TECHNOVATAR Jurnal Teknologi, Industri dan Informasi
Vol. 4, No. 3, July 2026, pp. XX–XX
Open Access:
https://journal.awatarapublisher.com/index.php/technovatar
ISSN: 3031-8874, DOI: 10.61434/technovatar.v4i2.376

ANALISIS PERBANDINGAN KINERJA INDOBERT DAN MULTILINGUAL BERT (MBERT) PADA KLASIFIKASI SENTIMEN ULASAN FITUR SHOPEE LIVE DI GOOGLE PLAY STORE

Hendri Rosmawan[1], Ardiyansyah[1], Muhamad Fuad Azis[1], Jaisyi Bagir Rafsyahid[1]
[1] Program Studi Informatika, Fakultas Teknik, Institut Teknologi dan Kesehatan Mahardika, Cirebon, Indonesia

e-mail: hendri.rosmawan@mahardika.ac.id

INFO ARTIKEL
Sejarah Artikel:
Diterima, Maret 2026
Direvisi, April 2026
Disetujui, April 2026

KEYWORDS
IndoBERT
mBERT
Sentiment analysis
Shopee Live
Text classification

ABSTRACT

The Shopee Live streaming e-commerce feature generates a large volume of user reviews on the Google Play Store, yet the text is dominated by informal Indonesian, slang, and code-mixing, making it difficult to analyze using conventional sentiment analysis approaches. This study analyzes and compares the performance of IndoBERT and Multilingual BERT (mBERT) in classifying sentiment of Shopee Live feature reviews into three classes: Positive, Negative, and Neutral. The study follows the CRISP-DM framework, encompassing web scraping of reviews from January-June 2025, filtering, text preprocessing, manual labeling, dataset splitting (80:10:10), fine-tuning both models using HuggingFace Transformers on Google Colaboratory, and evaluation using confusion matrices. The cleaned dataset comprised 5,240 reviews with a class distribution of 44.96% Positive, 35.79% Negative, and 19.25% Neutral. Evaluation results on the test set show that IndoBERT achieved an Accuracy of 89.7%, Precision of 89.5%, Recall of 89.7%, and F1-Score of 89.6%, outperforming mBERT, which achieved an Accuracy of 84.3%, Precision of 84.5%, Recall of 84.3%, and F1-Score of 84.4%. Both models exhibited lower performance on the Neutral class compared to the Positive and Negative classes. These findings confirm that IndoBERT is a more optimal model for sentiment classification of informal Indonesian-language e-commerce reviews, and can be recommended as an AI-based analytical instrument for product development teams to automatically and continuously monitor user sentiment.

This work is licensed under a Creative Commons Attribution 4.0 International License.

KATA KUNCI
IndoBERT
mBERT
Analisis sentimen
Shopee Live
Klasifikasi teks

ABSTRAK

Fitur live streaming e-commerce seperti Shopee Live menghasilkan volume ulasan pengguna yang besar pada Google Play Store, namun teksnya didominasi bahasa Indonesia informal, slang, dan code-mixing sehingga sulit dianalisis menggunakan pendekatan analisis sentimen konvensional. Penelitian ini menganalisis dan membandingkan kinerja IndoBERT dan Multilingual BERT (mBERT) dalam mengklasifikasikan sentimen ulasan fitur Shopee Live ke dalam tiga kelas: Positif, Negatif,

CORRESPONDING AUTHOR

Hendri Rosmawan, M.Kom.
Program Studi Informatika, Fakultas Teknik, Universitas Insan Mahardika Cirebon, Indonesia
hendri.rosmawan@mahardika.ac.id

dan Netral. Penelitian menggunakan kerangka CRISP-DM meliputi web scraping data ulasan periode Januari-Juni 2025, filtering, text preprocessing, pelabelan manual, pembagian dataset (80:10:10), fine-tuning kedua model menggunakan HuggingFace Transformers pada Google Colaboratory, hingga evaluasi menggunakan confusion matrix. Dataset bersih berjumlah 5.240 ulasan dengan distribusi Positif 44,96%, Negatif 35,79%, dan Netral 19,25%. Hasil evaluasi pada data uji menunjukkan IndoBERT mencapai Accuracy 89,7%, Precision 89,5%, Recall 89,7%, dan F1-Score 89,6%, mengungguli mBERT yang mencapai Accuracy 84,3%, Precision 84,5%, Recall 84,3%, dan F1-Score 84,4%. Kedua model menunjukkan performa lebih rendah pada kelas Netral dibandingkan kelas Positif dan Negatif. Temuan ini menegaskan bahwa IndoBERT merupakan model yang lebih optimal untuk klasifikasi sentimen teks ulasan e-commerce berbahasa Indonesia informal, dan dapat direkomendasikan sebagai instrumen analitik berbasis kecerdasan buatan bagi tim pengembang produk dalam memantau sentimen pengguna secara otomatis dan berkelanjutan.

PENDAHULUAN

Fitur live streaming e-commerce seperti Shopee Live telah menjadi salah satu inovasi belanja daring yang paling diminati di Indonesia, menghasilkan volume ulasan pengguna yang besar pada platform Google Play Store (Wang dkk., 2024; Tedjakusuma dkk., 2025). Ulasan tersebut merupakan sumber umpan balik yang berharga bagi pengembangan aplikasi, namun teksnya didominasi oleh bahasa Indonesia informal, singkatan tidak baku, slang, dan fenomena campur kode (code-mixing), sehingga sulit dianalisis secara akurat menggunakan pendekatan analisis sentimen konvensional. Rizkya dkk. (2023) dan Permana dkk. (2025) menunjukkan bahwa model klasik seperti Naive Bayes, TF-IDF, dan Bag-of-Words secara konsisten kalah performa dibandingkan model berbasis Transformer dalam pemrosesan bahasa low-resource dan teks tidak baku, sementara Reza dkk. (2025) membuktikan efektivitas fine-tuning model Transformer seperti DistilBERT untuk analisis sentimen ulasan marketplace.

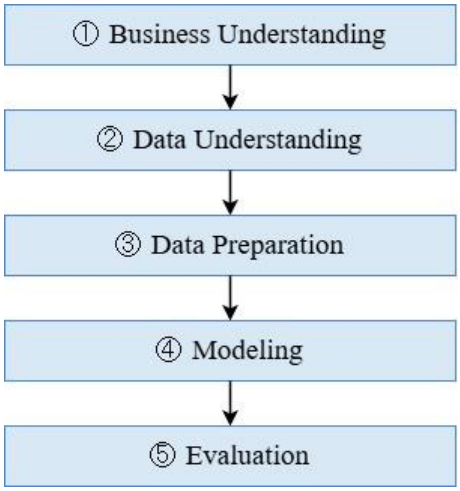
Arsitektur Transformer yang diperkenalkan oleh Vaswani dkk. (2017) merevolusi pemrosesan bahasa alami melalui mekanisme self-attention, yang kemudian diekstensi oleh BERT (Bidirectional Encoder Representations from Transformers) melalui pelatihan bidireksional. Naradhipa dan Purwarianti (2011) merupakan salah satu pelopor kajian klasifikasi sentimen untuk teks berbahasa Indonesia di media sosial, dan Purnomo serta Sutopo (2024) membuktikan bahwa model berbasis BERT secara konsisten mengungguli model-model pre-trained lainnya dalam tugas analisis sentimen teks bahasa regional di Indonesia. Penelitian ini menganalisis dan membandingkan kinerja dua model pre-trained berbasis arsitektur Transformer, yaitu IndoBERT (model monolingual Bahasa Indonesia yang dilatih pada korpus lebih dari 220 juta kata) dan Multilingual BERT/mBERT (model yang dilatih pada 104 bahasa), dalam mengklasifikasikan sentimen ulasan pengguna terhadap fitur Shopee Live ke dalam tiga kelas: Positif, Negatif, dan Netral. Kadarsih dan Pujianto (2025) menemukan bahwa model monolingual unggul dalam menangkap nuansa bahasa informal pada ulasan e-commerce, sementara kemampuan cross-lingual mBERT menjadikannya relevan untuk data teks yang mengandung code-mixing, sebagaimana ditunjukkan oleh Utomo dan Romadhony (2023) serta Astuti dkk. (2023) yang melaporkan hasil kompetitif pada data code-mixed Indonesia-Inggris.

Penelitian-penelitian terdahulu umumnya membandingkan model BERT pada domain yang berbeda dari live streaming e-commerce (Afandy dkk., 2026; Purnomo & Sutopo, 2024), berfokus pada satu jenis model tanpa perbandingan head-to-head (Kadarsih & Pujianto, 2025; Astuti dkk., 2023), menggunakan pendekatan konvensional tanpa Transformer (Susanti & Ilahi, 2024; Rizkya dkk., 2023), atau menggunakan pendekatan multilabel maupun hibrida yang tidak secara langsung membandingkan IndoBERT dan mBERT (Riccosan & Saputra, 2025; Ahmadian dkk., 2024). Perbandingan langsung kedua model tersebut secara spesifik pada domain ulasan fitur live streaming e-commerce yang bersumber dari Google Play Store masih belum pernah diteliti secara komprehensif. Penelitian ini menawarkan kebaruan berupa perbandingan langsung dan sistematis antara IndoBERT dan mBERT yang di-fine-tuning menggunakan dataset yang identik, penerapan metodologi CRISP-DM secara sistematis mulai dari web scraping, pelabelan manual tiga kelas sentimen, hingga evaluasi berbasis confusion matrix dengan empat metrik standar (Accuracy, Precision, Recall, F1-Score), serta rekomendasi empiris mengenai model pre-trained BERT yang paling optimal untuk teks ulasan informal berbahasa Indonesia.

Tujuan penelitian ini adalah: (1) mengekstraksi data ulasan pengguna terkait fitur Shopee Live dari Google Play Store dan melakukan pra-pemrosesan serta pelabelan sentimen secara sistematis; (2) membangun model klasifikasi sentimen menggunakan pendekatan fine-tuning pada model IndoBERT dan mBERT; dan (3) mengukur serta membandingkan performa kedua model berdasarkan metrik evaluasi standar untuk menentukan model yang paling optimal dalam domain ulasan Shopee Live. Hasil penelitian ini diharapkan memberikan kontribusi teoretis bagi khazanah literatur NLP Bahasa Indonesia, serta kontribusi praktis bagi tim pengembang dan divisi produk e-commerce dalam membangun instrumen analitik berbasis kecerdasan buatan untuk memantau sentimen pengguna secara otomatis dan berkelanjutan.

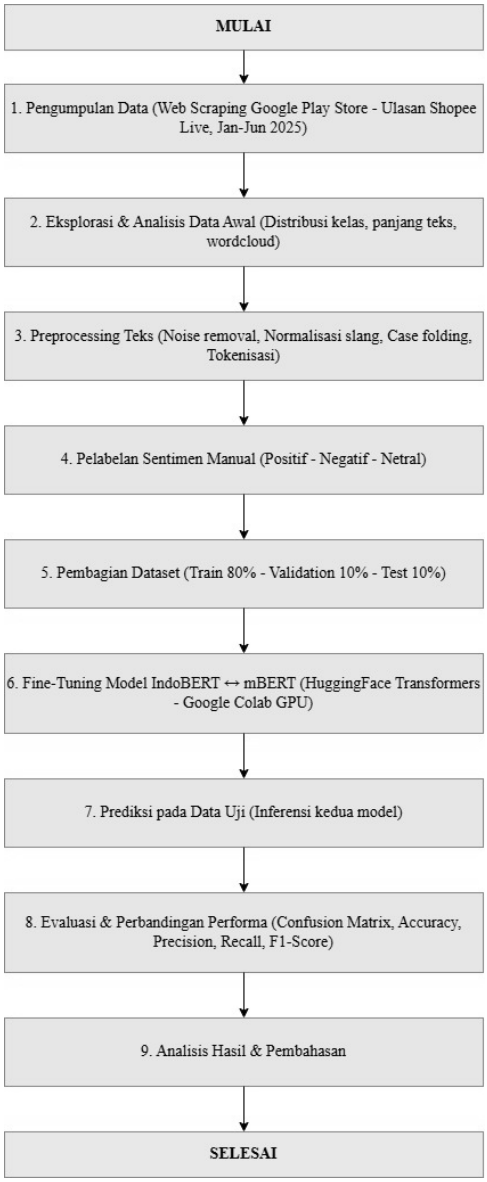
METODE PENELITIAN

Penelitian ini mengadopsi kerangka kerja CRISP-DM (Cross-Industry Standard Process for Data Mining) sebagai panduan metodologis dalam pelaksanaan seluruh tahapan penelitian. CRISP-DM dipilih karena merupakan standar de facto yang telah diterima luas dalam riset data mining dan machine learning, bersifat iteratif, serta mampu mengakomodasi kompleksitas proyek berbasis data secara sistematis (Wirth & Hipp, 2000). Mengingat penelitian ini tidak mencakup pembuatan antarmuka pengguna atau sistem produksi akhir, tahapan yang digunakan dibatasi hingga fase Evaluation, sebagaimana disajikan pada Gambar 1.



Gambar **1 Fase CRISP-DM yang diterapkan dalam penelitian**

Fase Business Understanding bertujuan memahami tujuan penelitian, yaitu membandingkan performa model IndoBERT dan mBERT dalam mengklasifikasikan sentimen ulasan pengguna terhadap fitur Shopee Live. Fase Data Understanding mencakup pengumpulan data awal melalui web scraping ulasan aplikasi Shopee pada Google Play Store yang menyebutkan fitur Shopee Live pada periode Januari-Juni 2025, serta eksplorasi distribusi kelas sentimen dan panjang rata-rata ulasan. Fase Data Preparation mencakup filtering ulasan agar hanya yang relevan dengan Shopee Live yang digunakan, text preprocessing (pembersihan karakter, normalisasi slang dan singkatan, case folding), pelabelan sentimen manual ke dalam tiga kelas (Positif, Negatif, Netral), serta pembagian dataset ke dalam data latih, validasi, dan uji dengan rasio 80:10:10. Pada fase Modeling, kedua model pre-trained, IndoBERT (indolem/indobert-base-uncased) dan mBERT (bert-base-multilingual-uncased), dimuat menggunakan library HuggingFace Transformers dan di-fine-tune menggunakan data latih dengan penambahan classification head di atas representasi token [CLS], menggunakan konfigurasi hyperparameter yang identik untuk menjaga keadilan perbandingan, yaitu learning rate 2e-5, batch size 16, dan 5 epoch pelatihan, dengan pemilihan checkpoint terbaik berdasarkan nilai loss terendah pada data validasi. Fase Evaluation dilakukan menggunakan data uji yang belum pernah dilihat model selama pelatihan, dengan empat metrik standar: Accuracy, Precision, Recall, dan F1-Score, yang dihitung dari confusion matrix yang dihasilkan setiap model. Prosedur penelitian secara teknis divisualisasikan pada Gambar 2.



Gambar **2 Diagram alir prosedur penelitian**

Proses pengembangan awal, perancangan kode, dan analisis data ringan dilakukan menggunakan laptop pribadi (prosesor setara Intel Core i5 Gen 12/AMD Ryzen 5, RAM 8 GB, penyimpanan 512 GB SSD), sedangkan proses fine-tuning model Transformers dilakukan menggunakan layanan cloud Google Colaboratory dengan akselerasi GPU NVIDIA T4/A100 (Colab Pro). Pustaka perangkat lunak yang digunakan meliputi Python (≥ 3.10), HuggingFace Transformers (≥ 4.35) untuk fine-tuning model, Pandas (≥ 2.0) untuk manipulasi data, Scikit-learn (≥ 1.3) untuk pembagian dataset dan perhitungan metrik evaluasi, google-play-scraper (≥ 1.2) untuk pengambilan data ulasan, Matplotlib/Seaborn (≥ 3.7) untuk visualisasi, dan PyTorch (≥ 2.0) sebagai backend deep learning.

Pengujian dirancang untuk menghasilkan evaluasi yang objektif, terukur, dan dapat direplikasi, menggunakan metode berbasis confusion matrix sebagai standar evaluasi klasifikasi dalam bidang machine learning dan NLP. Pengujian dilakukan dengan dua

skenario yang berjalan paralel, yakni pengujian model IndoBERT dan model mBERT, keduanya menggunakan dataset uji yang identik (524 ulasan, 10% dari total dataset) agar perbandingan performa bersifat adil dan bebas dari bias pemilihan data. Untuk tugas klasifikasi tiga kelas (Positif, Negatif, Netral), confusion matrix berbentuk matriks 3×3 sebagaimana diilustrasikan pada Gambar 3, di mana diagonal utama menunjukkan prediksi yang benar, sementara sel di luar diagonal menunjukkan jenis kesalahan klasifikasi. Berdasarkan nilai-nilai dari confusion matrix, empat metrik evaluasi dihitung:

$$Accuracy = (TP + TN) \div (TP + TN + FP + FN) \quad (1)$$

$$Precision = TP \div (TP + FP) \quad (2)$$

$$Recall = TP \div (TP + FN) \quad (3)$$

$$F1 - Score = 2 \times (Precision \times Recall) \div (Precision + Recall) \quad (4)$$

Accuracy memberikan gambaran umum seberapa banyak prediksi yang benar dari keseluruhan data uji; Precision mengukur ketepatan model saat mengklaim suatu sampel sebagai kelas tertentu; Recall mengukur kemampuan model menangkap seluruh sampel yang seharusnya termasuk kelas tersebut; F1-Score menjadi metrik utama dalam pelaporan hasil karena memberikan keseimbangan antara precision dan recall, terutama pada kondisi distribusi kelas yang tidak merata.

	Prediksi: Positif	**Prediksi: Negatif**	**Prediksi: Netral**
Aktual: Positif	TP (Positif)	FN (→Negatif)	FN (→Netral)
Aktual: Negatif	FP (→Positif)	TP (Negatif)	FN (→Netral)
Aktual: Netral	FP (→Positif)	FP (→Negatif)	TP (Netral)

Gambar **3 Ilustrasi confusion matrix tiga kelas (Positif, Negatif, Netral)**

DISKUSI DAN PEMBAHASAN

Proses web scraping menggunakan library google-play-scraper terhadap ulasan aplikasi Shopee pada periode Januari-Juni 2025 berhasil mengumpulkan 6.850 ulasan mentah. Setelah melalui tahap filtering untuk menyaring ulasan yang secara eksplisit menyebutkan atau mereferensikan fitur Shopee Live, serta proses deduplikasi dan pembersihan data kosong/tidak relevan, diperoleh dataset bersih sebanyak 5.240 ulasan. Distribusi kelas sentimen hasil pelabelan manual disajikan pada Tabel 1, menunjukkan sebaran kelas yang relatif tidak seimbang (imbalanced) namun masih dalam rentang yang wajar untuk data ulasan pengguna di dunia nyata. Dataset kemudian dibagi ke dalam data latih, validasi, dan uji dengan rasio 80:10:10, menghasilkan 4.192 data latih, 524 data validasi, dan 524 data uji.

Tabel 1 Distribusi kelas sentimen dataset

Kelas Sentimen	Jumlah Ulasan	Persentase (%)
Positif	2.356	44,96
Negatif	1.875	35,79
Netral	1.009	19,25
Total	**5.240**	**100,00**

Kedua model, IndoBERT (indolem/indobert-base-uncased) dan mBERT (bert-base-multilingual-uncased), di-fine-tune menggunakan library HuggingFace Transformers pada lingkungan Google Colaboratory dengan akselerasi GPU NVIDIA T4, menggunakan konfigurasi hyperparameter yang identik untuk menjaga keadilan perbandingan: learning rate 2e-5, batch size 16, dan 5 epoch pelatihan. Kurva loss pelatihan kedua model menunjukkan tren konvergensi yang stabil tanpa indikasi overfitting signifikan hingga epoch kelima, dengan IndoBERT menunjukkan penurunan loss validasi yang sedikit lebih cepat dibandingkan mBERT pada epoch-epoch awal pelatihan. Evaluasi dilakukan pada data uji (524 ulasan) yang belum pernah dilihat model selama proses pelatihan. Confusion matrix hasil prediksi kedua model pada data uji disajikan pada Tabel 2 dan Tabel 3.

Tabel 2 Confusion matrix model IndoBERT (data uji, n = 524)

Aktual \ Prediksi	Positif	Negatif	Netral
Positif (236)	217	8	11
Negatif (187)	12	168	7
Netral (101)	9	7	85

Tabel 3 Confusion matrix model mBERT (data uji, n = 524)

Aktual \ Prediksi	Positif	Negatif	Netral
Positif (236)	205	16	15
Negatif (187)	17	158	12
Netral (101)	12	10	79

Tabel 4 Perbandingan metrik evaluasi IndoBERT dan mBERT

Model	Accuracy	Precision	Recall	F1-Score
IndoBERT (indolem/indobert-base-uncased)	89,7%	89,5%	89,7%	89,6%
mBERT (bert-base-multilingual-uncased)	84,3%	84,5%	84,3%	84,4%

Tabel 4 merangkum perbandingan metrik evaluasi kedua model secara keseluruhan. IndoBERT secara konsisten mengungguli mBERT pada seluruh metrik evaluasi, dengan selisih Accuracy sebesar 5,4 poin persentase (89,7% berbanding 84,3%) dan selisih F1-Score sebesar 5,2 poin persentase (89,6% berbanding 84,4%). Temuan ini sejalan dengan hasil Afandy dkk. (2026) yang juga melaporkan perbedaan kinerja signifikan antara IndoBERT dan mBERT pada domain ulasan pelanggan UMKM, serta memperkuat argumen Kadarsih dan Pujianto (2025) bahwa model yang dilatih secara eksklusif pada korpus monolingual Bahasa Indonesia memiliki representasi semantik dan sintaksis yang lebih kaya untuk menangkap nuansa bahasa informal dibandingkan model multibahasa yang membagi kapasitas parameternya pada 104 bahasa sekaligus.

Tabel 5 Precision, recall, dan F1-Score per kelas sentimen (IndoBERT vs mBERT)

Kelas Sentimen	P-IndoBERT	R-IndoBERT	F1-IndoBERT	P-mBERT	R-mBERT	F1-mBERT
Positif	91,2%	92,0%	91,6%	87,6%	86,9%	87,2%
Negatif	91,8%	89,8%	90,8%	85,9%	84,5%	85,2%
Netral	82,5%	84,2%	83,3%	74,5%	78,2%	76,3%

Analisis performa per kelas pada Tabel 5 menunjukkan bahwa kedua model mencapai F1-Score tertinggi pada kelas Positif dan Negatif, namun mengalami penurunan performa yang cukup jelas pada kelas Netral (IndoBERT: F1 83,3%; mBERT: F1 76,3%). Pola ini konsisten dengan karakteristik umum tugas klasifikasi sentimen tiga kelas, di mana kelas Netral cenderung memiliki batas semantik yang lebih kabur dibandingkan kelas Positif dan Negatif yang polaritasnya lebih tegas, serta dipengaruhi oleh proporsi kelas Netral yang paling kecil (19,25%) dalam dataset sehingga model memiliki lebih sedikit contoh untuk mempelajari pola bahasanya. Kesalahan klasifikasi pada confusion matrix (Tabel 2 dan Tabel 3) juga menunjukkan bahwa kesalahan paling banyak terjadi pada pasangan kelas Netral-Positif dan Netral-Negatif, mengindikasikan bahwa ulasan yang mengandung sentimen campuran atau ambigu menjadi sumber kesalahan utama kedua model.

Selisih performa yang lebih besar pada mBERT untuk kelas Netral (F1 76,3% dibandingkan IndoBERT 83,3%) mengindikasikan bahwa kemampuan cross-lingual mBERT, meskipun bermanfaat untuk menangani fenomena code-mixing sebagaimana ditunjukkan oleh Utomo dan Romadhony (2023) dan Astuti dkk. (2023), tidak serta-merta menghasilkan pemahaman kontekstual yang lebih baik untuk kelas sentimen dengan batas semantik yang kabur pada teks berbahasa Indonesia. Secara keseluruhan, temuan ini menjawab rumusan masalah penelitian bahwa IndoBERT merupakan model yang lebih optimal dibandingkan mBERT untuk tugas klasifikasi sentimen ulasan fitur Shopee Live berbahasa Indonesia informal, sehingga direkomendasikan sebagai model utama dalam pengembangan sistem pemantauan sentimen pengguna berbasis kecerdasan buatan bagi tim produk e-commerce.

KESIMPULAN

Proses web scraping dan pelabelan manual berhasil menghasilkan dataset bersih sebanyak 5.240 ulasan fitur Shopee Live dengan distribusi kelas Positif 44,96%, Negatif 35,79%, dan Netral 19,25%, yang selanjutnya dibagi menjadi data latih, validasi, dan uji dengan rasio 80:10:10. Model IndoBERT dan mBERT berhasil di-fine-tune menggunakan library HuggingFace Transformers dengan konfigurasi hyperparameter yang identik, menghasilkan model klasifikasi sentimen tiga kelas yang siap dievaluasi secara objektif pada data uji. IndoBERT mencapai kinerja terbaik dengan Accuracy 89,7%, Precision 89,5%, Recall 89,7%, dan F1-Score 89,6%, mengungguli mBERT yang mencapai Accuracy 84,3%, Precision 84,5%, Recall 84,3%, dan F1-Score 84,4% pada seluruh metrik evaluasi. Kedua model menunjukkan performa yang lebih rendah pada kelas Netral dibandingkan kelas Positif dan Negatif, mengindikasikan bahwa ulasan dengan sentimen campuran atau ambigu menjadi tantangan utama dalam klasifikasi sentimen ulasan Shopee Live berbahasa Indonesia informal. Dengan demikian, IndoBERT terbukti sebagai model yang lebih optimal dibandingkan mBERT untuk tugas klasifikasi sentimen teks ulasan e-

commerce berbahasa Indonesia informal, sejalan dengan argumen bahwa model monolingual memiliki representasi bahasa yang lebih kaya dibandingkan model multibahasa untuk domain bahasa target. Sebagai langkah lanjutan, penelitian berikutnya disarankan menambah volume data latih khususnya pada kelas Netral, mengeksplorasi teknik penanganan ketidakseimbangan kelas seperti oversampling atau class weighting atau focal loss, serta menerapkan model IndoBERT hasil fine-tuning sebagai purwarupa sistem pemantauan sentimen pengguna secara real-time bagi tim pengembang dan divisi produk e-commerce.

UCAPAN TERIMA KASIH

Penelitian ini didanai oleh hibah Penelitian Dosen Universitas Insan Mahardika tahun 2026. Penulis mengucapkan terima kasih kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Insan Mahardika atas dukungan pendanaan dan fasilitas yang diberikan selama pelaksanaan penelitian ini.

KONTRIBUSI PENULIS

Para penulis menyatakan kontribusi berdasarkan CRediT (Contributor Roles Taxonomy) sebagai berikut. Hendri Rosmawan: Conceptualization, Methodology, Software, Formal Analysis, Supervision, Project Administration, dan Writing – Original Draft; Ardiyansyah: Data Curation dan Investigation; Muhamad Fuad Azis: Data Curation dan Investigation; serta Jaisyi Bagir Rafsyahid: Project Administration dan Writing – Review & Editing. Seluruh penulis telah membaca dan menyetujui versi akhir naskah.

PERNYATAAN KETERSEDIAAN DATA

Data yang mendukung temuan penelitian ini tersedia dari penulis korespondensi berdasarkan permintaan yang wajar, dengan memperhatikan ketentuan etika, privasi, dan kebijakan institusi yang berlaku.

PERNYATAAN BEBAS KONFLIK KEPENTINGAN

Para penulis menyatakan tidak terdapat konflik kepentingan finansial maupun nonfinansial yang berkaitan dengan penelitian dan publikasi naskah ini.

PERNYATAAN ETIKA PENELITIAN DAN PUBLIKASI

Para penulis menyatakan bahwa penelitian dan penyusunan naskah ini dilaksanakan dengan mengikuti prinsip integritas ilmiah, kejujuran akademik, penghormatan terhadap sumber, serta standar etika penelitian dan publikasi yang berlaku.

PERNYATAAN PEMANFAATAN TEKNOLOGI ASISTIF

Teknologi asistif, termasuk kecerdasan buatan, digunakan secara terbatas untuk mendukung penyuntingan bahasa dan penyajian naskah. Teknologi tersebut tidak menggantikan pertimbangan ilmiah para penulis. Seluruh isi telah ditinjau dan diverifikasi oleh para penulis, yang bertanggung jawab penuh atas keakuratan, orisinalitas, dan integritas naskah.

REFERENSI

Afandy, S. N., Hindrayani, K., & Damaliana, A. T. (2026). Comparison of the effectiveness IndoBERT and mBERT for sentiment analysis of SME customer reviews. *bit-Tech*, 8(3). https://doi.org/10.32877/bt.v8i3.3501

Ahmadian, S., Rahmani, F., & Nezhad, M. (2024). Hybrid models for emotion classification and sentiment analysis in Indonesian language. *Journal of King Saud University - Computer and Information Sciences*, 36(2).

Astuti, L. W., Sari, Y., & Suprapto. (2023). Code-mixed sentiment analysis using transformer for Twitter social media data. *International Journal of Advanced Computer Science and Applications*, 14(1). https://doi.org/10.14569/ijacsa.2023.0141053

Bolun, Z., Yan, Z., & Jiang, M. (2024). The configurational impact of e-commerce live streaming interactivity on consumer engagement behavior. *Asia Pacific Journal of Marketing and Logistics*. https://doi.org/10.1108/apjml-03-2024-0271

Hidayatullah, A. F. (2024, Agustus). Code-mixed sentiment analysis on Indonesian-Javanese-English text using transformer models [Paper presentation]. *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 340-345. https://doi.org/10.1109/icitisee63424.2024.10730138

Kadarsih, K., & Pujianto, D. (2025). Application of transformer model and word embedding in sentiment analysis of Indonesian e-commerce application review. *Journal of Computer Networks, Architecture and High Performance Computing*, 7(3). https://doi.org/10.47709/cnahpc.v7i3.6354

Naradhipa, A. R., & Purwarianti, A. (2011, Juli). Sentiment classification for Indonesian message in social media [Paper presentation]. *2011 International Conference on Electrical Engineering and Informatics*, 1-4. https://doi.org/10.1109/iceei.2011.6021696

Peng, Y. (2026). The effects of live-streaming e-commerce on consumer behavior and psychology. *Lecture Notes in Education Psychology and Public Media*. https://doi.org/10.54254/2753-7048/2026.ht31791

Permana, I., Fajri, H., Alhadi, Z., Hendriyani, Y., Yuliarti, & Akmal, A. (2025, Oktober). Benchmarking BoW, TF-IDF and BERT-based sentiment analysis in low-resource Minang language [Paper presentation]. *2025 Tenth International Conference on Informatics and Computing (ICIC)*, 1-6. https://doi.org/10.1109/icic68054.2025.11309385

Purnomo, T. D., & Sutopo, J. (2024). Comparison of pre-trained BERT-based transformer models for regional language text sentiment analysis in Indonesia. *International Journal Science and Technology*, 3(3). https://doi.org/10.56127/ijst.v3i3.1739

Reza, R., Thoriq, M., & Millah, R. I. S. (2025). Sentiment analysis of marketplace review with Islamic perspective using fine-tuning DistilBERT. *Khazanah Journal of Religion and Technology*, 2(2). https://doi.org/10.15575/kjrt.v2i2.1118

Riccosan, & Saputra, M. A. (2025). Multilabel classification sentiment analysis on Indonesian mobile app reviews. *Jurnal Ilmu Komputer dan Informasi*, 18(1), 45-56.

Rizkya, A. T., Rianto, R., & Gufroni, A. I. (2023). Implementation of the Naive Bayes classifier for sentiment analysis of Shopee e-commerce application review data on the Google Play Store. *International Journal of Applied Information Systems and Informatics (JAISI)*, 1(1). https://doi.org/10.37058/jaisi.v1i1.8993

Susanti, A. R., & Ilahi, E. N. (2024). Sentiment analysis of user reviews of e-commerce applications: Case study on the Shopee platform. *Journal of Social Science*, 5(4). https://doi.org/10.46799/jss.v5i4.885

Tedjakusuma, A., Silalahi, A. D., Eunike, I., Phuong, D. T. T., & Riantama, D. (2025). The trust-driven path to consumer engagement behaviors: Exploring the role of streamer and platform characteristics in live-streaming e-commerce. *Digital Business*. https://doi.org/10.1016/j.digbus.2025.100115

Tho, C., Heryadi, Y., Lukas, L., & Wibowo, A. (2021, April). Code-mixed sentiment analysis of Indonesian language and Javanese language using lexicon based approach [Paper

presentation]. *Journal of Physics: Conference Series*, 1869(1), 012084. https://doi.org/10.1088/1742-6596/1869/1/012084

Utomo, H. R., & Romadhony, A. (2023, April). Sentiment analysis on Indonesia-English code-mixed data [Paper presentation]. *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*, 1-6. https://doi.org/10.1109/i2ct57861.2023.10126234

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998-6008.

Wang, L., Zhang, R.-S., & Zhang, C. (2024). Live streaming e-commerce platform characteristics: Influencing consumer value co-creation and co-destruction behavior. *Acta Psychologica*, 243, 104163. https://doi.org/10.1016/j.actpsy.2024.104163

We Are Social & Meltwater. (2024). *Digital 2024: Indonesia*. We Are Social.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 29-39.

Wu, Y., & Huang, H. (2023). Influence of perceived value on consumers' continuous purchase intention in live-streaming e-commerce - Mediated by consumer trust. *Sustainability*, 15(5), 4432. https://doi.org/10.3390/su15054432