitas Affiliate Positif	0,91	0,68	0,78	31
Konten & Pengaruh Kreator Negatif	1,00	0,94	0,97	33
Konten & Pengaruh Kreator Netral	1,00	0,90	0,95	20
Konten & Pengaruh Kreator Positif	0,94	1,00	0,97	60
Kualitas Produk Negatif	0,98	0,92	0,95	125
Kualitas Produk Netral	0,67	0,81	0,73	47
Kualitas Produk Positif	0,90	0,88	0,89	91
Pengalaman Belanja & Pengiriman Negatif	1,00	0,50	0,67	2
Pengalaman Belanja & Pengiriman Netral	0,93	1,00	0,97	14
Pengalaman Belanja & Pengiriman Positif	1,00	1,00	1,00	1
Accuracy	—	—	0,90	519
Macro Average	0,86	0,81	0,82	519
Weighted Average	0,91	0,90	0,90	519

Berdasarkan Tabel 10, F1-Score tertinggi diperoleh pada kelas Pengalaman Belanja & Pengiriman | Positif (1,00), Harga & Promosi | Positif (0,97), serta Konten & Pengaruh Kreator | Negatif (0,97). Sebaliknya, performa terendah terdapat pada Kepercayaan & Kredibilitas *Affiliate* | Netral (0,00) akibat jumlah sampel yang sangat terbatas ($n = 2$). Secara umum, kelas dengan *support* besar menunjukkan performa lebih stabil dibandingkan kelas minoritas, sehingga ketidakseimbangan data masih menjadi tantangan utama. Meskipun *class weighting* membantu meningkatkan kinerja, model belum optimal pada kelas dengan distribusi sangat kecil. Untuk memperoleh pemahaman yang lebih mendalam terhadap pola kesalahan klasifikasi, dilakukan analisis menggunakan confusion matrix dalam dua bentuk, yaitu matriks jumlah (*absolute*) dan matriks proporsi (*normalized*). Matriks jumlah digunakan untuk mengidentifikasi distribusi kesalahan secara aktual, sedangkan matriks proporsi digunakan untuk mengevaluasi performa model secara relatif pada setiap kelas, sebagaimana ditunjukkan pada Gambar 4.

Hasil analisis menunjukkan bahwa sebagian besar prediksi berada pada diagonal utama, yang mengindikasikan akurasi tinggi. Kesalahan klasifikasi antar aspek hampir tidak ditemukan, sehingga model terbukti mampu memahami konteks aspek dengan baik. Namun, kesalahan masih terjadi pada polaritas sentimen dalam aspek yang sama, terutama antara positif–negatif pada aspek *Kepercayaan & Kredibilitas Affiliate* serta netral–positif pada *Kualitas Produk*. Performa rendah pada kelas minoritas, seperti *Kepercayaan & Kredibilitas Affiliate* | *Netral*, menunjukkan bahwa ketidakseimbangan data masih menjadi kendala utama. Secara keseluruhan, model sangat baik dalam identifikasi aspek, tetapi masih memerlukan peningkatan dalam klasifikasi sentimen, khususnya pada data yang ambigu dan tidak seimbang.



Gambar 4. Confusion Matrix

3.7. Evaluasi Robustness Stratified 5-Fold Cross-Validation

Untuk menguji stabilitas model, dilakukan *stratified 5-fold cross-validation* pada seluruh dataset berlabel ($n = 5.183$) dengan melatih ulang IndoBERT pada setiap *fold*. Hasil evaluasi disajikan pada Tabel 12.

Tabel 12. Performa keseluruhan *fine-tuned* IndoBERT

Fold	Akurasi	Presisi (W)	Recall (W)	F1-Score (W)	Presisi (M)
1	90,26%	90,60%	90,26%	90,06%	82,67%
2	91,71%	92,13%	91,71%	91,60%	86,78%
3	91,71%	91,60%	91,71%	91,48%	80,00%
4	90,54%	91,40%	90,54%	90,45%	83,06%
5	90,93%	91,28%	90,93%	90,80%	85,99%
Rata-rata	91,03%	91,40%	91,03%	90,88%	83,70%
Std. Deviasi	$\pm 0,66\%$	$\pm 0,55\%$	$\pm 0,66\%$	$\pm 0,66\%$	$\pm 2,74\%$

Hasil menunjukkan konsistensi performa yang tinggi pada seluruh *fold*, dengan akurasi pada rentang 90,26%–91,71%. Nilai simpangan baku yang rendah ($\pm 0,66\%$) menunjukkan bahwa model stabil dan tidak bergantung pada pembagian data tertentu. Sebaliknya, variasi yang lebih besar pada metrik macro ($\pm 2,8\%$) mengindikasikan bahwa performa antar kelas masih bervariasi, terutama pada kelas minoritas. Untuk memperkuat interpretasi ini, analisis dilanjutkan dengan estimasi interval kepercayaan 95% pada metrik utama.

3.8. Confidence Interval 95 Persen

Berdasarkan hasil *stratified 5-fold cross-validation*, *confidence interval* (CI) 95% dihitung menggunakan distribusi-t dengan empat derajat kebebasan $t_{0,975,4} = 2.776$. Hasil estimasi disajikan pada Tabel 13.

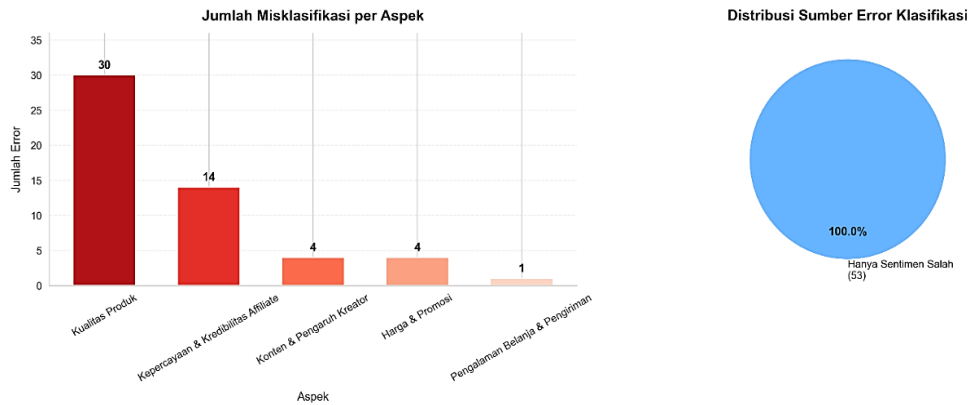
Tabel 13. Performa keseluruhan *fine-tuned* IndoBERT

Metrik	Rata-rata $\pm$ SD	95% CI
Akurasi	$91,03\% \pm 0,66\%$	[90,20% ; 91,85%]
F1-Score (<i>Weighted</i>)	$90,88\% \pm 0,66\%$	[90,06% ; 91,70%]
Presisi (<i>Weighted</i>)	$91,40\% \pm 0,55\%$	[90,71% ; 92,09%]
Recall (<i>Weighted</i>)	$91,03\% \pm 0,66\%$	[90,20% ; 91,85%]
Metrik	Rata-rata $\pm$ SD	95% CI
F1-Score (<i>Macro</i>)	$81,93\% \pm 2,84\%$	[78,40% ; 85,46%]
Presisi (<i>Macro</i>)	$83,70\% \pm 2,74\%$	[80,30% ; 87,10%]
Recall (<i>Macro</i>)	$82,12\% \pm 2,86\%$	[78,56% ; 85,67%]

Metrik *weighted* memiliki rentang *confidence interval* yang sempit ($\approx 1,65\%$), menunjukkan estimasi performa yang stabil dan andal. Sebaliknya, metrik macro memiliki rentang yang lebih lebar, mencerminkan variasi performa pada kelas minoritas. Selisih kecil antara hasil hold-out test set (89,79%) dan cross-validation (91,03%) menunjukkan tidak adanya indikasi *overfitting*.

3.9. Analisis Kesalahan

Dari 519 sampel pada test set, sebanyak 53 sampel (10,2%) mengalami misklasifikasi dan seluruhnya berasal dari kesalahan prediksi sentimen, tanpa kesalahan pada identifikasi aspek. Hal ini menunjukkan bahwa model sangat baik dalam mengenali aspek, namun masih menghadapi tantangan pada klasifikasi sentimen. Kesalahan utama meliputi kebingungan positif-negatif pada aspek Kepercayaan & Kredibilitas Affiliate, positif-netral pada Kualitas Produk, serta negatif-positif pada Harga & Promosi. Distribusi misklasifikasi per aspek dan sumber kesalahan ditunjukkan pada Gambar 5.



Gambar 5. Distribusi misklasifikasi per aspek dan sumber kesalahan.

Distribusi misklasifikasi pada Gambar 5 menunjukkan bahwa kesalahan paling banyak terjadi pada aspek Kualitas Produk, diikuti oleh Kepercayaan & Kredibilitas Affiliate, sedangkan aspek lainnya relatif rendah. Secara keseluruhan, hasil ini menegaskan bahwa peningkatan performa model selanjutnya perlu difokuskan pada kemampuan memahami nuansa sentimen yang lebih kompleks dan kontekstual.

3.10. Contoh Prediksi Model

Contoh prediksi model pada beberapa komentar baru disajikan pada Tabel 15, yang juga menampilkan tingkat kepercayaan (*confidence score*) untuk setiap prediksi.

Tabel 15. Contoh prediksi model *fine-tuned* IndoBERT pada komentar baru

Teks Komentar	Aspek Prediksi	Sentimen	Confidence
produk bagus banget kualitas oke sesuai deskripsi	Kualitas Produk	Positif	0,9459
kecewa barang tidak sesuai gambar yang di review kreator	Kualitas Produk	Negatif	0,9931
harga murah tapi kualitas lumayan worth it lah	Harga & Promosi	Positif	0,9674
pengiriman lama banget sudah seminggu belum sampai	Kualitas Produk	Negatif	0,4275
kreator ini jujur reviewnya terpercaya recommended	Kepercayaan & Kredibilitas Affiliate	Positif	0,6266
affiliate penipu produk palsu tidak sesuai	Kualitas Produk	Negatif	0,9954
konten review nya informatif dan menarik	Konten & Pengaruh Kreator	Positif	0,3666
seller tidak responsif chat tidak dibalas	Konten & Pengaruh Kreator	Negatif	0,7228

Berdasarkan Tabel 15, sebagian besar prediksi menunjukkan *confidence* yang tinggi pada komentar dengan sentimen yang jelas. Namun, beberapa prediksi memiliki *confidence* rendah, terutama pada komentar yang ambigu, seperti terkait pengiriman atau konten. Selain itu, terdapat indikasi overlap antar aspek pada beberapa kasus, yang menunjukkan bahwa batas antar aspek masih belum sepenuhnya terpisah dalam representasi model. Nilai *confidence* yang rendah berpotensi dimanfaatkan sebagai indikator ketidakpastian dalam skema *human-in-the-loop*.

3.11. Diskusi

Hasil penelitian ini menunjukkan performa yang kompetitif dibandingkan dengan penelitian sebelumnya yang menggunakan IndoBERT pada analisis sentimen berbahasa Indonesia. Widyandana *et al.* memperoleh akurasi sebesar 94,1% pada ulasan *e-commerce* [5], sedangkan Dhendra dan Utomo memperoleh *F1-score* sebesar 0,882 pada data layanan *e-government* [6]. Performa pada penelitian ini sedikit lebih rendah karena tugas yang dilakukan lebih kompleks, yaitu *Aspect-Based Sentiment Classification* (ABSC) dengan 15 kombinasi kelas aspek-sentimen pada komentar TikTok yang bersifat singkat, informal, ambigu, dan mengandung *slang*. Namun demikian, hasil yang diperoleh tetap menunjukkan bahwa IndoBERT mampu

memahami konteks bahasa Indonesia pada media sosial dengan baik. Temuan ini juga selaras dengan penelitian ABSC berbasis BERT pada data berbahasa Inggris [12], yang menunjukkan bahwa model *transformers* efektif dalam klasifikasi sentimen berbasis aspek.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem *Aspect-Based Sentiment Classification* (ABSC) pada komentar TikTok terkait *affiliate marketing* menggunakan model *fine-tuned* IndoBERT. Hasil penelitian menunjukkan bahwa pendekatan hibrida berbasis *keyword matching* dan *TF-IDF cosine similarity* mampu mengidentifikasi aspek komentar secara efektif, sedangkan IndoBERT menunjukkan performa yang baik dalam klasifikasi sentimen berbasis aspek pada data media sosial berbahasa Indonesia dengan akurasi sebesar 89,79%. Evaluasi model juga menunjukkan performa yang stabil dan konsisten pada pengujian *cross-validation*, meskipun dataset memiliki ketidakseimbangan kelas yang cukup signifikan. Analisis kesalahan menunjukkan bahwa model lebih baik dalam mengenali aspek daripada menentukan polaritas sentimen, terutama pada komentar yang ambigu dan kelas minoritas. Secara keseluruhan, penelitian ini menunjukkan bahwa IndoBERT memiliki kemampuan yang baik dalam memahami konteks komentar TikTok yang informal dan tidak terstruktur. Penelitian selanjutnya disarankan untuk melakukan augmentasi data pada kelas minoritas, mengeksplorasi model *transformer* yang lebih besar, serta memperluas pengujian pada platform media sosial dan *e-commerce* lainnya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Teknokrat Indonesia, khususnya Fakultas Teknik dan Ilmu Komputer, atas dukungan akademik dan fasilitas yang diberikan selama proses penelitian ini. Penulis juga mengapresiasi kontribusi berbagai pihak yang telah memberikan masukan, wawasan, serta dukungan teknis yang sangat membantu dalam penyusunan dan penyempurnaan penelitian ini.

REFERENSI

- [1] Ms. Mahdi Faza, "Penerapan Model Indobert Untuk Deteksi Potensi Sumber Stres Dalam Teks Media Sosial," *Jurnal Rekayasa Sistem Informasi Dan Teknologi*, Vol. 3, No. 1, 2025, Doi: <https://doi.org/10.70248/Jrsit.V3i1.3046>.
- [2] T. Putri, Siti Nurhaliza, And Dodi Vionanda, "Analisis Sentimen Penggunaan Aplikasi Youtube Menggunakan Metode Naïve Bayes," *Unp Journal Of Statistics And Data Science*, Vol. 3, No. 1, Pp. 60–66, Feb. 2025, Doi: 10.24036/Ujds/Vol3-Iss1/343.
- [3] I. Koswara, "Global Komunika Diskursus Digital Dan Suara Publik: Komentar Youtube Dalam Kasus #Indonesiagelap," *Jurnal Global Komunika*, Vol. 8, No. 2, Pp. 101–110, 2025, Doi: 10.5678/Gk.V8i2.2025.
- [4] F. A. D. P. Febrianti, F. Hamami, And R. Y. Fa'rifah, "Aspect-Based Sentiment Analysis Terhadap Ulasan Aplikasi Flip Menggunakan Pembobotan Term Frequency-Inverse Document Frequency (Tf-Idf) Dengan Metode Klasifikasi K-Nearest Neighbors (K-Nn)," *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, Vol. 4, No. 3, Pp. 1858–1873, Sep. 2023, Doi: 10.35870/Jimik.V4i3.429.
- [5] 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.
- [6] Dhendra And V. Gayuh Utomo, "Benchmarking Indobert And Transformer Models For Sentiment Classification On Indonesian E-Government Service Reviews," *Jurnal Transformatika*, Vol. 23, No. 1, Pp. 86–95, Jul. 2025, Doi: 10.26623/Transformatika.V23i1.12095.
- [7] K. Athfaal, R. Ramadhani, And M. R. Nashrulloh, "Public Sentiment Analysis Of Marketplace Tax Policy Using Indobert," *Jurnal Sistem Informasi, Teknologi Informasi Dan Komunikasi*, No. 2, P. 56, Nov. 2025, Doi: 10.33364/Sistematik/V.1-1.2403.
- [8] D. K. Putri, G. Karyono, And I. Tahyudin, "Analisis Sentimen E-Commerce Dengan Optimasi Indobert Menggunakan Bayesian Optuna," *Jurnal Pendidikan Dan Teknologi Indonesia*, Vol. 5, No. 12, Pp. 3538–3547, Dec. 2025, Doi: 10.52436/1.Jpti.1208.
- [9] M. D. R. Wahyudi, "Penerapan Algoritma Cosine Similarity Pada Text Mining Terjemah Al-Qur'an Berdasarkan Keterkaitan Topik," *Semesta Teknika*, Vol. 22, No. 1, 2019, Doi: 10.18196/St.221235.
- [10] J. C. Setiawan, K. M. Lhaksana, And B. Bunyamin, "Sentiment Analysis Of Indonesian Tiktok Review Using Lstm And Indobertweet Algorithm," *Jipi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, Vol. 8, No. 3, Pp. 774–780, Aug. 2023, Doi: 10.29100/Jipi.V8i3.3911.
- [11] M. I. Abidin And E. W. Pamungkas, "Analisis Sentimen Terhadap Timnas Indonesia Di Piala Asia 2023 Dengan Model Transformer Berbahasa Indonesia," *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, Vol. 10, No. 2, Pp. 482–496, Jul. 2025, Doi: 10.36341/Rabit.V10i2.6142.
- [12] R. I. Perwira, V. A. Permadi, D. I. Purnamasari, And R. P. Agusdin, "Domain-Specific Fine-Tuning Of Indobert For Aspect-Based Sentiment Analysis In Indonesian Travel User-Generated Content," *Journal*

- Of Information Systems Engineering And Business Intelligence*, Vol. 11, No. 1, Pp. 30–40, Feb. 2025, Doi: 10.20473/Jisebi.11.1.30-40.
- [13] G. Maharani, L. Putri, And R. S. Hayu, “The Influence Of Affiliate Marketing Content, Fyp Algorithm, And E-Wom On Tiktok Social Media On Generation Z Purchasing Decisions,” *Journal Of Digital Marketing And Business*, Vol. 4, No. 2, Pp. 112–125, 2025, Doi: 10.5678/Jdmb.V4i2.2025.
 - [14] U. Krishna, “Youtube Comments Sentiments Analysis,” *Int. J. Res. Appl. Sci. Eng. Technol.*, Vol. 13, No. 1, Pp. 875–880, Jan. 2025, Doi: 10.22214/Ijrasnet.2025.66475.
 - [15] J. Olusegun, E. Chris, B. Favor, And T. Ramon, “Aspect-Based Sentiment Analysis Of Customer Reviews In E-Commerce,” *International Journal Of Computer Science And Data Analytics*, Vol. 5, No. 1, Pp. 15–24, 2024, Doi: 10.4567/Ijcsda.V5i1.2024.
 - [16] S. A. Nugraha, “Penerapan Lexicon Based Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Danantara,” *Jurnal Sistem Cerdas Indonesia*, Vol. 2, No. 1, Pp. 10–18, 2025, Doi: 10.7891/Jsci.V2i1.2025.
 - [17] A. Wazira And K. Azmi, “Pengujian Instrument Kekedisiplinan Guru Terhadap Prestasi Akademik Peserta Didik,” *Jurnal Riset Edukasi Vokasi, Inovasi, Dan Sosial (J-Revisi)*, Vol. 1, No. 2, Pp. 45–53, Dec. 2024, Doi: 10.1234/Jrevisi.V1i2.2024
 - [18] M. Bais, A. Hakiki, And Y. Darmi, “Penerapan Algoritma Machine Learning Svm Dan Nbc Pada Sentimen Analisis Komentar Youtube Program Pengaduan Masyarakat Lapor Mas Wapres,” *Jurnal Ilmu Manajemen*, Vol. 4, No. 1, Pp. 66–74, 2025, Doi: 10.38035/Jim.V4i1.
 - [19] M. S. Munir, J. Wati, R. Wulan, And S. B. Wati, “Pengaruh Affiliate Marketing Terhadap Keputusan Pembelian Di Tiktok Shop,” *Riggs: Journal Of Artificial Intelligence And Digital Business*, Vol. 4, No. 2, Pp. 5092–5097, Jul. 2025, Doi: 10.31004/Riggs.V4i2.1390.
 - [20] A. Irianti, H. Halimah, S. Sutedi, And M. Agariana, “Integration Of Bert And Svm In Sentiment Analysis Of Twitter/X Regarding Constitutional Court Decision No. 60/Puu-Xxii/2024,” *Jurnal Teknik Informatika (Jutif)*, Vol. 6, No. 2, Pp. 469–482, Apr. 2025, Doi: 10.52436/1.Jutif.2025.6.2.4068.
 - [21] A. R. Andriawan, M. Mustakim, And R. Novita, “Sentiment Analysis Classification Of Political Parties On Twitter Using Gated Recurrent Unit Algorithm And Natural Language Processing,” *Journal Of Informatics And Telecommunication Engineering*, Vol. 7, No. 2, Pp. 514–522, Jan. 2024, Doi: 10.31289/Jite.V7i2.10709.
 - [22] M. Jamil, H. Hadiyanto, And R. Sanjaya, “Sentiment Analysis: Classifying Public Comments On Youtube In Disaster Management Simulation In Indonesia Using Naïve Bayes And Support Vector Machine,” *Ingenierie Des Systemes D’information*, Vol. 29, No. 2, Pp. 437–446, Apr. 2024, Doi: 10.18280/Isi.290205.
 - [23] Y. A. Harahap And H. Sazali, “Digital Marketing Tiktok Affiliate Pada Pemasaran Produk Fashion,” *R2j*, Vol. 6, No. 5, 2024, Doi: 10.38035/Rrj.V6i5.
 - [24] T. Tanika Marunduri And S. Sandino Berutu, “Implementasi Aspect-Based Sentiment Analysis Berbasis Indobert Pada Program Sekolah Rakyat,” *Jurnal Algoritme*, Vol. 6, No. 2, Pp. 395–406, 2026, Doi: 10.35957/Algoritme.V6i2.15734.
 - [25] F. Salamah, S. Fatimah, N. Permatasari, M. Afrina, And A. Ibrahim, “Metode Aspect-Based Sentiment Analysis (Absa) Pada Ulasan Songket Di Google Maps Menggunakan Indobert,” *Prosiding Seminar Nasional Informatika Dan Data Science*, Vol. 3, No. 1, Pp. 88–96, 2026, Doi: 10.5678/Snids.V3i1.2026.