Vol. 6, No. 1 - Juni 2026, Hal. 51-64
DOI: 10.59395/jitp.v6i1.197

Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Roblox Berbahasa Indonesia Menggunakan IndoBERT dengan Augmentasi Data Berbasis LLM

Yolanda Gloria[1], Dedi Saputra[2], Siti Nurdiani[3], Eva Meilinda[4]
[1234]Program Studi Informatika, Universitas Bina Sarana Informatika, Pontianak, Kalimantan Barat, Indonesia
e-mail: 15220466@bsi.ac.id[1], dedi.dst@bsi.ac.id[2], siti.sxd@bsi.ac.id[3], eva.emd@bsi.ac.id[4]

Received:	Revised:	Accepted:	Available online:
07.06.2026	19.06.2026	25.06.2026	30.06.2026

Abstract: *Roblox stands out as one of the most widely adopted online gaming platforms in Indonesia, generating a large and continuously growing volume of Indonesian-language reviews on Google Play Store. Earlier studies typically assigned a single sentiment label per review, leaving specific aspectss users praise or criticize unidentified. This study implements Aspect-Based Sentiment Analysis (ABSA) using IndoBERT fine-tuned independently for four aspects: technical, system, content, and social. The dataset consists of 3,619 cleaned and labeled reviews annotated through a hybrid strategy combining the Llama-4-Scout-17B model with manual verification. Neutral-class scarcity was resolved by generating 550 synthetic samples via Llama 3.3 70B. Results show average F1-Macro improved substantially from 0.5739 to 0.8550 after augmentation, with all aspects surpassing the 0.80 threshold and outperforming the XLM-RoBERTa baseline (0.8035). An ablation study confirms data augmentation as the most critical pipeline component, statistically validated through McNemar's Test ($p < 0.05$). LIME-based interpretability reveals contextual linguistic patterns, while error analysis identifies vulnerabilities in underrepresented classes. Sentiment distribution shows the System aspect is the most negatively reviewed (93.8%), while Content receives the most positive feedback (86.3%). This study validates the effectiveness of combining IndoBERT with LLM-based augmentation for multidimensional sentiment analysis of informal and imbalanced Indonesian-language game reviews.*
Keywords: *Aspect-Based Sentiment Analysis; IndoBERT; Transformer; Game Application Reviews; Data Augmentation; Roblox.*

Abstrak: Roblox merupakan salah satu platform permainan daring yang paling banyak digunakan di Indonesia, menghasilkan volume ulasan berbahasa Indonesia yang besar dan terus bertambah di Google Play Store. Kajian terdahulu umumnya hanya menetapkan satu label sentimen tunggal per ulasan tanpa diferensiasi aspek. Penelitian ini mengimplementasikan Aspect-Based Sentiment Analysis (ABSA) menggunakan model IndoBERT yang di-fine-tune secara independen untuk empat aspek: teknis, sistem, konten, dan sosial. Dataset terdiri dari 3.619 ulasan berlabel menggunakan strategi anotasi hibrida antara model Llama-4-Scout-17B dan verifikasi manual. Kelangkaan kelas netral diatasi dengan membangkitkan 550 sampel sintetis melalui Llama 3.3 70B. Hasil eksperimen menunjukkan rata-rata F1-Macro meningkat dari 0,5739 menjadi 0,8550 pascaaugmentasi, seluruh aspek melampaui ambang 0,80 dan mengungguli baseline XLM-RoBERTa (0,8035). Ablation study mengkonfirmasi augmentasi data sebagai komponen terpenting, terkonfirmasi secara statistik melalui McNemar's Test ($p < 0,05$). Analisis LIME mengungkap pola kebahasaan kontekstual, sedangkan analisis error mengidentifikasi kelemahan pada kelas minoritas. Aspek Sistem paling banyak mendapat ulasan negatif (93,8%), sementara aspek Konten paling banyak mendapat apresiasi positif (86,3%). Penelitian ini memvalidasi efektivitas kombinasi IndoBERT dengan augmentasi berbasis LLM untuk analisis sentimen multidimensi pada ulasan game berbahasa Indonesia.
Kata kunci: Analisis Sentimen Berbasis Aspek; IndoBERT; Transformer; Ulasan Aplikasi Game; Augmentasi Data; Roblox.

1. PENDAHULUAN

Pertumbuhan ekosistem permainan daring di Indonesia berlangsung dengan laju yang sangat pesat, seiring semakin meluasnya penetrasi smartphone di berbagai kalangan masyarakat. Di tengah persaingan platform yang semakin ketat, Roblox berhasil menorehkan tingkat popularitas yang sangat tinggi, khususnya di tengah komunitas generasi muda. Besarnya basis pengguna aktif tersebut secara otomatis menghasilkan volume ulasan baru yang besar di Google Play Store ulasan yang menyimpan pendapat otentik terkait pengalaman bermain, penilaian terhadap fitur-fitur yang tersedia, maupun berbagai permasalahan yang dihadapi pengguna. Sitorus et al. (2024) menegaskan bahwa data ulasan pengguna game di Google Play Store dapat dimanfaatkan sebagai sumber yang andal untuk mengukur kualitas layanan dan pengalaman pengguna secara lebih terstruktur dan berbasis data.

Persoalan mendasar yang muncul adalah ketika volume ulasan telah mencapai puluhan ribu (Rafianto et al., 2026), analisis yang dilakukan secara konvensional menjadi sesuatu yang tidak memungkinkan untuk dijalankan. Pendekatan berbasis komputasi pun menjadi suatu keharusan yang tidak dapat dielakkan. Sayangnya, sebagian besar penelitian yang telah ada masih bertumpu pada

Vol. 6, No. 1 - Juni 2026, Hal. 51-64
DOI: 10.59395/jitp.v6i1.197

metode machine learning konvensional seperti SVM, Naive Bayes, atau Logistic Regression yang seluruhnya hanya mampu menghasilkan satu label sentimen tunggal tanpa mampu membedakan aspek tertentu yang dibicarakan (Setiawan et al., 2025). Seorang pengguna yang memuji kekayaan konten sekaligus mengeluhkan masalah lag dalam satu ulasan yang sama hanya akan mendapatkan satu label tunggal, sehingga kandungan informasi berharga di dalamnya menjadi tidak tergali.

Aspect-Based Sentiment Analysis (ABSA) muncul sebagai solusi yang tepat atas keterbatasan pendekatan konvensional tersebut. Dengan kapasitasnya dalam memetakan sentimen secara terpisah untuk setiap dimensi atau aspek yang terdapat dalam satu ulasan yang sama, ABSA mampu menghasilkan wawasan yang jauh lebih mendalam dan dapat langsung diaplikasikan untuk pengambilan keputusan (Zhu et al., 2022). Untuk konteks bahasa Indonesia yang kaya akan ragam informal, akronim, serta fenomena alih kode, model transformer IndoBERT yang dibangun dari korpus empat miliar kata berbahasa Indonesia terbukti melampaui performa pendekatan TF-IDF maupun model transformer multibahasa (Wilie et al., 2020).

Sejumlah penelitian terdahulu telah mengeksplorasi analisis sentimen pada ulasan Roblox, namun dengan pendekatan yang masih terbatas. Setiawan et al. (2025) mengklasifikasikan sentimen dari 20.000 ulasan Roblox menggunakan SVM dan memperoleh akurasi 82,51%, meskipun hanya mencakup dua kelas sentimen tanpa diferensiasi aspek. Rafianto et al. (2026) mengintegrasikan LDA dan SVM pada 40.089 ulasan Roblox dan berhasil mencapai akurasi 85,22%, namun analisisnya masih berupa sentimen umum per topik dan belum menyentuh level ABSA yang sesungguhnya. Pada ranah permainan yang berbeda, Bachtiar et al. (2026) membandingkan Naive Bayes dan Logistic Regression untuk ulasan Free Fire, di mana Logistic Regression meraih hasil terbaik 81,63% juga tanpa pembedaan aspek. Maulana et al. (2020) meningkatkan akurasi analisis sentimen ulasan film menggunakan SVM berbasis Information Gain dan memperoleh akurasi 85,65%, namun juga masih terbatas pada dua kelas sentimen tanpa diferensiasi aspek.

Di sisi penelitian ABSA yang memanfaatkan IndoBERT, Jazuli et al. (2025) berhasil menerapkannya pada ulasan mahasiswa dengan F1-score 97,4%, sementara Alfiana et al. (2026) mengaplikasikannya pada ulasan aplikasi transportasi dengan F1-score 78,5% pada dua kelas sentimen. Kesenjangan yang belum terisi adalah penerapan ABSA berbasis IndoBERT dengan tiga kelas sentimen khusus pada ulasan game Roblox berbahasa Indonesia yang penuh dengan ekspresi informal, termasuk penanganan terhadap ketimpangan distribusi kelas netral. Strategi augmentasi berbasis LLM diusulkan sebagai solusi atas permasalahan ketimpangan distribusi tersebut. Bahri et al. (2023) telah menunjukkan bahwa ABSA berbasis transfer learning, termasuk IndoBERT, unggul dibanding metode machine learning konvensional dalam konteks ulasan berbahasa Indonesia.

Penelitian ini mengimplementasikan ABSA berbasis IndoBERT pada 3.619 ulasan Roblox berbahasa Indonesia yang diperoleh dari Google Play Store, dengan cakupan empat aspek utama: teknis, sistem, konten, dan sosial. Tantangan ketimpangan distribusi kelas netral diatasi melalui strategi augmentasi data berbasis LLM. Untuk memvalidasi keunggulan model domain-spesifik terhadap model transformer multibahasa modern, penelitian ini juga menggunakan XLM-RoBERTa sebagai model baseline pembanding, mengacu pada efektivitasnya yang telah terbukti dalam berbagai tugas NLP lintas bahasa (Wilie et al., 2020). Kontribusi utama penelitian ini meliputi: (1) perancangan pipeline ABSA yang komprehensif mencakup augmentasi berbasis LLM dan evaluasi komparatif terhadap baseline XLM-RoBERTa untuk domain ulasan game berbahasa Indonesia, (2) penggabungan augmentasi berbasis LLM dengan fine-tuning IndoBERT secara per-aspek, serta (3) pemetaan distribusi sentimen per aspek yang dapat menjadi rujukan konkret bagi pengembang dalam menetapkan prioritas perbaikan produk. Lebih dari itu, penelitian ini turut membuktikan bahwa perbedaan performa antara IndoBERT dan XLM-RoBERTa bukan sekadar kebetulan, melainkan terkonfirmasi secara statistik melalui McNemar's Test. Untuk membuka jendela ke dalam logika prediksi model, analisis LIME diterapkan guna mengidentifikasi kata-kata yang paling berpengaruh dalam setiap keputusan klasifikasi sehingga hasil penelitian dapat dipertanggungjawabkan secara ilmiah.

2. METODE

Penelitian ini merancang dua skenario paralel yang berjalan secara bersamaan skenario tanpa augmentasi dan dengan augmentasi agar pengaruh augmentasi terhadap kinerja model dapat diukur

Vol. 6, No. 1 - Juni 2026, Hal. 51-64
DOI: 10.59395/jitp.v6i1.197

secara langsung dan terukur. Keseluruhan alur kerja penelitian tersusun atas lima tahap utama sebagaimana diilustrasikan pada Gambar 1.

Gambar 1. Alur Penelitian

1.1 Pengumpulan dan Preprocessing Data

Pengumpulan data dilaksanakan dari halaman resmi aplikasi Roblox di Google Play Store dengan menggunakan pustaka google-play-scraper yang dikonfigurasi dengan parameter lang='id' dan country='id', sebagaimana direkomendasikan oleh Yasin et al. (2024) untuk pengumpulan ulasan dari Google Play Store. Proses scraping berhasil menghimpun 6.000 ulasan mentah sebagai bahan baku awal. Tahapan preprocessing terdiri dari delapan langkah berurutan yang diterapkan secara sistematis: (1) konversi seluruh karakter menjadi huruf kecil, (2) penghapusan tautan atau URL, (3) eliminasi emoji dan karakter non-ASCII, (4) pembuangan semua karakter numerik, (5) penghilangan simbol yang bukan merupakan huruf, (6) pemotongan pengulangan karakter yang berlebihan, (7) normalisasi kata tidak baku menggunakan kamus yang berisi 97 entri hasil kurasi secara manual, dan (8) normalisasi spasi antar kata.

Proses penyaringan data diterapkan dengan empat kriteria seleksi: ulasan yang tersusun dari kurang dari 5 kata langsung dibuang, data duplikat dieliminasi, ulasan yang melampaui 100 kata disisihkan, dan deteksi bahasa menggunakan pustaka langdetect memastikan hanya ulasan dalam bahasa Indonesia yang lolos seleksi. Dari total 6.000 ulasan awal, serangkaian proses tersebut menghasilkan 3.631 ulasan yang bersih dan siap digunakan (ekuivalen dengan 60,5% dari keseluruhan data awal), yang kemudian menjadi 3.619 ulasan setelah tahap pelabelan mengeliminasi 12 ulasan yang tidak relevan dengan aspek mana pun. Tabel 1 menyajikan contoh ulasan yang berhasil dikumpulkan dan lolos proses penyaringan.

Vol. 6, No. 1 - Juni 2026, Hal. 51-64
DOI: 10.59395/jitp.v6i1.197

Tabel 1. Dataset

Nama User	Bintang	Tanggal	Ulasan
Sanur Oke	5	25/04/2026	seru banget
Janitra Dwi	5	25/04/2026	Game ini sangat bagus aku suka dengan gamenya karena ad rp(brookheaven), menara, tower, story(dll) bisa mabar juga rekomend buat kalian jdi mainin game nya sekarang jugaa >Loveroblox<
Sukianto	1	25/04/2026	tidak seimbangnya penyesuaian harga robux di item marketplace, untuk akun yang lama rata rata lebih mahal. sedangkan akun akun yang baru dibuat rata rata diberi harga dibawahnya.
syaa corbella	1	25/04/2026	ga habis pikir sama ni Roblox, SETIAP MAU VERIFIKASI GA BISA TERUS... harga itemnya pilih kasih, celana klasik gua di hapus tapi robux ga dibalikin.
Sabrina Anugrah	5	25/04/2026	gem nya bagus dan seru untuk di mainkan dan aku menyukainya gem Roblox dan kalo di situ aku paling suka main brokheven

1.2 Pelabelan Data

Setiap ulasan dalam dataset diberikan label sentimen untuk keempat aspek secara simultan Teknis, Sistem, Konten, dan Sosial dengan tiga pilihan polaritas yang tersedia: negatif, netral, atau positif (Wilie et al., 2020). Proses pelabelan menerapkan pendekatan anotasi hibrida yang mengombinasikan otomatisasi berbasis LLM dengan verifikasi manual, sejalan dengan temuan (Pangestu et al., 2025) yang membuktikan bahwa metode ini mampu memproses volume data yang besar dengan mempertahankan konsistensi yang terjaga dengan baik. Model Llama-4-Scout-17B diakses melalui API Groq dalam kelompok sepuluh ulasan per permintaan. Dari 3.631 ulasan yang tersedia, sebanyak 3.180 ulasan dengan tingkat kepercayaan tinggi dan sedang langsung diterima, sedangkan 444 ulasan berconfidence rendah diverifikasi secara manual oleh peneliti sendiri sebagai satu-satunya annotator manual dalam penelitian ini, dan 12 ulasan yang tidak relevan dieliminasi, menghasilkan dataset final sebanyak 3.619 entri berlabel. (R. Nugroho et al., 2025) menerapkan strategi anotasi berbantuan model bahasa yang serupa pada ulasan Mobile JKN dan memperoleh konsistensi pelabelan yang tinggi.

Konsistensi hasil anotasi divalidasi melalui pengujian intra-rater reliability menggunakan Cohen's Kappa. Nilai rata-rata kappa yang diperoleh sebesar 0,378 (termasuk dalam kategori Sedang) dengan akurasi anotasi rata-rata mencapai 88,2%. Nilai kappa yang relatif rendah pada aspek Teknis dan Sistem turut dipengaruhi oleh dominasi kelas negatif dalam subset yang diuji, sebagaimana dibuktikan oleh akurasi aspek Sistem yang justru tertinggi (96,8%) meskipun demikian, nilai kappa pada kategori Sedang ini tetap diakui sebagai keterbatasan dalam konsistensi pelabelan yang perlu menjadi perhatian pada penelitian selanjutnya. Tabel 2 menyajikan contoh-contoh hasil pelabelan yang dihasilkan.

Tabel 2. Contoh Hasil Pelabelan Sentimen Per Aspek pada Dataset

Ulasan Bersih	Teknis	Sistem	Konten	Sosial	Keyakinan
game ini sangat bagus aku suka dengan gamenya karena ad rp brookheaven menara tower story bisa mabar juga rekomend buat kalian loveroblox	positif	-	positif	positif	high
game nya bagus dan okelah cuman kadang itu suka ngelag gitu kayak pas main agak keganggu sedikit it s okay	negatif	-	positif	-	medium
hapus saja ini aplikasi game anak anak	-	-	negatif	-	low

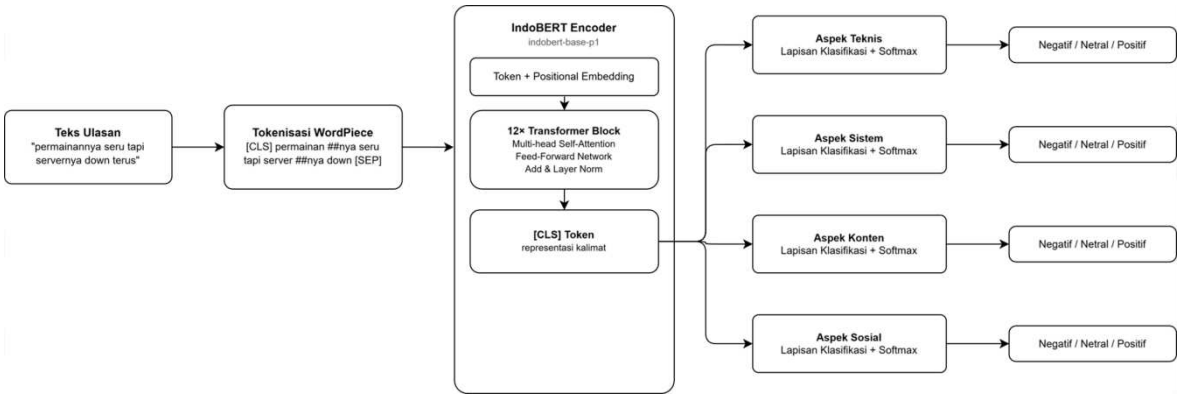
Vol. 6, No. 1 - Juni 2026, Hal. 51-64
DOI: 10.59395/jitp.v6i1.197

1.3 Augmentasi Data

Augmentasi difokuskan secara spesifik pada kelas netral yang menempati posisi sebagai kelas minoritas di seluruh aspek. Pendekatan berbasis LLM dipilih karena terbukti mampu menghasilkan teks sintetis dengan kualitas yang lebih koheren dan keragaman linguistik yang lebih baik dibandingkan teknik konvensional lainnya Zidan et al. (2026). Sebanyak 550 ulasan sintetis berhasil dibangkitkan menggunakan model Llama 3.3 70B melalui OpenRouter API dengan distribusi berikut: 150 untuk aspek Teknis, 150 untuk aspek Sistem, 100 untuk aspek Konten, dan 150 untuk aspek Sosial. Data sintetis yang dihasilkan hanya disisipkan ke dalam data training data validasi dan data pengujian tidak dimodifikasi sama sekali guna menjaga integritas proses evaluasi.

1.4 Fine-Tuning IndoBERT

Proses fine-tuning dilaksanakan secara terpisah untuk setiap aspek, menggunakan model dasar indobenchmark/indobert-base-p1 yang diunduh dari Hugging Face (Lin et al., 2022) dengan lapisan klasifikasi tiga kelas yang ditambahkan di atasnya, mengikuti pendekatan fine-tuning BERT untuk ulasan aplikasi mobile Indonesia sebagaimana dilakukan (K. S. Nugroho et al., 2020). Arsitektur model yang digunakan dapat diamati pada Gambar 2. Pendekatan serupa telah diterapkan oleh Jayadianti et al. (2022) yang mengombinasikan IndoBERT dengan R-CNN untuk analisis sentimen ulasan berbahasa Indonesia dan memperoleh akurasi 95,16%.



Gambar 2. Arsitektur Model IndoBERT untuk Klasifikasi Sentimen Berbasis Aspek

Pembagian dataset dilakukan dengan stratified split 80:10:10 (training:validasi:testing) untuk memastikan proporsi setiap kelas terjaga secara merata di ketiga subset, sebagaimana direkomendasikan (Pranatawijaya et al., 2024) dalam konteks dataset ABSA yang tidak seimbang. Pangestu et al. (2025) mengonfirmasi bahwa pembagian data dengan stratified split pada tugas ABSA multikelas berbahasa Indonesia menghasilkan distribusi kelas yang lebih stabil dan evaluasi yang lebih representatif. Konfigurasi hyperparameter yang diterapkan meliputi: maksimum 10 epoch, ukuran batch 16 (training) dan 32 (evaluasi), learning rate 2e-5, warmup ratio 0,1, dan weight decay 0,01. Ketidakseimbangan kelas ditangani melalui pembobotan otomatis menggunakan compute_class_weight='balanced' dari pustaka scikit-learn. Mekanisme early stopping dengan patience 3 epoch diterapkan untuk memantau F1-Macro pada data validasi dan menghentikan pelatihan secara otomatis jika tidak ada peningkatan. Simanjuntak et al. (2024) menunjukkan bahwa penyesuaian hyperparameter IndoBERT secara sistematik memberikan dampak signifikan terhadap performa klasifikasi teks berbahasa Indonesia.

1.5 Fine-Tuning XLM-RoBERTa sebagai Baseline

Sebagai model baseline pembanding, eksperimen paralel dilaksanakan menggunakan XLM-RoBERTa (xlm-roberta-base) yang diakses melalui Hugging Face (Wilie et al., 2020). Model ini dipilih sebagai representasi transformer multibahasa modern yang telah terbukti kompetitif pada berbagai tugas NLP lintas bahasa. Konfigurasi fine-tuning yang diterapkan sepenuhnya identik dengan IndoBERT, dan eksperimen dijalankan pada kedua skenario baik tanpa maupun dengan augmentasi sehingga perbandingan berlangsung secara langsung dan adil.

Vol. 6, No. 1 - Juni 2026, Hal. 51-64
DOI: 10.59395/jitp.v6i1.197

1.6 Evaluasi Model

Evaluasi kinerja model menggunakan F1-Macro sebagai metrik primer, mengingat metrik ini memberikan bobot yang setara kepada setiap kelas tanpa dipengaruhi oleh dominasi kelas mayoritas sebuah pilihan yang sangat tepat untuk konteks dataset yang tidak seimbang. Ambang nilai F1-Macro sebesar 0,80 ditetapkan sebagai standar kinerja yang dianggap baik, mengacu pada penelitian ABSA IndoBERT sebelumnya (Rafianto et al., 2026). Metrik pelengkap yang turut dihitung mencakup accuracy, F1-Weighted, serta precision dan recall untuk masing-masing kelas. Adi et al. (2024) juga menegaskan bahwa pemilihan metrik F1-Macro sangat krusial pada dataset tidak seimbang agar evaluasi tidak bias terhadap kelas mayoritas. Linawati et al. (2020) juga menerapkan Confusion Matrix untuk mengevaluasi model klasifikasi Random Forest dan C4.5 dalam prediksi prestasi akademik mahasiswa dan memperoleh akurasi tertinggi 92,4% pada metode Random Forest. Perbandingan kinerja antara IndoBERT dan XLM-RoBERTa dilakukan pada kondisi yang sepenuhnya identik untuk memastikan validitas komparatif.

3. HASIL DAN PEMBAHASAN

1.7 Performa Model Tanpa Augmentasi

Tabel 3 merangkum hasil evaluasi pada kondisi tanpa augmentasi. Kontradiksi yang sangat mencolok terlihat pada data ini: accuracy rata-rata mencapai angka 91,44% namun F1-Macro rata-rata hanya berada di 0,5739. Kesenjangan yang signifikan ini merupakan tanda klasik dari bias model model cenderung mengandalkan prediksi kelas mayoritas secara dominan dan gagal dalam mengenali kelas yang memiliki jumlah sampel lebih sedikit.

Tabel 3. Metrik Evaluasi Model IndoBERT Tanpa Augmentasi

Aspek	Accuracy	F1-Macro	F1-Weighted	F1-Neg	F1-Net	F1-Pos
Teknis	0,9000	0,7001	0,8943	0,94	0,31	0,86
Sistem	0,9340	0,4330	0,9175	0,97	0,00	0,33
Konten	0,8776	0,5289	0,8741	0,65	0,00	0,93
Sosial	0,9459	0,6335	0,9373	0,96	0,00	0,94
Rata-rata	0,9144	0,5739	0,9058	0,88	0,08	0,77

Akar permasalahan menjadi gamblang ketika kinerja per kelas ditelaah lebih dalam: F1-score kelas netral pada aspek Sistem, Konten, dan Sosial masing-masing mencatat nilai 0,00 yang berarti model benar-benar tidak mampu mengidentifikasi satu pun ulasan netral pada ketiga aspek tersebut. Penyebab utamanya adalah sedikitnya jumlah sampel netral dalam data training (19 sampel untuk aspek Sistem, 15 untuk aspek Sosial). Kondisi ini menegaskan bahwa penanganan distribusi kelas netral merupakan prasyarat yang mutlak diperlukan agar model dapat berfungsi secara menyeluruh dan efektif.

1.8 Performa Model Dengan Augmentasi

Setelah 550 ulasan sintetis kelas netral disisipkan ke dalam data training, hasilnya menunjukkan perubahan yang sangat berarti. Tabel 4 memperlihatkan bahwa untuk pertama kalinya, seluruh aspek secara bersamaan berhasil melampaui nilai F1-Macro 0,80.

Tabel 4. Metrik Evaluasi Model IndoBERT Dengan Augmentasi

Aspek	Accuracy	F1-Macro	F1-Weighted	F1-Neg	F1-Net	F1-Pos
Teknis	0,8884	0,8319	0,8884	0,93	0,74	0,83
Sistem	0,9421	0,8123	0,9404	0,97	0,87	0,60
Konten	0,9098	0,8124	0,9106	0,73	0,76	0,95
Sosial	0,9603	0,9635	0,9601	0,97	1,00	0,92
Rata-rata	0,9252	0,8550	0,9249	0,90	0,84	0,83

Vol. 6, No. 1 - Juni 2026, Hal. 51-64
DOI: 10.59395/jitp.v6i1.197

Kelas netral yang sebelumnya tidak terdeteksi pada tiga aspek kini memperoleh F1-score bermakna yaitu 0,74 (Teknis), 0,87 (Sistem), 0,76 (Konten), dan 1,00 untuk aspek Sosial. Aspek Sosial tampil sebagai yang terbaik secara keseluruhan dengan F1-Macro mencapai 0,9635, di mana seluruh 17 sampel netral berhasil diklasifikasikan dengan benar. Pencapaian ini memvalidasi bahwa teks sintetis yang dibangkitkan oleh LLM mengandung pola kebahasaan yang cukup representatif dan relevan. (Yulianti & Nissa, 2024) turut membuktikan kemampuan IndoBERT dalam menangkap nuansa sentimen ulasan berbahasa Indonesia (Zidan et al., 2026).

1.9 Dampak Augmentasi: Perbandingan Dua Skenario

Tabel 5 merangkum selisih F1-Macro yang dihasilkan dari kedua skenario perbandingan. Pola yang muncul bersifat konsisten: semakin parah ketimpangan kelas pada kondisi awal, semakin besar lonjakan performa yang diperoleh melalui augmentasi.

Tabel 5. Perbandingan F1-Macro Sebelum dan Sesudah Augmentasi

Aspek	F1-Macro Tanpa Aug	F1-Macro Dg Aug	Delta	Keterangan
Teknis	0,7001	0,8319	+0,1318	Meningkat
Sistem	0,4330	0,8123	+0,3793	Meningkat signifikan
Konten	0,5289	0,8124	+0,2835	Meningkat signifikan
Sosial	0,6335	0,9635	+0,3300	Meningkat signifikan
Rata-rata	0,5739	0,8550	+0,2811	Meningkat

Aspek Sistem mencatat kenaikan terbesar (+0,3793), diikuti aspek Sosial (+0,3300) keduanya merupakan aspek yang sebelumnya hanya memiliki 19 dan 15 sampel netral dalam data training. Secara agregat, rata-rata F1-Macro bergerak dari 0,5739 ke 0,8550, setara dengan kenaikan relatif mendekati 49%.

1.10 Baseline Transformer Modern: XLM-RoBERTa

Guna memperkuat validitas temuan penelitian, dilakukan eksperimen tambahan menggunakan XLM-RoBERTa (xlm-roberta-base) sebagai representasi model transformer multibahasa modern. Eksperimen ini dilaksanakan pada kondisi yang sepenuhnya identik baik tanpa maupun dengan augmentasi sehingga perbandingan dengan IndoBERT berlangsung secara langsung dan adil.

Tabel 6. Hasil Evaluasi Model XLM Tanpa Augmentasi

Aspek	Accuracy	F1-Macro	F1-Neg	F1-Net	F1-Pos
Teknis	0,8750	0,6307	0,9236	0,1250	0,8434
Sistem	0,9340	0,3220	0,9659	0,0000	0,0000
Konten	0,8898	0,5466	0,6939	0,0000	0,9459
Sosial	0,9640	0,6477	0,9744	0,0000	0,9688
Rata-rata	0,9157	0,5367	0,8895	0,0313	0,6895

Tabel 7. Hasil Evaluasi Model XLM dengan Augmentasi

Aspek	Accuracy	F1-Macro	F1-Neg	F1-Net	F1-Pos
Teknis	0,8884	0,8331	0,9246	0,7391	0,8354
Sistem	0,9256	0,6077	0,9565	0,8667	0,0000
Konten	0,9137	0,8174	0,7458	0,7586	0,9479
Sosial	0,9524	0,9558	0,9610	1,0000	0,9062
Rata-rata	0,9200	0,8035	0,8970	0,8411	0,6724

Vol. 6, No. 1 - Juni 2026, Hal. 51-64
DOI: 10.59395/jitp.v6i1.197

Setelah augmentasi diterapkan, rata-rata F1-Macro XLM-RoBERTa meningkat menjadi 0,8035. Jika dibandingkan secara langsung, IndoBERT dengan augmentasi tetap mempertahankan keunggulannya (0,8550 vs 0,8035). Perbedaan ini secara nyata mencerminkan keunggulan model yang sejak tahap pelatihan awal difokuskan secara khusus pada teks berbahasa Indonesia, terutama dalam menangkap nuansa bahasa gaul dan berbagai ekspresi informal yang mendominasi ulasan game. Dhendra & Gayuh Utomo (2025) memperkuat temuan ini melalui benchmarking komprehensif yang membuktikan bahwa IndoBERT secara konsisten mengungguli model multibahasa pada tugas klasifikasi sentimen layanan berbahasa Indonesia (Lin et al., 2022).

1.11 Distribusi Sentimen per Aspek

Distribusi sentimen yang terpetakan dalam dataset berlabel menghasilkan informasi yang dapat langsung dimanfaatkan oleh tim pengembang Roblox, sebagaimana tersaji pada Tabel 8 berikut.

Tabel 8. Distribusi Sentimen pada Dataset Berlabel per Aspek

Aspek	Negatif	Netral	Positif	Total
Teknis	1.514 (75,8%)	77 (3,9%)	406 (20,3%)	1.997
Sistem	994 (93,8%)	19 (1,8%)	47 (4,4%)	1.060
Konten	259 (10,6%)	77 (3,1%)	2.112 (86,3%)	2.448
Sosial	758 (68,8%)	15 (1,4%)	328 (29,8%)	1.101

Sentimen negatif 93,8% pada aspek Sistem mencerminkan ketidakpuasan pengguna terhadap kebijakan verifikasi usia dan monetisasi Robux. Aspek Teknis juga didominasi keluhan (75,8%) terkait lag, crash, dan ketidakstabilan koneksi. Sebaliknya, aspek Konten menjadi kekuatan platform dengan 86,3% ulasan positif. Aspek Sistem menjadi prioritas paling mendesak bagi pengembang, sementara aspek Konten perlu dipertahankan.

1.12 Ablation Study

Untuk memahami seberapa besar kontribusi masing-masing komponen terhadap kinerja akhir model, penelitian ini merancang empat skenario eksperimen yang berbeda sebagaimana dirangkum pada Tabel 9 berikut.

Tabel 9. Ablation Study Kontribusi Komponen Model

No	Konfigurasi	Model	Aug	CW	F1 Teknis	F1 Sistem	F1 Konten	F1 Sosial	Rata-rata
1	Ablasi 1	IndoBERT	Ya	Tidak	0,8263	0,8362	0,8090	0,9699	0,8604
2	Ablasi 2	IndoBERT	Tidak	Ya	0,7001	0,4330	0,5289	0,6335	0,5739
3	Ablasi 3	XLM-R	Ya	Ya	0,8420	0,6077	0,7916	0,9549	0,7991
4	Full Model	IndoBERT	Ya	Ya	0,8319	0,8123	0,8124	0,9635	0,8550

Dampak paling dramatis tampak ketika komponen augmentasi dieliminasi. Ablasi 2 hanya sanggup menghasilkan rata-rata F1-Macro sebesar 0,5739 terpaut 0,2811 poin di bawah full model. Di antara seluruh komponen yang diuji, augmentasi data terbukti sebagai elemen yang paling tidak dapat ditinggalkan (Zidan et al., 2026). Ablasi 3 dengan XLM-RoBERTa menghasilkan rata-rata F1-Macro 0,7991, sejalan dengan argumentasi (Wilie et al., 2020) bahwa model yang dilatih secara spesifik pada korpus bahasa Indonesia jauh lebih efektif dalam menangani bahasa informal, singkatan, dan gejala alih kode. Merdiansah & Ridha (2024) mengonfirmasi keunggulan IndoBERT untuk analisis sentimen teks informal berbahasa Indonesia.

1.13 Uji Signifikansi Statistik (McNemar's Test)

Untuk memastikan bahwa perbedaan kinerja antar konfigurasi bukan sekadar variasi acak yang terjadi secara kebetulan, penelitian ini menerapkan uji signifikansi statistik menggunakan

Vol. 6, No. 1 - Juni 2026, Hal. 51-64
DOI: 10.59395/jitp.v6i1.197

McNemar's Test pada tingkat kepercayaan 95% (α = 0,05). Tabel 10 menyajikan hasil pengujian secara lengkap.

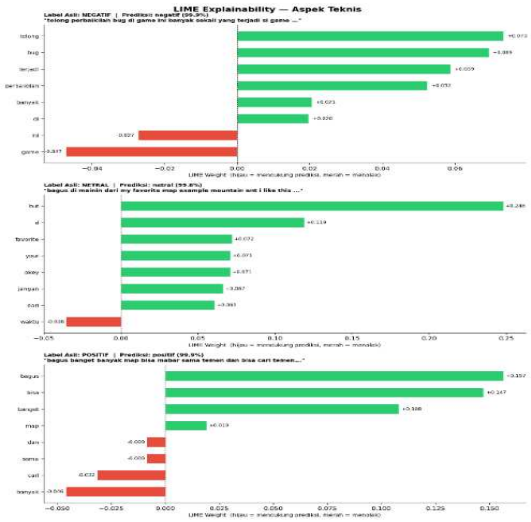
Tabel 10. Hasil McNemar's Test Antar Konfigurasi

Perbandingan	Aspek	n	p-value	Metode	Signifikan?	Full✓ Abl✗	Full✗ Abl✓
Full vs Ablasi 1 (efek Class Weight)	Teknis	215	1,0000	Exact	Tidak	10	9
	Sistem	121	1,0000	Exact	Tidak	0	1
	Konten	255	1,0000	Exact	Tidak	3	3
	Sosial	126	1,0000	Exact	Tidak	1	2
Full vs Ablasi 2 (efek Augmentasi)	Teknis	200	0,0000	Chi²	Ya	75	10
	Sistem	106	0,0010	Exact	Ya	14	1
	Konten	245	0,0000	Chi²	Ya	52	11
	Sosial	111	0,0000	Chi²	Ya	53	3
Full vs Ablasi 3 (efek Arsitektur)	Teknis	215	1,0000	Exact	Tidak	9	10
	Sistem	121	0,6250	Exact	Tidak	3	1
	Konten	255	1,0000	Exact	Tidak	7	8
	Sosial	126	1,0000	Exact	Tidak	3	2

Hanya perbandingan Full vs Ablasi 2 yang secara konsisten menghasilkan p < 0,05 pada seluruh aspek, membuktikan peningkatan kinerja augmentasi memiliki landasan statistik yang kuat. Efek pembobotan kelas dan arsitektur model tidak mencapai signifikansi statistik, meskipun tetap berkontribusi positif secara praktis.

1.14 Interpretabilitas Model: Analisis LIME

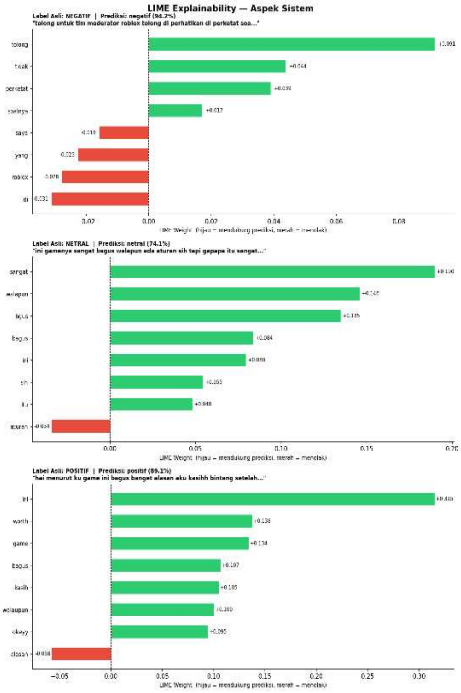
Guna menelusuri lebih jauh alasan di balik setiap prediksi yang dihasilkan model, penelitian ini melengkapi evaluasi kuantitatif dengan analisis interpretabilitas menggunakan LIME (Local Interpretable Model-agnostic Explanations) (Perikos & Diamantopoulos, 2024). LIME bekerja dengan menjelaskan mengapa model menghasilkan prediksi tertentu pada level kalimat individual, melalui identifikasi kata-kata yang paling mendorong atau menolak suatu label klasifikasi.



Gambar 3. LIME Explainability Aspek Teknis

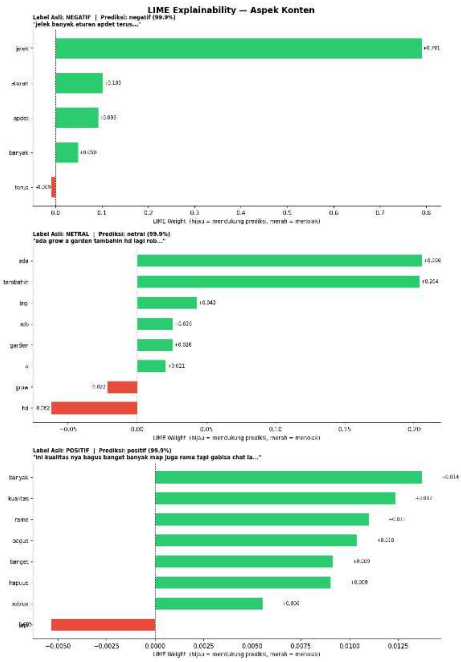
Vol. 6, No. 1 - Juni 2026, Hal. 51-64
DOI: 10.59395/jitp.v6i1.197

Pada aspek Teknis (Gambar 3), kata-kata seperti 'bug', 'tolong', dan 'perbaikalah' berperan sebagai sinyal negatif yang sangat kuat, sementara 'bagus', 'bisa', dan 'banget' mendominasi prediksi positif. Hal yang menarik ditemukan pada kelas netral kata-kata berbahasa Inggris seperti 'but' dan 'favorite' tampil sebagai fitur paling berpengaruh, mengindikasikan bahwa ulasan yang mengandung fenomena alih kode cenderung memiliki orientasi netral.



Gambar 4. LIME Explainability Aspek Sistem

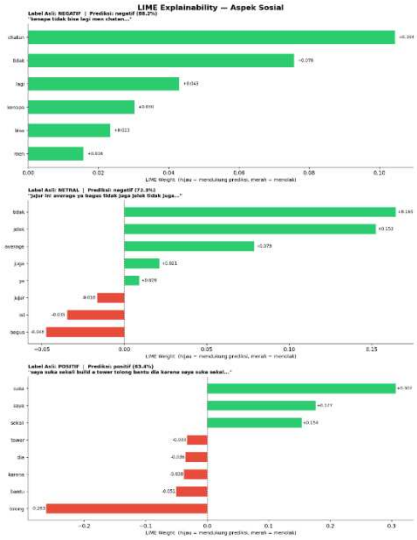
Pada aspek Sistem (Gambar 4), kata 'tolong' muncul sebagai sinyal negatif terkuat dengan bobot +0,091. Untuk kelas positif, kata 'ini' (+0,315) yang secara leksikal tidak bermuatan sentimen apapun justru tampil sebagai fitur paling dominan sebuah indikasi bahwa model mampu menangkap konteks keseluruhan kalimat secara holistik.



Gambar 5. LIME Explainability Aspek Konten

Vol. 6, No. 1 - Juni 2026, Hal. 51-64
DOI: 10.59395/jitp.v6i1.197

Pada aspek Konten (Gambar 5), kata 'jelek' memperoleh bobot +0,791 pada kelas negatif jauh melampaui kontribusi fitur-fitur lainnya. Pada kelas netral, kata 'ada' (+0,206) dan 'tambahin' (+0,204) menjadi pendorong terkuat karena ulasan yang berisi permintaan penambahan fitur secara alami memiliki orientasi yang netral.

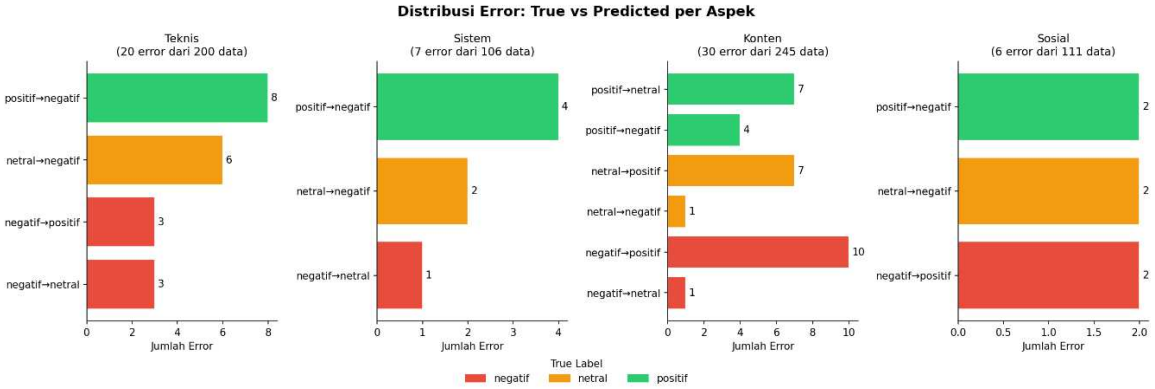


Gambar 6. LIME Explainability Aspek Sosial

Pada aspek Sosial (Gambar 6), kata 'chatan' (+0,104) dan 'tidak' (+0,076) berfungsi sebagai penanda negatif yang dominan. Temuan paling menarik pada kelas positif adalah kata 'tolong' yang mencatat bobot negatif yang cukup besar (−0,263), yang sebagian menjelaskan pola False Negative yang diamati pada aspek Sosial (Wilie et al., 2020).

1.15 Analisis Error

Tabel 11 merangkum pola kesalahan yang paling sering terjadi per aspek, sementara distribusi keseluruhan error divisualisasikan pada Gambar 7.



Gambar 7. Distribusi Error

Tabel 11. Ringkasan Pola Kesalahan

Aspek	Total Error	Error Rate	Pola Dominan	Jumlah
Teknis	20/200	10,0%	positif → negatif	8 kasus
Sistem	7/106	6,6%	positif → negatif	4 kasus
Konten	30/245	12,2%	negatif → positif	10 kasus
Sosial	6/111	5,4%	Merata (3 pola)	2 kasus/pola

Vol. 6, No. 1 - Juni 2026, Hal. 51-64
DOI: 10.59395/jitp.v6i1.197

Pada seluruh aspek, model memperlihatkan kecenderungan untuk terdorong ke arah kelas yang memiliki keterwakilan paling besar dalam data. Meskipun augmentasi berhasil memperbaiki keterwakilan kelas netral secara signifikan, kelas minoritas lain seperti positif pada aspek Sistem dan negatif pada aspek Konten masih belum mendapat penanganan yang memadai (Zidan et al., 2026). Perluasan cakupan augmentasi ke seluruh kelas yang kurang terwakili dapat menjadi langkah perbaikan paling konkret dan terukur untuk penelitian selanjutnya. Fadilah & Priyanta (2022) memperlihatkan bahwa augmentasi data berbasis generasi teks terbukti efektif mengatasi keterbatasan sampel pada tugas klasifikasi berbahasa Indonesia.

1.16 Perbandingan dengan Penelitian Sebelumnya
1.17

Tabel 12. Perbandingan dengan Penelitian Terdahulu

Peneliti	Metode	Domain	Bahasa	Kelas	Akurasi	F1
(Setiawan et al., 2025)	SVM + TF-IDF	Ulasan Roblox	Indonesia	2	82,51%	-
(Rafianto et al., 2026)	LDA + SVM	Ulasan Roblox	Indonesia	2	85,22%	86%
(Bachtiar et al., 2026)	LR + TF-IDF	Ulasan Free Fire	Indonesia	2	81,63%	87,91%
(Jazuli et al., 2025)	ABSA+IndoBERT	Ulasan Mahasiswa	Indonesia	3	97,9%	97,4%
(Alfiana et al., 2026)	ABSA+IndoBERT	Ulasan KAI	Indonesia	2	92,8%	78,5%
(Rahman et al., 2024)	LDA + BERT	Ulasan Gojek	Inggris	3	97,27%	96,67%
Penelitian ini	ABSA+IndoBERT+LLM	Ulasan Roblox	Indonesia	3	92,52%	85,50%

Penelitian ini berhasil meraih F1-Macro sebesar 85,50% dengan cakupan tiga kelas sentimen melampaui Alfiana et al. (2026) yang menerapkan ABSA IndoBERT meskipun hanya pada dua kelas sentimen. Selisih dengan Jazuli et al. (2025) dan Rahman et al. (2024) yang mencatat F1-score di atas 96% lebih tepat dipahami sebagai perbedaan karakteristik data yang dihadapi: ulasan game Roblox sarat dengan sarkasme, bahasa gaul, dan fenomena alih kode yang menjadikannya jauh lebih menantang untuk diklasifikasikan secara akurat. Putri & Ardiansyah (2023) juga menunjukkan bahwa model berbasis BERT konsisten mengungguli pendekatan machine learning konvensional untuk analisis sentimen teks berbahasa Indonesia pada berbagai domain.

4. KESIMPULAN

Penelitian ini memvalidasi bahwa IndoBERT yang di-fine-tune per aspek, dikombinasikan dengan augmentasi data berbasis LLM, merupakan pendekatan yang sangat efektif untuk menjalankan ABSA pada ulasan game berbahasa Indonesia. Hasil ini selaras dengan temuan Kono et al. (2025) yang membuktikan bahwa fine-tuning IndoBERT secara domain-spesifik konsisten menghasilkan kinerja tertinggi dibandingkan metode klasifikasi lainnya. Tiga temuan pokok yang layak untuk mendapat perhatian khusus dapat diuraikan sebagai berikut.

Pertama, augmentasi data kelas netral berbasis LLM memberikan dampak yang sangat signifikan terhadap kinerja model. Rata-rata F1-Macro meningkat hampir 49% secara relatif dari 0,5739 menjadi 0,8550 dan seluruh aspek untuk pertama kalinya berhasil melampaui ambang nilai 0,80 secara bersamaan. Kelas netral yang sebelumnya tidak dapat terdeteksi sama sekali pada tiga aspek kini berhasil dikenali dengan F1-score yang bermakna. Signifikansi statistik dari temuan ini dikonfirmasi secara kuat melalui McNemar's Test ($p < 0,05$) pada seluruh aspek.

Kedua, pemetaan distribusi sentimen per aspek menghasilkan wawasan yang dapat langsung dioperasionalkan oleh tim pengembang. Aspek Sistem merupakan prioritas paling kritis yang

Vol. 6, No. 1 - Juni 2026, Hal. 51-64
DOI: 10.59395/jitp.v6i1.197

memerlukan perhatian segera (93,8% negatif), khususnya terkait kebijakan verifikasi dan mekanisme monetisasi, sedangkan aspek Konten merupakan kekuatan inti platform (86,3% positif) yang wajib dipertahankan dan terus dikembangkan ke depannya.

Ketiga, IndoBERT terbukti kompetitif dan unggul untuk domain ulasan game informal berbahasa Indonesia dengan tiga kelas sentimen dan empat aspek sekaligus, melampaui F1-score penelitian ABSA IndoBERT sebelumnya pada konteks yang lebih terstruktur sebesar 78,5% (Alfiana et al., 2026), dan mengungguli baseline XLM-RoBERTa (0,8035). Hierarki kepentingan komponen yang berhasil diidentifikasi augmentasi > arsitektur > pembobotan kelas dapat dijadikan panduan praktis yang berharga bagi peneliti yang menghadapi tantangan serupa di masa mendatang. Penelitian lanjutan dapat mengeksplorasi arsitektur yang lebih besar (IndoBERT-large, XLM-R-large), memperluas cakupan augmentasi ke seluruh kelas yang kurang terwakili, serta merancang sistem pemantauan sentimen secara real-time yang responsif terhadap setiap pembaruan aplikasi (Reddy et al., 2025). Penelitian ini juga memiliki keterbatasan pada proses verifikasi manual yang hanya dilakukan oleh satu annotator (peneliti sendiri) serta nilai Cohen's Kappa yang masih berada pada kategori Sedang ($\kappa = 0{,}378$), sehingga penelitian lanjutan disarankan melibatkan annotator tambahan untuk memungkinkan pengujian inter-rater reliability yang lebih komprehensif dan meningkatkan konsistensi pelabelan.

DAFTAR PUSTAKA

Adi, S., Mola, S., Luttu, Y. C., & Rumlaklak, D. N. (2024). *Perbandingan Metode Machine Learning dalam Analisis Sentimen Komentar Pengguna Aplikasi InDriver pada Dataset Tidak Seimbang*. *03*. https://doi.org/10.21456/vol14iss3pp247-255

Alfiana, H. N., Doewes, A., & Widoyono, B. (2026). *Aspect-Based Sentiment Analysis of Access by KAI Application Reviews Using IndoBERT for Multi-Label Classification Tasks*. *7*(1), 286–306.

Bachtiar, A., Hawali, M. J., Simbolon, N. C., Maulana, D. A., Anggraini, R. A., & Kunci, K. (2026). *ANALISIS SENTIMEN ULASAN FREE FIRE MENGGUNAKAN NAIVE BAYES LOGISTIC REGRESSION Abstraksi Keywords : Pendahuluan Tinjauan Pustaka*. *7*(2).

Bahri, C. A., Suadaa, L. H., Studi, P., Statistik, K., Regression, L., Review, G. M., & Learning, T. (2023). *Aspect-Based Sentiment Analysis in Bromo Tengger Semeru National Park Indonesia Based on Google Maps User Reviews*. *17*(1), 79–90.

Dhendra, & Gayuh Utomo, V. (2025). Benchmarking IndoBERT and Transformer Models for Sentiment Classification on Indonesian E-Government Service Reviews. *Jurnal Transformatika*, *23*(1), 86–95. https://doi.org/10.26623/transformatika.v23i1.12095

Fadilah, N., & Priyanta, S. (2022). *Automatic Essay Scoring Using Data Augmentation in Bahasa Indonesia*. *16*(4), 401–410.

Jayadianti, H., Kaswidjanti, W., Utomo, A. T., Saifullah, S., Dwiyanto, F. A., & Drezewski, R. (2022). Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN. *ILKOM Jurnal Ilmiah*, *14*(3), 348–354. https://doi.org/10.33096/ilkom.v14i3.1505.348-354

Jazuli, A., Widowati, & Kusumaningrum, R. (2025). Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback. *Applied Sciences*, *15*(1), 172. https://doi.org/10.3390/app15010172

Kono, M. F., Fajri, I. N., & Pristyanto, Y. (2025). Public Sentiment Analysis on Corruption Issues in Indonesia Using IndoBERT Fine-Tuning, Logistic Regression, and Linear SVM. *Journal of Applied Informatics and Computing*, *9*(5), 2616–2628. https://doi.org/10.30871/jaic.v9i5.10537

Lin, T., Wang, Y., Liu, X., & Qiu, X. (2022). A survey of transformers. *AI Open*, *3*(1), 111–132. https://doi.org/10.1016/j.aiopen.2022.10.001

Linawati, S., Nurdiani, S., Handayani, K., & Latifah, L. (2020). Prediksi Prestasi Akademik Mahasiswa Menggunakan Algoritma Random Forest Dan C4.5. *Jurnal Khatulistiwa Informatika*, *8*(1), 47–52. https://doi.org/10.31294/jki.v8i1.7827

Maulana, R., Rahayuningsih, P. A., Irmayani, W., Saputra, D., & Jayanti, W. E. (2020). Improved Accuracy of Sentiment Analysis Movie Review Using Support Vector Machine Based Information Gain. *Journal of Physics: Conference Series*, *1641*(1). https://doi.org/10.1088/1742-6596/1641/1/012060

Merdiansah, R., & Ridha, A. A. (2024). *Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT*. *7*, 221–228.

Vol. 6, No. 1 - Juni 2026, Hal. 51-64
DOI: 10.59395/jitp.v6i1.197

Nugroho, K. S., Sukmadewa, A. Y., Bachtiar, F. A., & Yudistira, N. (2020). *BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews*. 1–10.

Nugroho, R., Azka, N., Sayudha, W., & Graha, R. (2025). *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi) Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google*. *9*(June), 495–505.

Pangestu, A. S., Christian, E., & Lestari, A. (2025). *Aspect-Based Sentiment Analysis pada Ulasan Pengguna Aplikasi Mobile JKN Menggunakan Model Berbasis Transformer*. *5*, 10–12.

Perikos, I., & Diamantopoulos, A. (2024). Explainable Aspect-Based Sentiment Analysis Using Transformer Models. *Big Data and Cognitive Computing*, *8*(11). https://doi.org/10.3390/bdcc8110141

Pranatawijaya, V. H., Sari, N. N. K., Rahman, R. A., Christian, E., & Geges, S. (2024). Unveiling User Sentiment: Aspect-Based Analysis and Topic Modeling of Ride-Hailing and Google Play App Reviews. *Journal of Information Systems Engineering and Business Intelligence*, *10*(3), 328–339. https://doi.org/10.20473/jisebi.10.3.328-339

Putri, N. A. R., & Ardiansyah. (2023). Analisis Sentimen Terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa. *Jurnal Sains Dan Informatika*, *9*(2), 136–145. https://doi.org/10.34128/jsi.v9i2.649

Rafianto, G., Hannie, & Ma'sum, A. (2026). Analisis Sentimen Berbasis Topik Ulasan Pengguna Aplikasi Roblox Menggunakan Integrasi LDA dan SVM. *Jurnal Informatika Dan Teknik Elektro Terapan*, *14*(2). https://doi.org/10.23960/jitet.v14i2.9336

Rahman, R. A., Pranatawijaya, V. H., & Sari, N. N. K. (2024). Analisis Sentimen Berbasis Aspek pada Ulasan Aplikasi Gojek. *KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi*, *4*(1), 70–82.

Reddy, Y. A., Agarwal, S., Parashar, V., & Arora, A. (2025). *Visualizing Public Opinion on X: A Real-Time Sentiment Dashboard Using VADER and DistilBERT*. http://arxiv.org/abs/2504.15448

Setiawan, P., Zhakaria, A., Mandey, F. F., & Kurniawati, I. (2025). *Analisis Sentimen Pengguna Aplikasi Roblox Berdasarkan Ulasan Menggunakan Metode Machine Learning*. *7*(3), 452–460.

Simanjuntak, A., Lumbantoruan, R., Sianipar, K., Gultom, R., Simaremare, M., & Situmeang, S. (2024). *Studi dan Analisis Hyperparameter Tuning IndoBERT Dalam Pendeteksian Berita Palsu*. *13*, 60–67.

Sitorus, A. R., John, W., Ziraluho, P., & S, I. M. S. (2024). *ANALISIS SENTIMEN GAME MOBILE LEGENDS BERDASARKAN REVIEW PENGGUNA DI PLAYSTORE MENGGUNAKAN ALGORITMA NAIVE BAYES*. *4*(2), 108–113.

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 843–857. https://doi.org/10.18653/v1/2020.aacl-main.85

Yasin, A., Fatima, R., Ghazi, A. N., & Wei, Z. (2024). Python data odyssey: Mining user feedback from google play store. *Data in Brief*, *54*, 110499. https://doi.org/10.1016/j.dib.2024.110499

Yulianti, E., & Nissa, N. K. (2024). ABSA of Indonesian customer reviews using IndoBERT: single-sentence and sentence-pair classification approaches. *Bulletin of Electrical Engineering and Informatics*, *13*(5), 3579–3589. https://doi.org/10.11591/eei.v13i5.8032

Zhu, L., Xu, M., Bao, Y., Xu, Y., & Kong, X. (2022). Deep learning for aspect-based sentiment analysis: a review. *PeerJ Computer Science*, *8*. https://doi.org/10.7717/PEERJ-CS.1044

Zidan, M., Pinandito, A., & Rusydi, A. N. (2026). *Analisis Perbandingan Efisiensi Metode Augmentasi Data EDA , Back- Translation Dan LLM Untuk Klasifikasi Topik Berita Berbahasa Indonesia Pada Kondisi Class Imbalance*. *10*(1), 1–6.