# ANALISIS SENTIMEN PADA ULASAN *E-COMMERCE* MENGGUNAKAN STRATEGI ADAPTASI *PRE-TRAINED MODEL* INDO-ELECTRA

**Muhammad Anbiya Fatah[1)], Ajeng Miranti[2)], Desvita Sarhani Koly[3)], dan Muhammad Fajar Syuhada[3)]**
[1,2,3)]Teknik Informatika, Fakultas Teknik, Komputer dan Desain, Universitas Nusa Putra
Jl. Raya Cibolang No.21, Cibolang Kaler, Kec.Cisaat, Kabupaten Sukabumi, Jawa Barat 43152
e-mail: muhammad.anbiya_ti23@nusaputra.ac.id[1)] , ajeng.miranti_ti22@nusaputra.ac.id[2)] , desvita.sarhani_ti23@nusaputra.ac.id[3)], muhammad.fajar_ti23@nusaputra.ac.id[4)]

Korespondensi: e-mail: muhammad.anbiya_ti23@nusaputra.ac.id

## ABSTRAK

Sektor *e-commerce* di Indonesia telah bertransformasi menjadi pilar utama dalam ekosistem ekonomi digital saat ini. Namun, besarnya volume ulasan pelanggan seringkali menjadi kendala bagi pelaku bisnis untuk mengekstraksi informasi kepuasan pelanggan secara manual. Tantangan signifikan dalam pemrosesan data ulasan adalah penggunaan bahasa yang tidak terstruktur, adanya kata-kata singkatan, serta ketidakseimbangan distribusi kelas sentimen yang ekstrem. Penelitian ini mengajukan strategi adaptasi *Transfer Learning* memanfaatkan model bahasa *pre-trained* Indo-ELECTRA yang di-fine-tune khusus untuk mengategorikan ulasan Tokopedia ke dalam tiga kategori: sentimen positif, netral, dan negatif. Dataset yang diuji mencakup 65,544 entri data. Tahapan evaluasi kinerja model diukur menggunakan instrumen *Confusion Matrix* dan kurva ROC-AUC. Hasil eksperimen memperlihatkan bahwa arsitektur Indo-ELECTRA memiliki ketajaman dalam menangkap konteks ulasan dengan capaian Akurasi sebesar 97,83% dan *F1-Score macro* sebesar 0,58. Di samping itu, perolehan nilai AUC sebesar 0,96 menegaskan reliabilitas model dalam mengidentifikasi perbedaan antar kelas sentimen. Riset ini mengonfirmasi bahwa optimalisasi model transformer melalui teknik *fine-tuning* sangat kompeten dalam mendukung sistem pemantauan reputasi layanan di industri *e-commerce.*

**Kata Kunci**: Analisis Sentimen, Tokopedia, Klasifikasi Teks, *Transfer Learning*, Indo-ELECTRA.

## ABSTRACT

*The e-commerce sector in Indonesia has transformed into a major pillar of the current digital economic ecosystem. However, the massive volume of customer reviews often becomes an obstacle for business actors to manually extract customer satisfaction information. Significant challenges in processing review data are the use of unstructured language, the presence of abbreviations, and extreme class distribution imbalance. This study proposes a Transfer Learning adaptation strategy utilizing the Indo-ELECTRA pre-trained language model, which is fine-tuned to categorize Tokopedia reviews into three categories: positive, neutral, and negative sentiment. The tested dataset includes 65,544 data entries. The model performance evaluation stages were measured using Confusion Matrix and ROC-AUC curve instruments. Experimental results show that the Indo-ELECTRA architecture has the sharpness to capture review context with an Accuracy achievement of 97.83% and a macro F1-Score of 0.58. In addition, the AUC value of 0.96 confirms the model's reliability in identifying differences between sentiment classes. This research confirms that optimizing transformer models through fine-tuning techniques is highly competent in supporting service reputation monitoring systems in the e-commerce industry.*

***Keywords***: *Sentiment Analysis, Tokopedia, Text Classification, Transfer Learning, Indo-ELECTRA.*

## I. PENDAHULUAN

Integrasi teknologi informasi telah mengubah paradigma transaksi ekonomi masyarakat di Indonesia. Sistem belanja yang sebelumnya bersifat luring, perlahan beralih menuju pola transaksi berbasis daring melalui berbagai platform digital. Perubahan perilaku ini menjadikan aktivitas belanja terasa lebih praktis, fleksibel, dan efisien [1]. Tokopedia menjadi salah satu penyedia layanan *e-commerce* dengan basis massa pengguna yang sangat masif di tanah air. Platform Tokopedia menyediakan ruang bagi konsumen untuk memberikan umpan balik melalui fitur ulasan produk. Fitur ini menjadi referensi krusial bagi calon pembeli maupun penjual dalam menilai kualitas sebuah layanan [2]. Meskipun pengguna diberikan kebebasan berpendapat, besarnya volume data dan beragamnya gaya penulisan seringkali memunculkan kendala dalam pemantauan kualitas ulasan secara langsung. Analisis sentimen hadir sebagai solusi

komunikasi digital yang dilakukan untuk mengekstraksi opini dari ulasan pelanggan, baik berupa pujian, kritik, maupun keluhan terkait aspek produk, harga, dan layanan pengiriman [3]. Jika data ulasan ini tidak dikelola dengan tepat, pelaku bisnis akan kesulitan merespons ketidakpuasan pelanggan secara cepat. Oleh sebab itu, diperlukan sebuah kerangka kerja berbasis kecerdasan buatan (*Artificial Intelligence*) untuk memproses ulasan tersebut secara otomatis.

Inovasi kecerdasan buatan dapat diterapkan untuk melakukan klasifikasi sentimen pada ribuan komentar di platform Tokopedia. Salah satu pendekatan yang paling mutakhir dalam pemrosesan bahasa alami adalah metode *transfer learning* dengan menerapkan teknik *fine-tuning*. Strategi ini bekerja dengan mengadopsi model saraf tiruan yang telah melalui pelatihan awal pada korpus data besar, lalu menyesuaikannya kembali untuk kebutuhan klasifikasi tertentu [4]. Beberapa penelitian terdahulu telah mengimplementasikan deteksi sentimen pada ulasan daring menggunakan algoritma konvensional. Sebagai contoh, penerapan metode *Support Vector Machine (SVM)* pada data ulasan layanan digital menghasilkan akurasi yang cukup baik, namun masih kesulitan dalam mengenali hubungan semantik antar kata yang kompleks [5]. Penelitian lainnya menguji performa *random forest* untuk pemetaan kepuasan pelanggan, akan tetapi masih menemukan hambatan dalam menangani ambiguitas makna pada kalimat yang bersifat sarkastik [6]. Melihat adanya keterbatasan pada algoritma klasifikasi tradisional dalam membedakan nuansa bahasa Indonesia, dibutuhkan pendekatan yang lebih peka terhadap konteks linguistik yang dalam. Maka dari itu, studi ini mengusulkan implementasi *transfer learning* melalui proses *fine-tuning* pada model *pre-trained* Indo-ELECTRA untuk mengategorikan ulasan Tokopedia ke dalam kelas positif, netral, dan negatif. Strategi ini diharapkan mampu memberikan akurasi yang lebih presisi serta membantu perusahaan dalam meningkatkan efisiensi sistem penanganan keluhan pelanggan di sektor *e-commerce*.

## II. KAJIAN TEORI

### *A. Analisis Sentimen*

Analisis sentimen merupakan proses pengolahan data teks untuk mengidentifikasi sikap, pendapat, atau emosi yang terkandung dalam ulasan, baik yang bersifat subjektif maupun objektif terhadap suatu produk [7]. Konsep ini bertujuan untuk mengelompokkan teks ke dalam polaritas sentimen tertentu guna melihat kecenderungan persepsi publik. Analisis ini sangat krusial bagi dunia industri karena membantu dalam pengambilan keputusan strategis berdasarkan masukan langsung dari konsumen.

### *B. Natural Language Processing (NLP)*

*Natural Language Processing (NLP)* adalah disiplin ilmu yang memadukan teknik komputer, kecerdasan buatan, dan linguistik untuk memungkinkan mesin memahami serta menginterpretasikan bahasa manusia secara otomatis [8]. Melalui pendekatan NLP, komputer dapat mengenali pola kalimat yang tidak standar pada ulasan *e-commerce*, seperti penggunaan *slang* atau bahasa gaul. Implementasi NLP yang tepat memungkinkan sistem memberikan respons yang sangat akurat dan menyerupai logika pemahaman manusia.

### *C. Transfer Learning*

*Transfer Learning* didefinisikan sebagai teknik pembelajaran mesin di mana model yang telah dilatih untuk tugas umum digunakan sebagai titik awal untuk tugas spesifik yang baru [9]. Dengan memanfaatkan basis pengetahuan yang sudah ada, model tidak perlu dilatih dari nol, sehingga mampu meningkatkan performa klasifikasi secara signifikan meskipun dataset yang tersedia untuk tugas spesifik tersebut relatif terbatas.

### *D.* Indo-ELECTRA

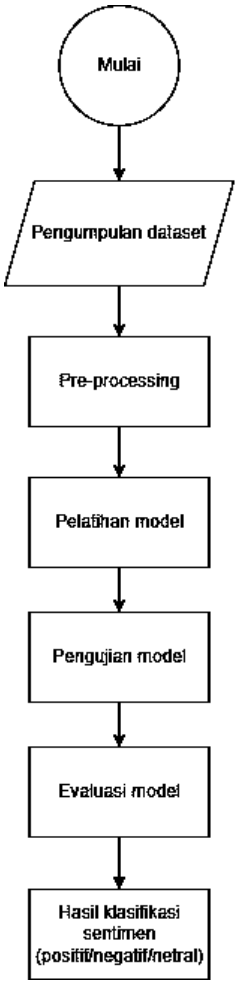
Model ELECTRA bekerja dengan mekanisme unik yang membedakannya dari arsitektur *transformer* lainnya melalui penggunaan sistem diskriminator untuk mendeteksi penggantian token (*replaced token detection).* Pada tahun 2020, dikembangkan varian model bahasa Indonesia khusus yang dikenal sebagai Indo-ELECTRA [10]. Arsitektur ini dirancang secara optimal untuk menangani struktur bahasa Indonesia, menjadikannya sangat kompetitif dalam menjalankan berbagai tugas klasifikasi teks

dibandingkan model pendahulunya.

### *E. Fine-Tuning*

*Fine-tuning* merupakan tahap penyesuaian parameter pada model *pre-trained* agar lebih selaras dengan tugas klasifikasi sentimen melalui proses pelatihan pada dataset yang spesifik [11]. Dalam penelitian ini, *fine-tuning* dilakukan untuk menyempurnakan kemampuan model dalam mengenali karakteristik ulasan pada Tokopedia sehingga mampu memberikan prediksi yang lebih akurat sesuai dengan tujuan riset.

## III. METODE




**Gambar I.** Diagram Alur Sistem
Sumber: Hasil Penelitian (2026)

### *A. Pengumpulan Dataset*

Pada tahap ini, dilakukan pengumpulan data ulasan produk dari platform *e-commerce* Tokopedia. Dataset yang digunakan diperoleh dari repositori publik kaggle berjudul "Tokopedia *Product Review* 2025". Dataset ini mencakup ulasan pelanggan dari berbagai kategori produk seperti Makanan & Minuman, Kesehatan, Elektronik, Pertukangan, dan Olahraga. Total dataset yang digunakan adalah 65,544 data, berikut adalah tabel distribusi kelas dari dataset tersebut:

**Tabel I.** Jumlah Dataset
Sumber: Hasil Penelitian (2026)

| Kelas | Jumlah Data |
|---|---|
| Sentimen Positif | 58,068 |

| | |
|---|---|
| Sentimen Netral | 4,547 |
| Sentimen Negatif | 2,928 |

*B. Preprocessing*

Tahap *pra*-pemrosesan data bertujuan untuk mengubah data teks mentah menjadi format terstruktur yang dapat dipahami oleh model *Deep Learning*. Mengingat data yang digunakan berasal dari ulasan pelanggan di platform Tokopedia yang memiliki karakteristik tidak terstruktur (penggunaan bahasa gaul, singkatan, dan emoji), tahapan ini sangat krusial untuk memastikan kualitas input model. Pada penelitian ini, tahapan *pra*-pemrosesan meliputi empat proses utama: penyeragaman huruf, pembersihan data, normalisasi kata, dan tokenisasi.

*C. Training*

Penelitian ini menerapkan metode *fine-tuning* pada model Indo-ELECTRA (menggunakan varian ChristopherA08/IndoELECTRA) dengan menambahkan lapisan klasifikasi linear pada output token. Lapisan ini bertugas memetakan representasi fitur model Indo-ELECTRA ke dalam tiga output sentimen, yaitu positif, netral, dan negatif, dengan meminimalkan fungsi kerugian *Cross-Entropy Loss.*

Proses pelatihan dilakukan menggunakan *framework* PyTorch dan pustaka Hugging Face *Transformers.* Proses ini dilakukan dengan menyesuaikan bobot model *pra*-latih agar spesifik terhadap karakteristik bahasa ulasan pada platform *e-commerce* Tokopedia. Konfigurasi hiperparameter yang digunakan dalam proses pelatihan dirinci pada Tabel II berikut:

**Tabel II**. Parameter Pelatihan Indo-ELECTRA
Sumber: Hasil Penelitian (2026)

| Nama | Parameter |
|---|---|
| *Optimizer* | AdamW |
| *Learning Rate* | 2e-5 |
| *Epoch* | 3 |
| *Batch Size* | 16 |
| *Max Sequence Length* | 128 |

Model dievaluasi pada setiap akhir *epoch*, dan pos-el (*checkpoint*) model dengan nilai *F1-Score* terbaik akan disimpan sebagai model final untuk digunakan pada tahap pengujian sistem.

*D.* Pengujian Model

Pengujian model dilakukan menggunakan data uji sebesar 20% dari total dataset yang telah dipisahkan sebelum proses pelatihan dimulai. Langkah ini bertujuan untuk mengukur kemampuan generalisasi model Indo-ELECTRA dalam mengklasifikasikan data ulasan baru dari Tokopedia yang belum pernah dipelajari sebelumnya selama fase pelatihan

Evaluasi hasil pengujian dilakukan dengan menggunakan instrumen *Confusion Matrix* untuk memetakan prediksi model terhadap label sebenarnya secara detail. Melalui instrumen ini, kinerja model diukur secara komprehensif menggunakan metrik *Accuracy, Precision, Recall,* dan *F1-Score.*

*E.* Evaluasi Model

Evaluasi kinerja didasarkan pada metrik Akurasi untuk mengukur ketepatan prediksi secara global, serta *F1-Score* sebagai metrik utama. Penggunaan *F1-Score* sangat krusial mengingat karakteristik dataset ulasan yang sering kali tidak seimbang (*imbalanced*), sehingga rata-rata harmonik antara *Precision* dan *Recall* memberikan gambaran performa yang lebih objektif.

Selain itu, analisis kesalahan dilakukan menggunakan *Confusion Matrix* untuk memetakan tingkat *False Positive* dan *False Negative*, serta kurva ROC-AUC untuk mengukur kemampuan model dalam membedakan antar kelas (Positif, Netral, Negatif) pada berbagai ambang batas. Berikut adalah rumus

pengukuran evaluasi performa yang digunakan:

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

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ Score = \frac{2\ x\ Precision\ x\ Recall}{Precision\ x\ Recall} \quad (4)$$

Keterangan :

- *Accuracy* : Rasio antara jumlah prediksi benar dengan total jumlah prediksi yang dibuat.
- *Precision* : Rasio antara jumlah prediksi kelas tertentu yang benar dengan total prediksi untuk kelas tersebut oleh model.
- *Recall* : Rasio antara jumlah prediksi kelas tertentu yang benar dengan total jumlah data asli pada kelas tersebut.
- *F1-Score* : Nilai yang menggabungkan *Precision* dan *Recall* untuk memberikan Gambaran keseluruhan tentang kinerja model pada tiap label sentimen

## IV. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah ulasan produk dari pengguna platform *e-commerce* Tokopedia yang diperoleh dari repositori Kaggle (dataset "Tokopedia *Product Reviews*"). Bagian ini membahas hasil pengujian dan analisis kinerja model analisis sentimen yang dikembangkan menggunakan strategi adaptasi *pre-trained* model Indo-ELECTRA (model ChristopherA08/IndoELECTRA dari Hugging Face).

Data dibagi menjadi 80% data latih dan 20% data validasi. Data validasi digunakan sebagai data evaluasi untuk menghitung *confusion matrix, classification report*, dan ROC–AUC. Proporsi ini dipilih agar model memperoleh data latih yang cukup besar untuk mempelajari pola bahasa informal, sekaligus menyediakan data validasi yang memadai untuk evaluasi kinerja secara obyektif. Meskipun pembagian ini tidak sepenuhnya mengatasi ketidakseimbangan kelas, penyajian distribusi data memberikan konteks yang lebih jelas saat menafsirkan metrik seperti *F1-score* dan AUC.

**Tabel III.** Pembagian Dataset
Sumber: Hasil Penelitian (2026)

| Subset | Jumlah Data | Persentase |
|---|---|---|
| Data Latih | 52,434 | 80% |
| Data Validasi | 13,109 | 20% |
| **Total** | **65,544** | **100%** |

Pembagian ini mengikuti praktik umum pada penelitian analisis sentimen berbasis NLP untuk menjaga generalisasi model, dengan tetap mempertahankan karakteristik distribusi kelas asli. Namun, dominasi kuat kelas positif menyebabkan model cenderung bias ke kelas mayoritas sehingga evaluasi menggunakan *macro-averaged F1-score* menjadi penting untuk menilai performa pada kelas negatif dan netral.

### *A. Hasil Pelatihan Model Indo-ELECTRA*

Proses *fine-tuning* model Indo-ELECTRA dilakukan selama 3 *epoch* dengan *learning rate* 2e-5 dan *batch size* 16 pada data latih ulasan Tokopedia. Nilai *training loss* dan *validation loss* pada setiap *epoch* ditunjukkan pada Tabel IV, yang menggambarkan konvergensi model selama proses pelatihan.

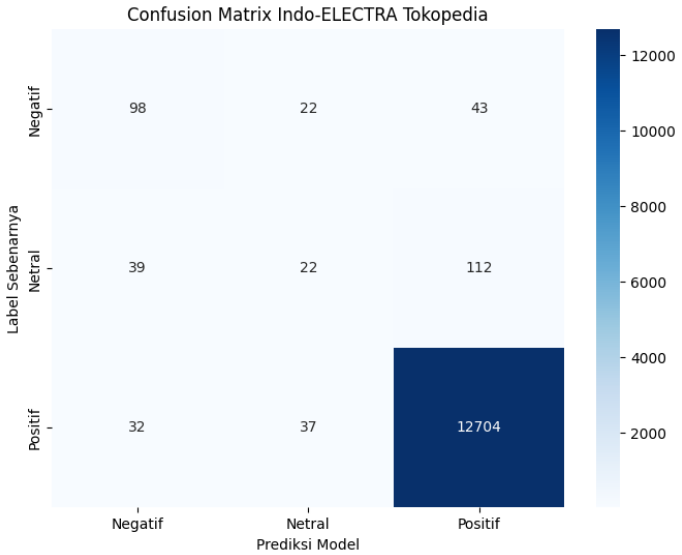
**Tabel IV.** Hasil Pelatihan Model Indo-ELECTRA
Sumber: Hasil Penelitian (2026)

| *Epoch* | *Training Loss* | *Validation Loss* |
|---|---|---|
| 1 | 0,065900 | 0,086044 |
| 2 | 0,056600 | 0,080509 |
| 3 | 0,041200 | 0,089740 |

Nilai *training loss* konsisten menurun dari 0,0659 menjadi 0,0412, menunjukkan model berhasil mempelajari pola data secara bertahap. *Validation loss* sempat turun hingga *epoch* ke-2 dan sedikit naik pada *epoch* ke-3, yang mengindikasikan model mulai mendekati titik jenuh (*early overfitting* ringan) tetapi masih dalam batas yang dapat diterima.

### *B. Evaluasi kinerja model berdasarkan Confusion Matrix*

Evaluasi pertama dilakukan dengan menganalisis *Confusion Matrix* pada *test set* untuk melihat distribusi prediksi benar dan salah.




**Gambar II.** *Confusion Matrix* Model Indo-ELECTRA
Sumber: Hasil Penelitian (2026)

*Confusion matrix* menunjukkan bahwa model sangat akurat pada kelas positif dengan *true positive (TP)* mencapai 12,704. Namun, kesalahan terbanyak terjadi pada kelas netral yang sering salah diklasifikasikan sebagai positif sebanyak 112 kasus, serta ulasan negatif yang terdeteksi sebagai positif sebanyak 43 kasus. Hal ini mengindikasikan dampak dari data imbalance, di mana kelas positif mendominasi, menyebabkan bias model terhadap kelas mayoritas.

### *C. Evaluasi kinerja model berdasarkan Laporan Klasifikasi*

Untuk mengukur kualitas klasifikasi secara keseluruhan, dihitung metrik *Precision, Recall, F1-Score*, dan *Accuracy (macro-averaged* untuk menangani imbalance).

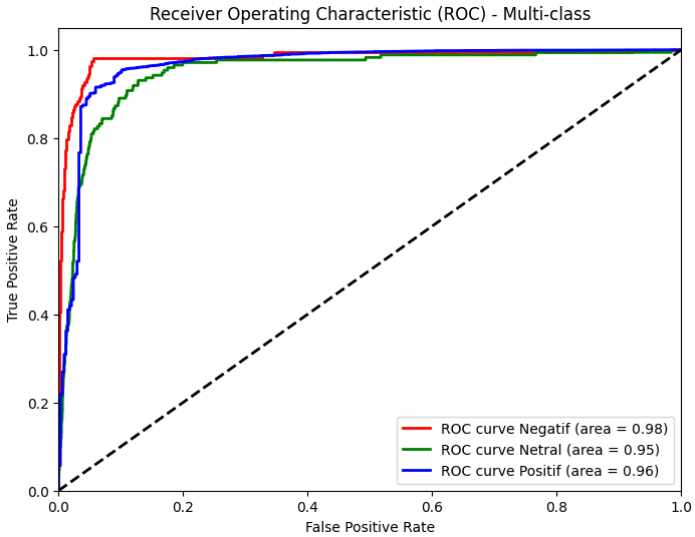
**Tabel V.** Laporan Klasifikasi
Sumber: Hasil Penelitian (2026)

| Kelas | *Precision* | *Recall* | *F1-Score* | *Support* |
|---|---|---|---|---|
| Negatif | 0,58 | 0,60 | 0,59 | 163 |
| Netral | 0,27 | 0,13 | 0,17 | 173 |
| Positif | 0,99 | 0,99 | 0,99 | 12,773 |
| *Macro Avg* | 0,61 | 0,57 | 0,58 | 13,109 |

Model mencapai *accuracy* sebesar 97,83% dan *F1-score macro* sebesar 0,58. Nilai *accuracy* yang tinggi dipengaruhi oleh dominasi kelas positif, di mana model menunjukkan performa hampir sempurna pada kelas tersebut (*F1-score* 0,99). Namun, performa pada kelas netral masih relatif rendah (*F1-score* 0,17), yang mengindikasikan kesulitan model dalam membedakan ulasan yang bersifat ambigu, meskipun proses fine-tuning menggunakan Indo-ELECTRA telah meningkatkan kemampuan model dalam memahami konteks bahasa informal.

### *D. Analisis kurva ROC dan Nilai AUC*

Performa model dievaluasi lebih lanjut menggunakan kurva *Receiver Operating Characteristic (ROC)* dengan pendekatan *One-vs-Rest* untuk *multi-class*.




**Gambar III.** Kurva ROC-AUC Model Indo-ELECTRA
Sumber: Hasil Penelitian (2026)

Nilai AUC *macro* sebesar 0,96, mendekati 1,0 mengindikasikan bahwa model memiliki kemampuan diskriminatif yang sangat baik dalam memisahkan kelas sentimen, terutama pada kelas negatif dan positif. AUC netral sedikit lebih rendah, yaitu 0,95 karena imbalance data, tetapi secara keseluruhan kurva yang melengkung tajam ke sudut kiri atas menunjukkan sensitivitas tinggi dengan *false positive rate* rendah, bahkan pada data *noisy* seperti ulasan *e-commerce.*

### *E. Pembahasan Umum Hasil Penelitian*

Adaptasi model Indo-ELECTRA melalui *fine-tuning* terbukti efektif untuk tugas analisis sentimen tiga kelas pada ulasan *e-commerce* berbahasa Indonesia, dengan *accuracy* 97,83%, *F1-macro* 0,58, dan AUC 0,96. Nilai *accuracy* yang tinggi dipengaruhi oleh dominasi kelas positif, sedangkan *F1-macro* yang relatif rendah menyoroti permasalahan ketidakseimbangan data, di mana performa pada kelas minoritas (negatif dan netral) lebih rendah dibandingkan kelas mayoritas.

Dibandingkan penelitian serupa, seperti analisis ulasan GoPay menggunakan Indo-ELECTRA [1] yang memperoleh *accuracy* 95% dan *F1-score* 0,95 pada dataset yang lebih seimbang, hasil penelitian ini menunjukkan *accuracy weighted* yang lebih tinggi namun nilai *macro-average* yang lebih rendah akibat ketidakseimbangan kelas yang ekstrem, dengan proporsi sentimen positif mencapai sekitar 97% dari total data. Keunggulan Indo-ELECTRA terlihat dari penurunan *training loss* yang cepat dari 0,0659 menjadi 0,0412, yang didukung oleh mekanisme *replaced token detection* yang efektif dalam menangani teks informal, *slang*, dan campuran bahasa.

Tahapan pra-pemrosesan, tokenisasi dengan *max sequence length* 128, penggunaan lowercase, serta pengaturan hiperparameter seperti *learning rate* 2e-5 dan *batch size* 16 berkontribusi dalam mengurangi *noise* pada data. Namun demikian, kesalahan klasifikasi pada ulasan sarkastik dan sentimen netral masih ditemukan, yang diduga disebabkan oleh jumlah *support* kelas minoritas yang relatif rendah, yaitu hanya berkisar antara 163-173 sampel.

Model ini berpotensi diimplementasikan pada platform *e-commerce* seperti Tokopedia untuk mendukung deteksi otomatis keluhan pelanggan, monitoring tingkat kepuasan, serta rekomendasi perbaikan produk atau layanan, misalnya melalui prioritas penanganan ulasan negatif. Bagi pelaku bisnis, penerapan model ini dapat membantu meningkatkan kualitas layanan, rating toko, dan penjualan. Sebagai pengembangan ke depan, penelitian selanjutnya dapat menerapkan teknik penanganan data tidak seimbang seperti *class weighting, oversampling*, atau penggunaan *focal los*s guna meningkatkan performa klasifikasi pada kelas minoritas, khususnya sentimen negatif dan netral, sehingga model lebih optimal untuk diterapkan pada skenario aplikasi *real-time*.

## V. KESIMPULAN

Penelitian ini berhasil mengadaptasi model *Pre-trained* Indo-ELECTRA untuk menganalisis sentimen ulasan *e-commerce* di Tokopedia berbahasa Indonesia, dengan hasil yang cukup memuaskan. Model yang telah di-fine-tune selama 3 *epoch* menggunakan *learning rate* 2e-5 dan *batch size* 16 mampu mencapai akurasi 97,83%, *F1-score macro* 0,58, dan AUC 0,96. Angka-angka tersebut menunjukkan kemampuan model dalam membedakan berbagai kelas sentimen secara efektif. Kelebihan model terlihat dari penurunan secara konsisten nilai *training loss* dari 0,0659 menjadi 0,0412, serta performa yang hampir sempurna pada kelas sentimen positif dengan *F1-score* mencapai 0,99. Mekanisme deteksi token yang digantikan pada Indo-ELECTRA terbukti efektif dalam menghadapi teks yang informal, menggunakan *slang*, atau menggabungkan berbagai jenis bahasa yang sering ditemukan pada ulasan *e-commerce*.

Namun, penelitian ini juga mengungkapkan tantangan utama berupa ketidakseimbangan data yang sangat signifikan, di mana sekitar 97% dari total data bersifat positif. Hal ini menyebabkan model cenderung berbias terhadap kelas mayoritas, sehingga performa pada kelas minoritas, terutama sentimen netral dan negatif, masih relatif rendah. *F1-score* untuk kelas netral hanya mencapai 0,17, sementara untuk kelas negatif sekitar 0,59. Kesalahan klasifikasi yang paling banyak terjadi adalah ulasan netral yang dianggap sebagai ulasan positif oleh model.

Model ini memiliki kemungkinan untuk digunakan dalam berbagai penerapan praktis, seperti mendeteksi keluhan pelanggan secara otomatis, memantau kepuasan pengguna, dan menentukan prioritas dalam menangani ulasan negatif di platform *e-commerce*. Untuk pengembangan di masa depan, dianjurkan untuk menerapkan teknik penanganan data yang tidak seimbang, seperti *class weighting, oversampling,* atau *focal loss,* agar performa model pada kelas minoritas meningkat dan model dapat dioptimalkan untuk digunakan dalam aplikasi *real-time*.

## VI. DAFTAR PUSTAKA

[1] Temasek and Bain & Company, “e-Conomy SEA 2023: Reaching new heights: Navigating the path to profitability,” 2023.

[2] M. S. Rosyadi, D. S. Rusdianto, and C. A. Sari, “Analisis sentimen ulasan produk pada marketplace tokopedia menggunakan algoritma naive bayes,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 1, 2023.

[3] Liu and Bing, *Sentiment analysis: mining opinions, sentiments, and emotions*. 2015.

[4] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. 2008.

[5] A. S. Ritonga and D. Purwanti, "Sentiment Analysis of *E-commerce* Product Reviews Using Support Vector Machine," *Journal of Information System Exploration and Research*, vol. 1, no. 1, pp. 13–14, 2023.

[6] P. Pratama and A. S. Girsang, "Sentiment Analysis of Indonesian *E-commerce* Reviews Using Random Forest and SMOTE," in *International Conference on Information Management and Technology (ICIMTech)*, 2021, pp. 540–545.

[7] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014.

[8] D. Jurafsky and J. H. Martin, *Speech and Language Processing*. 2023.

[9] S. J. Pan and Q. Yang, "A Survey on *Transfer Learning*," *IEEE Trans Knowl Data Eng*, vol. 22, no. 10, pp. 1345–1359, 2010.

[10] B. Wilie *et al.*, "IndoBenchmark: Towards a Unified Benchmark for Indonesian Natural Language Processing," in *Proceedings of the 1st Conference on Asia-Pacific Chapter of the Association for Computational Linguistics*, 2020, pp. 249–260.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, 2019, pp. 4171–4186.