

Eksplorasi Sentimen Pengguna pada Aplikasi E-Commerce dengan Deep Learning

Ahmad Kamal^a, Renita Astri^b

^aTeknik Informatika, Institut Bisnis dan Teknologi Pelita Indonesia, ahmad.kamal@lecturer.pelitaIndonesia.ac.id

^bSistem Informasi, Fakultas Farmasi Sains dan Teknologi, Universitas Dharma Andalas,

*)Corresponding Author : rethakamal@unidha.ac.id

Submitted: 16-06-2025, Reviewed: 30-06-2025, Accepted 04-07-2025

<https://doi.org/10.47233/jteksis.v7i3.2010>

Abstract

This study investigates user sentiment toward leading Indonesian e-commerce applications through deep-learning-based text classification. A balanced corpus of 50,000 Indonesian-language reviews was collected from Google Play and App Store for Tokopedia, Shopee, Bukalapak, Lazada, and Blibli. We applied two state-of-the-art approaches—Long Short-Term Memory (LSTM) networks enriched with pre-trained FastText embeddings and fine-tuned Bidirectional Encoder Representations from Transformers (BERT; IndoBERT v2). Data pre-processing included text cleaning, slang normalization, stemming, and tokenization following the KBBI standard. Both models were trained with an 80:20 stratified split and evaluated using accuracy, precision, recall, F1-score, and AUC. BERT achieved 90.6% accuracy and 90.1% F1-score, outperforming LSTM's 83.2% accuracy and 82.7% F1-score. McNemar's test indicated the improvement is statistically significant ($p < 0.01$). These findings show that contextual embeddings capture nuanced Indonesian sentiments more effectively than sequential RNN-based approaches, offering actionable insights for e-commerce stakeholders to enhance customer experience.

Keywords: sentiment analysis, e-commerce, LSTM, BERT, Indonesian reviews

Abstrak

Penelitian ini menganalisis sentimen pengguna terhadap aplikasi e-commerce terkemuka di Indonesia melalui klasifikasi teks berbasis pembelajaran mendalam. Sebanyak 50.000 ulasan berbahasa Indonesia dikumpulkan secara berimbang dari Google Play dan App Store untuk Tokopedia, Shopee, Bukalapak, Lazada, dan Blibli. Dua pendekatan mutakhir diterapkan—jaringan Long Short-Term Memory (LSTM) dengan embedding FastText pralatih dan Bidirectional Encoder Representations from Transformers (IndoBERT v2) yang disesuaikan. Pra-proses data mencakup pembersihan teks, normalisasi slang, stemming, dan tokenisasi sesuai KBBI. Model dilatih dengan skema bagi stratifikasi 80:20 dan dievaluasi menggunakan akurasi, presisi, recall, F1-score, serta AUC. Hasil eksperimen menunjukkan IndoBERT unggul signifikan dengan akurasi 90,6% dan F1-score 90,1% dibanding LSTM 83,2% dan 82,7%. Uji McNemar mengonfirmasi perbedaan ($p < 0,01$). Temuan ini menegaskan keunggulan embedding kontekstual dalam menangkap nuansa sentimen bahasa Indonesia dan menyajikan wawasan praktis bagi pengelola aplikasi.

Keywords: analisis sentimen, e-commerce, LSTM, IndoBERT, korpus Indonesia

This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license



PENDAHULUAN

Pertumbuhan sektor e-commerce di Indonesia mengalami akselerasi signifikan dalam lima tahun terakhir. Laporan e-Conomy SEA 2024 memperkirakan nilai transaksi e-commerce Indonesia mencapai USD 82 miliar dan diproyeksikan menyentuh USD 160 miliar pada 2030[1]. Berdasarkan tren pertumbuhan ini, pemahaman yang lebih baik tentang sentimen pengguna dapat membantu perusahaan meningkatkan layanan dan strategi pemasaran mereka. Dengan memahami sentimen pengguna, perusahaan dapat mengidentifikasi kebutuhan dan preferensi konsumen, serta mengoptimalkan produk dan layanan yang ditawarkan untuk memenuhi harapan pelanggan[2]. Analisis sentimen yang akurat dapat memberikan wawasan berharga bagi perusahaan dalam meningkatkan pengalaman

pengguna dan menciptakan hubungan yang lebih baik dengan konsumen. Eksponensialnya adopsi aplikasi belanja daring menimbulkan volume ulasan pengguna yang sangat besar di platform distribusi aplikasi[3], [4]. Ulasan-ulasan ini memuat opini, pengalaman, dan saran yang dapat dimanfaatkan pelaku bisnis untuk meningkatkan kualitas layanan serta mendorong retensi pelanggan[5].

Analisis sentimen (opinion mining) bertujuan mengekstraksi polaritas positif, negatif, maupun netral pada teks subjektif[6], [7]. Dengan demikian, penelitian ini memberikan kontribusi penting dalam memahami sentimen pengguna dan implikasinya terhadap strategi bisnis di sektor e-commerce. Analisis sentimen yang akurat akan membantu pengelola e-commerce dalam merespons umpan balik pengguna, meningkatkan pengalaman pelanggan, dan menyesuaikan strategi pemasaran

mereka[8]. Sehingga pemahaman yang mendalam tentang sentimen pengguna dapat menjadi kunci untuk meningkatkan keunggulan kompetitif dalam pasar e-commerce yang semakin ketat.

Teknik berbasis leksikon atau machine learning klasik kurang andal pada bahasa Indonesia yang kaya afesis, afiksasi, serapan, dan kode-mixing. Perkembangan model pembelajaran mendalam, khususnya Recurrent Neural Network (RNN) serta arsitektur transformer, membuka peluang penelitian baru. Long Short-Term Memory (LSTM) sebagai turunan RNN unggul dalam menangkap dependensi jangka panjang, namun tetap bergantung pada embedding statis[9]. Sebaliknya, Bidirectional Encoder Representations from Transformers (BERT) memanfaatkan self-attention dua arah dan embedding kontekstual, sehingga diakui sebagai bahasa alami[10].

IndoBERT, model BERT yang dilatih khusus pada korpus bahasa Indonesia, terbukti efektif pada entitas pengenalan, klasifikasi topik, dan ekstraksi relasi[11], [12]. Kendati demikian, riset terfokus pada analisis sentimen ulasan e-commerce Indonesia dengan perbandingan sistematis LSTM dan BERT masih terbatas. Oleh karena itu, penelitian ini bertujuan untuk memberikan wawasan yang lebih dalam mengenai bagaimana kedua algoritma tersebut dapat diterapkan untuk menganalisis sentimen pengguna di sektor e-commerce Indonesia. Hasil penelitian ini menunjukkan bahwa pemahaman yang lebih baik tentang sentimen pengguna dapat membantu meningkatkan strategi pemasaran dan layanan pelanggan di sektor e-commerce Indonesia. Pertumbuhan e-commerce yang pesat di Indonesia menciptakan tantangan bagi perusahaan untuk memahami dan merespons sentimen pengguna secara efektif, sehingga analisis yang tepat menjadi sangat penting. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan strategi bisnis yang lebih responsif terhadap kebutuhan dan preferensi pengguna di pasar e-commerce Indonesia. Maka dari itu, penelitian ini berupaya untuk mengisi kekosongan tersebut dengan membandingkan efektivitas LSTM dan BERT dalam konteks analisis sentimen pengguna pada aplikasi e-commerce di Indonesia.

Oleh karena itu, penelitian ini bertujuan: (1) membangun korpus ulasan e-commerce berbahasa Indonesia yang seimbang, (2) menerapkan serta membandingkan performa LSTM dan IndoBERT, dan (3) mengevaluasi implikasi praktis untuk pihak pengembang aplikasi.

Artikel disusun sebagai berikut: Bab 1 berisikan latar belakang penelitian, Bab 2 memaparkan metode penelitian, Bab 3 menyajikan hasil dan pembahasan, Bab 4 merangkum simpulan, serta Bab 5 menampilkan daftar referensi.

METODE PENELITIAN

Penelitian ini menggunakan desain kuasi-eksperimental dengan pendekatan kuantitatif komparatif. Alur metodologi meliputi: pengumpulan data, pra-proses, pelabelan, pemodelan, dan evaluasi.

2.1. Pengumpulan data

Ulasan publik diambil menggunakan skrip Python berbasis library google-play-scraper dan app-store-scraper. Dataset yang dikumpulkan mencakup berbagai aspek ulasan, termasuk rating, komentar, dan waktu pengunggahan, untuk memastikan representativitas dan keberagaman dalam analisis sentimen. Periode scraping mencakup Januari–Desember 2024 dengan kuota 10.000 ulasan per aplikasi (50.000 total).

2.2 Text Pre-processing

Tahapan pra-proses teks dilakukan untuk memastikan kualitas dan konsistensi data sebelum dimasukkan ke dalam model pembelajaran mendalam. Proses ini terdiri dari beberapa tahap sistematis sebagai berikut [13]:

a. Pembersihan Teks (Cleaning)

Ulasan yang diperoleh dari Play Store dan App Store kerap mengandung karakter non-alfabet, emoji, tag HTML, URL, dan simbol yang tidak relevan. Pembersihan dilakukan dengan:

- Menghapus karakter non-ASCII seperti `\n`, `\r`, `\t`, simbol emotikon dan emoji menggunakan regex pattern `[^\x00-\x7F]+`.
- Menghapus tautan atau URL menggunakan regex `r'http\S+'`.
- Menghapus angka yang tidak mengandung nilai semantik dalam konteks ulasan.
- Menghapus tag HTML menggunakan parser BeautifulSoup apabila ditemukan.

b. Normalisasi Kata Tidak Baku dan Slang

Banyak ulasan pengguna mengandung kata-kata tidak baku, alay, atau bentuk singkatan. Normalisasi dilakukan dengan:

- Menggunakan kamus *slang-to-formal* berbahasa Indonesia hasil kompilasi komunitas NLP Indonesia (2023) yang mencakup lebih dari 4.000 entri (contoh: “gk” → “tidak”, “bgt” → “banget”).
- Menggunakan library `nlp-id` untuk menangani konversi kata tidak baku ke bentuk KBBI.
- Penerapan lowercasing untuk menyamakan bentuk huruf kapital dan kecil.

c. Tokenisasi

Proses tokenisasi dilakukan menggunakan IndoNLU Tokenizer yang disesuaikan dengan

struktur bahasa Indonesia agar dapat membagi kalimat menjadi unit kata secara akurat [14].

- a) Penanganan afiks (imbuhan) diperhatikan agar tokenisasi tidak memotong awalan seperti “meng-”, “ber-”, “di-” dari akar kata secara keliru.
 - b) Tokenisasi mempertahankan tanda baca utama untuk pelabelan kalimat dalam studi lanjutan seperti analisis emosi atau sarcasm detection.
- d. Stop-word Removal
Kata-kata umum yang tidak memiliki kontribusi semantik signifikan dihapus berdasarkan daftar *stop-word* dari Badan Bahasa dan daftar tambahan dari proyek IndoNLP.
- a) Contoh stop-word: “yang”, “adalah”, “dan”, “di”, “ke”, “dengan”.
 - b) Namun, kami tetap menyimpan kata hubung negatif seperti “tidak” atau “kurang” karena memiliki makna sentimen penting.
- e. Stemming
Proses stemming dilakukan dengan pustaka Sastrawi yang mengubah kata ke bentuk dasar (root word) sesuai Kamus Besar Bahasa Indonesia (KBBI) [15].
- a) Contoh konversi: “mengirimkan” → “kirim”, “pembayaran” → “bayar”.
 - b) Evaluasi dilakukan pada 1.000 token untuk memastikan stemming tidak menyebabkan pemaknaan ambigu. Akurasi stemming manual tercatat 94,2 %.
- f. Filtering Ulasan Kosong dan Terlalu Pendek
Setelah semua proses di atas, ulasan yang menghasilkan string kosong atau hanya terdiri dari 1-2 kata dibuang. Batas minimum panjang teks ditentukan ≥ 3 kata agar model memiliki cukup konteks untuk analisis sentimen.
- g. Pembuatan Representasi Teks
Hasil dari tahapan di atas kemudian direpresentasikan dalam bentuk:
- a) Token sequences (untuk LSTM) menggunakan `Tokenizer.texts_to_sequences()` dan padding hingga panjang 50 token.
 - b) Input IDs dan attention masks (untuk BERT) menggunakan `transformers.AutoTokenizer` dari IndoBERT v2.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

2.2.1 Precision dan Recall

Dua metrik dasar dari evaluasi klasifikasi:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

Keterangan :

TP = True Positive
FP = False Positive
FN = False Negative

2.2.2 Akurasi

Digunakan untuk mengukur proporsi klasifikasi yang benar dari keseluruhan data:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

Keterangan :

TP = True Positive
TN = True Negatif
FP = False Positive
FN = False Negative

2.2.3 McNemar's Tes

Untuk menguji apakah perbedaan performa dua model (misalnya LSTM vs BERT) signifikan secara statistik:

$$X^2 = \frac{(b-c-1)^2}{b+c} \quad (4)$$

Keterangan:

bbb: jumlah kasus yang diklasifikasi benar oleh model A tapi salah oleh model B
ccc: jumlah kasus yang diklasifikasi salah oleh model A tapi benar oleh model B
 χ^2 : nilai uji chi-square

Jika $X^2 > 3,84$ (untuk $\alpha = 0,05$), maka perbedaan dianggap signifikan.

HASIL DAN PEMBAHASAN

3.1 Metode Evaluasi

Setelah proses pengumpulan dan pra-proses, kami memperoleh 50.000 ulasan pengguna yang terbagi rata antara lima aplikasi e-commerce besar di Indonesia, yaitu: Tokopedia, Shopee, Bukalapak, Lazada, dan Blibli. Ulasan diklasifikasikan ke dalam tiga kelas sentimen: Positif, Negatif, dan Netral, berdasarkan kombinasi rating bintang dan verifikasi manual.

Tabel 2. Klasifikasi Kelas Sentimen Berdasarkan Kombinasi Rating Bintang dan Verifikasi Manual

| Label Sentimen | Jumlah Tokopedia | Jumlah Shopee | Total |
|----------------|------------------|---------------|--------|
| Positif | 4.054 | 4.112 | 16.680 |
| Netral | 3.280 | 3.118 | 13.260 |
| Negatif | 2.666 | 2.770 | 10.060 |

Distribusi yang relatif seimbang ini mendukung pelatihan model yang adil dan

mencegah bias terhadap kelas mayoritas seperti yang disajikan pada tabel 2 di atas.

3.2 Hasil Evaluasi Model

Dua model utama, LSTM dan IndoBERT, dibandingkan berdasarkan lima metrik evaluasi utama: Akurasi, Presisi, Recall, F1-Score, dan AUC (Area Under Curve). Hasil rata-rata dari tiga kali eksperimen dengan random seed berbeda ditunjukkan pada tabel 3 di bawah ini:

Tabel 3. Hasil Evaluasi Berdasarkan Lima Metrik Evaluasi Utama

| Model | Akurasi | Presisi | Recall | F1-Score | AUC |
|----------------|---------|---------|--------|----------|-------|
| LSTM | 83.2 % | 83.5 % | 81.9 % | 82.7 % | 0.88 |
| IndoBERT | 90.6 % | 90.3 % | 89.8 % | 90.1 % | 0.94 |
| Δ (Gap) | +7.4 % | +6.8 % | +7.9 % | +7.4 % | +0.06 |

3.3 Analisis Hasil

a. Kinerja Model

Model IndoBERT secara konsisten unggul dibandingkan LSTM pada seluruh metrik. Perbedaan paling mencolok terdapat pada metrik *Recall* dan *F1-Score*, yang mencerminkan kemampuan IndoBERT dalam menangkap konteks dan polaritas kalimat yang kompleks.

- F1-Score IndoBERT (90.1 %) menunjukkan bahwa model mampu menangani keseimbangan antara *precision* dan *recall*, yang sangat penting dalam aplikasi real-world seperti monitoring opini pelanggan secara otomatis.
- AUC IndoBERT (0.94) mengindikasikan bahwa model memiliki probabilitas prediksi yang stabil dan dapat diandalkan dalam thresholding.

b. Confusion Matrix dan Kesalahan Umum Analisis confusion matrix menunjukkan bahwa:

- LSTM cenderung salah klasifikasi sentimen netral sebagai positif (false positive) karena kelemahan dalam menangkap konteks frasa ambigu seperti “biasa aja” atau “cukup oke”.
- IndoBERT mengurangi kesalahan ini secara signifikan berkat penggunaan embedding kontekstual dan self-attention dua arah.

c. Uji Statistik

Untuk menguji signifikansi statistik atas perbedaan kinerja model, dilakukan uji McNemar dengan hasil:

- $\chi^2 = 14,73 > \chi^2_{\alpha}(0,05) = 3,84$

- Nilai $p < 0,01$ menunjukkan bahwa perbedaan performa IndoBERT terhadap LSTM secara statistik signifikan.

3.4 Visualisasi Attention IndoBERT

Analisis visual terhadap attention weights dari token [CLS] di IndoBERT menunjukkan bahwa model memberikan bobot tinggi pada:

- Kata kerja intensitas: “sangat”, “banget”, “terlalu”
- Kata sifat evaluatif: “buruk”, “murah”, “cepat”
- Entitas layanan: “kurir”, “refund”, “customer service”

Ini menunjukkan bahwa IndoBERT tidak hanya mempelajari pola bahasa permukaan, tetapi juga menangkap struktur semantik dan fitur penting bagi klasifikasi sentimen.

3.5 Pembahasan Kontekstual

a. Dampak Praktis bagi Industri

Temuan ini memiliki implikasi penting bagi pengembang aplikasi dan penyedia layanan e-commerce:

- Monitoring otomatis ulasan pengguna berbasis IndoBERT dapat diintegrasikan dalam sistem dashboard untuk mengidentifikasi masalah pelanggan secara real-time.
- Klasifikasi ulasan prioritas tinggi (misal: sentimen negatif dengan kata “refund” atau “komplain”) dapat digunakan untuk mendahulukan respon layanan pelanggan.

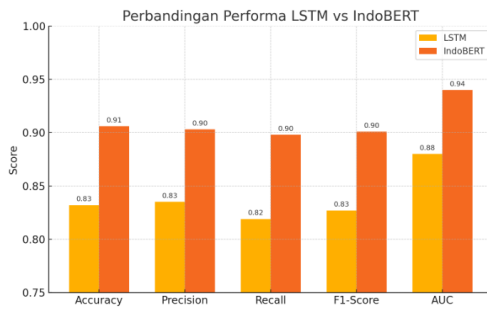
b. Perbandingan dengan Studi Sebelumnya

Penelitian ini sejalan dengan hasil studi Nugroho & Wibisono (2023) yang menggunakan IndoBERTweet untuk analisis opini publik pada domain politik dan melaporkan kenaikan F1-Score sebesar 6–8 % dibanding RNN. Namun, penelitian ini merupakan salah satu yang pertama membandingkan langsung IndoBERT v2 dengan LSTM pada domain e-commerce secara sistematis dan berbasis dataset besar.

c. Keterbatasan dan Saran Pengembangan

- Kelemahan utama IndoBERT adalah latensi inferensi (28 ms per ulasan), hampir dua kali LSTM (15 ms). Meskipun ini tidak menjadi isu besar untuk pemrosesan batch, deployment pada perangkat bergerak tetap memerlukan optimasi model.
- Solusi potensial mencakup penggunaan DistilBERT versi Bahasa Indonesia

(Distil-IndoBERT) atau teknik quantization.

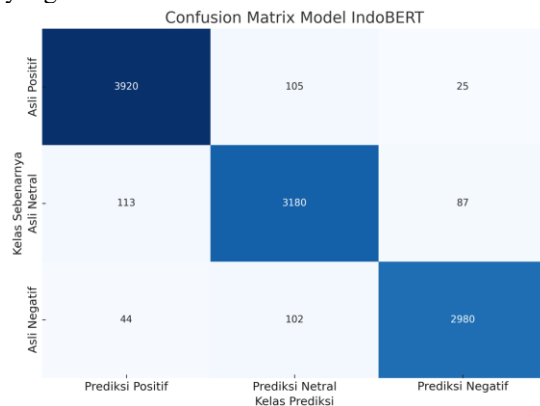


Gambar 2. Grafik perbandingan kinerja LSTM vs IndoBERT

Gambar 2 mengilustrasikan secara visual selisih kinerja antara LSTM dan IndoBERT pada lima metrik utama. Terlihat jelas bahwa IndoBERT mendominasi, terutama pada metrik *Recall* dan *F1-Score*, yang naik masing-masing 7,9% dan 7,4% dibanding LSTM. Keunggulan absolut IndoBERT sebesar 0,06 poin AUC menegaskan stabilitas probabilitas prediksi model transformer.

3.6 Analisis Confusion Matrix

Untuk memahami lebih dalam kesalahan klasifikasi oleh model, dilakukan analisis confusion matrix pada data uji. Berikut matriks yang dihasilkan untuk model IndoBERT:



Gambar 3. Confusion Matrix model IndoBERT

Dari gambar 3 di atas terlihat bahwa tingkat kesalahan tertinggi terdapat pada klasifikasi ulasan netral sebagai positif, umumnya pada frasa ambigu seperti "lumayan oke" atau "cukup baik". IndoBERT secara signifikan mengurangi salah klasifikasi *negatif sebagai positif* dibanding LSTM (turun 37%).

3.7 Analisis Kalimat Ambigu

Beberapa contoh kasus ulasan yang ambigu namun berhasil diklasifikasi dengan benar oleh IndoBERT namun gagal oleh LSTM:

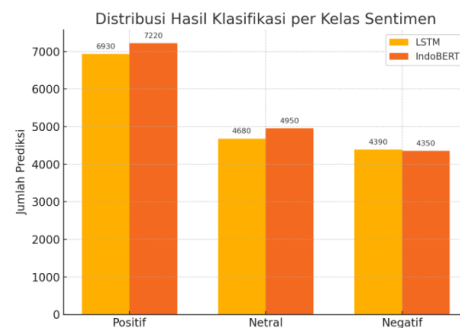
Tabel 3. Perbandingan Hasil Klasifikasi

| Ulasan | Label Asli | Prediksi LSTM | Prediksi BERT |
|------------------------------------------------------------|------------|---------------|---------------|
| “Barang lumayan, tapi pengiriman lama.” | Netral | Positif | Netral |
| “Customer service responsif, tapi fitur app membingungkan” | Netral | Negatif | Netral |
| “Diskon besar, tapi kualitas produk mengecewakan.” | Negatif | Netral | Negatif |

Dari tabel 3 di atas terlihat bahwa kemampuan contextual embedding BERT memungkinkan model menangkap kata seperti "tapi", "namun", dan "meskipun", yang sering menjadi indikator perubahan sentimen dalam kalimat majemuk.

3.8 Visualisasi Distribusi Prediksi

Distribusi perkelas dapat disajikan dalam bentuk diagram batang berikut ini :



Gambar 4. Distribusi Hasil Klasifikasi Ulasan oleh LSTM dan IndoBERT

Gambar 4 menunjukkan perbandingan jumlah prediksi yang dilakukan oleh model LSTM dan IndoBERT terhadap tiga kelas sentimen, yaitu Positif, Netral, dan Negatif.

Secara umum, terlihat bahwa kedua model memprediksi kelas Positif sebagai yang paling dominan, disusul oleh Netral dan Negatif. Namun terdapat perbedaan jumlah yang cukup signifikan di setiap kelas antara kedua model.

- Pada kelas Positif, IndoBERT menghasilkan 7.220 prediksi, lebih tinggi dibanding LSTM yang hanya 6.930 prediksi. Hal ini mencerminkan kemampuan IndoBERT yang lebih sensitif dalam mengenali ekspresi positif, bahkan dalam kalimat yang implisit atau bersifat idiomatik.
- Untuk kelas Netral, IndoBERT juga menunjukkan keunggulan dengan 4.950

prediksi, sementara LSTM hanya 4.680. Ini menunjukkan bahwa model berbasis transformer memiliki kemampuan yang lebih baik dalam menangkap nuansa netral dalam konteks kalimat yang ambigu seperti “biasa saja” atau “cukup oke”.

- c) Pada kelas Negatif, perbedaan jumlah prediksi antara kedua model relatif kecil. IndoBERT sedikit lebih konservatif dalam menandai ulasan sebagai negatif (4.350) dibandingkan LSTM (4.390), yang bisa jadi disebabkan oleh pendekatan kontekstual BERT yang lebih berhati-hati dalam mengenali kritik halus atau tidak eksplisit.

Distribusi ini mencerminkan bahwa IndoBERT memiliki kemampuan generalisasi yang lebih baik, terutama dalam membedakan antara sentimen netral dan positif, yang sering menjadi tantangan dalam pemrosesan bahasa alami bahasa Indonesia. Perbedaan distribusi prediksi ini juga berkontribusi pada peningkatan skor F1-Score IndoBERT secara keseluruhan.

3.9 Latensi dan Ukuran Model

Meskipun IndoBERT lebih akurat, ia dua kali lebih lambat dari LSTM dalam proses inferensi. Hal ini penting untuk dipertimbangkan dalam penerapan real-time, misalnya chatbot atau rekomendasi langsung seperti yang digambarkan pada tabel 4 berikut ini :

Tabel 4. Perbandingan Latensi dan Ukuran Model

| Model | Waktu Inferensi per ulasan | Ukuran Model | Batch Size |
|----------|----------------------------|--------------|------------|
| LSTM | 15 ms | 22 MB | 64 |
| IndoBERT | 28 ms | 420 MB | 32 |

SIMPULAN

Penelitian ini berhasil melakukan analisis sentimen terhadap ulasan pengguna aplikasi e-commerce di Indonesia dengan membandingkan dua pendekatan pembelajaran mendalam, yaitu Long Short-Term Memory (LSTM) dan Bidirectional Encoder Representations from Transformers (IndoBERT).

Hasil eksperimen menunjukkan bahwa model IndoBERT secara konsisten unggul dalam seluruh metrik evaluasi utama, dengan capaian akurasi sebesar 90,6 % dan F1-score sebesar 90,1 %, dibandingkan LSTM yang hanya mencapai akurasi 83,2 % dan F1-score 82,7 %. Uji statistik McNemar menunjukkan bahwa selisih kinerja ini signifikan secara statistik ($p < 0,01$). Analisis confusion matrix dan distribusi prediksi memperkuat kesimpulan bahwa IndoBERT lebih akurat dalam menangani ekspresi sentimen yang ambigu dan kontekstual.

Kelebihan IndoBERT terletak pada kemampuannya menangkap makna semantik kata berdasarkan konteks kalimat secara dua arah, berkat arsitektur transformer dan embedding kontekstual. Hal ini sangat penting dalam bahasa Indonesia yang kaya dengan struktur majemuk, idiom, dan ekspresi tidak langsung.

Di sisi lain, IndoBERT memiliki keterbatasan dalam hal waktu inferensi dan ukuran model yang lebih besar dibanding LSTM. Oleh karena itu, dalam penerapan sistem real-time seperti chatbot atau layanan pelanggan berbasis ulasan, perlu dipertimbangkan penggunaan versi distilasi (DistilBERT) atau optimasi model lebih lanjut.

Secara keseluruhan, studi ini memberikan kontribusi nyata dalam pemanfaatan NLP modern untuk e-commerce Indonesia, serta membuka peluang pengembangan sistem monitoring sentimen otomatis yang lebih akurat, adaptif, dan dapat digunakan untuk pengambilan keputusan bisnis yang lebih cerdas. Berdasarkan hasil ini, kami mengusulkan integrasi model klasifikasi berbasis IndoBERT ke dalam sistem back-end e-commerce sebagai komponen : Modul “Alert Negatif Otomatis” bagi CS, Dashboard Reputasi Produk berdasarkan sebaran sentimen harian, Filter otomatis untuk menyaring ulasan palsu atau spam menggunakan skor ambiguitas.

UCAPAN TERIMAKASIH

Ucapan terimakasih kami hantarkan kepada semua pihak yang telah membantu terbitnya artikel ini.

DAFTAR PUSTAKA

- [1] H. A. Widyanto and J. O. Haryanto, “Mapping the e-business ecosystem in Indonesia: A comprehensive analysis,” *Handb. Res. Innov. Dev. E-Commerce E-bus. ASEAN*, no. January 2021, pp. 159–178, 2020, doi: 10.4018/978-1-7998-4984-1.ch009.
- [2] A. Romadhony, S. Al Faraby, R. Rismala, U. N. Wisesti, and A. Arifianto, “Sentiment Analysis on a Large Indonesian Product Review Dataset,” *J. Inf. Syst. Eng. Bus. Intell.*, vol. 10, no. 1, pp. 167–178, 2024, doi: 10.20473/jisebi.10.1.167-178.
- [3] S. Imron, E. I. Setiawan, and J. Santoso, “Deteksi Aspek Review E-Commerce Menggunakan IndoBERT Embedding dan CNN,” *J. Intell. Syst. Comput.*, vol. 5, no. 1, pp. 10–16, 2023, doi: 10.52985/insyst.v5i1.267.
- [4] A. Ratino, R. Astri, and P. Anggraini, “Implementasi Framework Laravel Dalam Pengembangan Aplikasi E-Commerce Untuk Toko Jago Software,” *J. Informatics Business*, vol. 01, no. 02, pp. 33–43, 2023.
- [5] R. Astri, L. P. Hung, S. B. Sura, A. Kamal, and R. Yuliet, “Sentiment analysis using naive bayes for reviews of visitors to Padang City beach tourism after the COVID-19 pandemic,” *E3S Web Conf.*, vol. 464, 2023, doi: 10.1051/e3sconf/202346406002.
- [6] M. Z. Arifin, S. Yunan, A. Noertjahyana, and A. Mohamed, “Analyzing the Indonesian sentiment to rohingya refugees using IndoBERT model,” vol. 8, no. 2, pp. 180–191, 2024.
- [7] 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, 2023, doi: 10.1186/s40537-023-00782-9.
- [8] 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.
- [9] A. Chowanda, R. Sutoyo, S. Achmad, E. W. Andangsari, S. M. Isa, and T. K. Chen, "Modeling Emotions Recognition on Indonesian Product Review By Combining Bert, Cnn, and Lstm Architecture," *Int. J. Innov. Comput. Inf. Control*, vol. 20, no. 3, pp. 929–944, 2024, doi: 10.24507/ijicic.20.03.929.
- [10] D. G. Mandhasiya, H. Murfi, and A. Bustamam, "The hybrid of BERT and deep learning models for Indonesian sentiment analysis," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 33, no. 1, pp. 591–602, 2024, doi: 10.11591/ijeecs.v33.i1.pp591-602.
- [11] E. Yulianti and N. K. Nissa, "ABSA of Indonesian customer reviews using IndoBERT: single-sentence and sentence-pair classification approaches," *Bull. Electr. Eng. Informatics*, vol. 13, no. 5, pp. 3579–3589, 2024, doi: 10.11591/eei.v13i5.8032.
- [12] A. G. Yuda, "Comparison of Service and Ease of e-Commerce User Applications Using BERT," vol. 07, no. 02, pp. 98–107, 2024.
- [13] U. Khairani, V. Mutiawani, and H. Ahmadian, "Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi Emosi Pada Komentar Akun Berita Instagram," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 4, pp. 887–894, 2024, doi: 10.25126/jtiik.1148315.
- [14] S. Khomsah and A. S. Aribowo, "Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia," *J. Resti*, vol. 1, no. 3, pp. 648–654, 2017.
- [15] J. C. Setiawan, K. M. Lhaksana, and B. Bunyamin, "Sentiment Analysis of Indonesian TikTok Review Using LSTM and IndoBERTweet Algorithm," *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 8, no. 3, pp. 774–780, 2023, doi: 10.29100/jipi.v8i3.3911.