KAMPUS AKADEMIK PUBLISING
Jurnal Sains Student Research (JSSR)
Vol.4, No.6 Desember 2026
e-ISSN: 3025-9851, p-ISSN : 3025-986X, Hal 1289-1303
DOI : https://doi.org/10.61722/jssr.v4i6.13347

Analisis Sentimen Menggunakan Model IndoBERT Pada Game Violence District Roblox

Zahra Meilani H[1*], Wina Witanti[2], Melina[3]
[1]Program Studi Teknik Informatika, Fakultas Sains Dan Informatika, Universitas Jenderal Achmad Yani
[2]Program Studi Teknik Informatika, Fakultas Sains Dan Informatika, Universitas Jenderal Achmad Yani
[3]Program Studi Teknik Informatika, Fakultas Sains Dan Informatika, Universitas Jenderal Achmad Yani
*Penulis Korespondensi: zahrameilanihermawan@gmail.com

Abstract. *User-generated content-based gaming platforms such as Roblox have experienced a significant increase in the number of users as well as the intensity of digital interaction. One of the fairly popular games is Violence District, which is often a topic of discussion on the TikTok social media platform. Player reviews are unstructured, subjective, and highly diverse, so an analytical method is needed to systematically understand player perceptions. This study aims to analyze the sentiment of TikTok users' comments on the game Violence District into three classes, namely positive, negative, and neutral, using the IndoBERT model through a fine-tuning approach. A total of 6,970 comments were collected using a web scraping technique through ExportTok, then processed through preprocessing stages, leaving 6,337 data points after 633 data points with empty text were removed, and a duplicate removal process removed 967 data points, resulting in 5,370 clean data points. Labeling was carried out semi-automatically using a lexicon-based approach with the InSet dictionary combined with a Violence District-specific domain dictionary, followed by manual validation, resulting in a class distribution with positive as the dominant class (2,236 data points, 41.64%), followed by negative (1,746 data points, 32.51%), and neutral (1,388 data points, 25.85%). The data was then split using stratified hold-out validation with an 80:10:10 ratio into 4,296 training data points, 537 validation data points, and 537 test data points. The IndoBERT model was fine-tuned with a configuration of learning rate 2e-5, batch size 16, AdamW optimizer (weight decay 0.01), a maximum of 10 epochs, and an early stopping mechanism (patience 3) to prevent overfitting. The model evaluation results on the 537 test data points show that the model achieved an accuracy of 75.61%, macro precision of 76.36%, macro recall of 72.51%, and macro F1-score of 73.33%. Based on the confusion matrix analysis, the model correctly predicted 200 out of 228 data points for the positive class, 127 out of 171 data points for the neutral class, and 81 out of 138 data points for the negative class, resulting in a total of 408 correct predictions out of 537 data points. The best performance was achieved by the positive class with an F1-score of 84.40%, followed by the neutral class with an F1-score of 76.53%, and the negative class with an F1-score of 60.46%. The largest classification errors occurred in the negative class, most of which were incorrectly predicted as positive (34 data points), followed by the neutral class incorrectly predicted as negative (32 data points), and the negative class incorrectly predicted as neutral (23 data points), indicating semantic ambiguity between negative comments and both positive and neutral sentiment expressions. The relatively lower performance on the negative class suggests that the characteristics of negative comments, which tend to overlap with positive and neutral sentiment expressions, require further optimization strategies, such as data augmentation or refinement of the labeling process for that class.*

Keywords: *Sentiment Analysis; IndoBERT; Fine-Tuning; Roblox; Violence District.*

Abstrak. Platform game berbasis *user-generated content* seperti Roblox mengalami peningkatan signifikan dalam jumlah pengguna serta intensitas interaksi digital. Salah satu game yang cukup populer adalah Violence District, yang sering menjadi topik diskusi pada media sosial TikTok. Ulasan pemain bersifat tidak terstruktur, subjektif, dan sangat beragam, sehingga diperlukan metode analitis untuk memahami persepsi pemain secara sistematis. Penelitian ini bertujuan menganalisis sentimen komentar pengguna

Naskah Masuk:01 November 2026; Revisi: 02 November 2026; Diterima: 03 November 2026; Terbit: 11 November 2026

TikTok terhadap game Violence District ke dalam tiga kelas, yaitu positif, negatif, dan netral, menggunakan model IndoBERT melalui pendekatan *fine-tuning*. Sebanyak 6.970 komentar dikumpulkan menggunakan teknik *web scraping* melalui ExportTok, kemudian diproses melalui tahapan *preprocessing* sehingga menyisakan 6.337 data setelah 633 data dengan teks kosong dihapus, dan proses *duplicate removal* menghapus 967 data sehingga diperoleh 5.370 data bersih. Pelabelan dilakukan secara semi-otomatis menggunakan pendekatan *lexicon-based* dengan kamus InSet yang dipadukan dengan kamus domain khusus Violence District, kemudian dilakukan validasi manual, menghasilkan distribusi kelas positif sebagai kelas dominan (2.236 data, 41,64%), diikuti negatif (1.746 data, 32,51%), dan netral (1.388 data, 25,85%). Data kemudian dibagi menggunakan *stratified hold-out validation* dengan proporsi 80:10:10 menjadi 4.296 data latih, 537 data validasi, dan 537 data uji. Model IndoBERT di-*fine-tune* dengan konfigurasi *learning rate* 2e-5, *batch size* 16, *optimizer* AdamW (*weight decay* 0,01), maksimum 10 epoch, dan mekanisme *early stopping* (*patience* 3) untuk mencegah *overfitting.* Hasil evaluasi model terhadap 537 data uji menunjukkan bahwa model berhasil mencapai akurasi 75,61%, *precision macro* 76,36%, *recall macro* 72,51%, dan *F1-score macro* 73,33%. Berdasarkan analisis *confusion matrix*, model memprediksi dengan benar sebanyak 200 dari 228 data kelas positif, 127 dari 171 data kelas netral, dan 81 dari 138 data kelas negatif, sehingga total prediksi benar sebanyak 408 dari 537 data. Performa terbaik dicapai pada kelas positif dengan F1-score 84,40%, diikuti kelas netral dengan F1-score 76,53%, dan kelas negatif dengan F1-score 60,46%. Kesalahan klasifikasi terbesar terjadi pada kelas negatif yang sebagian besar salah diprediksi sebagai positif (34 data), diikuti kelas netral yang salah diprediksi sebagai negatif (32 data) dan kelas negatif yang salah diprediksi sebagai netral (23 data), mengindikasikan adanya ambiguitas semantik antara komentar negatif dan ekspresi sentimen positif maupun netral. Performa pada kelas negatif yang relatif lebih rendah menunjukkan bahwa karakteristik komentar negatif yang cenderung bercampur dengan ekspresi sentimen positif maupun netral memerlukan strategi optimasi lebih lanjut, seperti augmentasi data atau penyempurnaan proses pelabelan pada kelas tersebut.

Kata kunci: Analisis Sentimen*;* IndoBERT*; Fine-Tuning, Roblox; Violence District.*

1. LATAR BELAKANG

Perkembangan teknologi digital telah membawa perubahan besar dalam perilaku masyarakat terutama pada sektor hiburan berbasis internet. Industri game daring mengalami pertumbuhan yang signifikan dengan hadirnya platform interaktif seperti Roblox, yang menjadi salah satu ekosistem permainan berbasis user-generated content terbesar di dunia. Roblox merupakan salah satu platform game berbasis user-generated content yang mencatat lebih dari 54 juta pengguna aktif harian secara global pada 2022 (Dewi, Sudipa, Sunarya, Kusuma Dewi, & Kusuma, 2025). Popularitas Roblox tidak hanya didorong oleh variasi permainannya, tetapi juga kemampuannya sebagai ruang sosial, tempat pemain dapat berinteraksi, berkreasi, dan berbagi pengalaman. Kondisi ini menjadikan Roblox sebagai media yang kaya untuk dianalisis, terutama dalam konteks opini dan perilaku pengguna yang terekspresi melalui ulasan digital (Yuningsih & Jaizul, 2025).

Salah satu game yang cukup dikenal dalam platform Roblox adalah Violence District yang menjadi salah satu permainan yang cukup populer dan sering dibicarakan oleh pemain. Popularitasnya mendorong banyak pengguna untuk membagikan

pengalaman dan opini mereka secara terbuka, termasuk melalui komentar di platform TikTok. Telah banyak penelitian yang mengkaji analisis sentimen pengguna terhadap game digital dan platform daring berbasis komentar media sosial, baik untuk menilai tingkat kepuasan, persepsi pengguna, maupun potensi dampak sosial dari sebuah permainan (Manalu, n.d.). Namun demikian, permasalahan utama yang muncul adalah belum adanya penelitian sistematis yang menganalisis sentimen pengguna terhadap game *Violence District* secara khusus. Ulasan mengenai game ini dapat ditemukan pada platform TikTok. Ulasan pengguna game menyediakan masukan berharga bagi pengembang untuk meningkatkan kualitas fitur, menanggapi keluhan, dan memahami preferensi pemain (Alkindi & Nasution, 2024).

Penelitian oleh Dewi yang menganalisis ulasan pengguna terhadap aplikasi Roblox di Google Play Store dengan memanfaatkan pembobotan TF-IDF dan lexicon InSet. Hasil penelitian tersebut menunjukkan bahwa metode SVM yang dipadukan dengan pendekatan leksikal mampu mencapai akurasi 87,03% dalam mengolah ulasan pengguna secara real-time (Dewi et al., 2025). Perkembangan metode berbasis Transformer mulai menggantikan pendekatan tradisional. Penelitian oleh Ferdiansyah yang menganalisis sentimen ulasan pengguna game Roblox berbahasa Indonesia menerapkan model IndoBERT berbasis Transformer dengan optimasi SMOTE untuk mengatasi ketidakseimbangan kelas. Dataset yang digunakan berjumlah 25.000 ulasan yang dikumpulkan dari Google Play Store dan diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Hasil evaluasi menunjukkan bahwa model IndoBERT mampu mencapai akurasi sebesar 94% sebelum optimasi dan meningkat menjadi 95% setelah penerapan SMOTE, dengan nilai precision, recall, dan F1-score yang konsisten tinggi di setiap kelas sentiment (Suryono, 2025). Penelitian oleh Haris. yang menerapkan IndoBERT untuk analisis sentimen ulasan game eFootball menunjukkan bahwa fine-tuning IndoBERT dengan teknik validasi tambahan mampu meningkatkan akurasi hingga 92%, yang membuktikan superioritas model berbasis Transformer dalam memahami konteks semantik ulasan pengguna dibandingkan metode machine learning klasik (Haris, Suharso, & Nurkifli, 2024). Selanjutnya, penelitian oleh Nuryadi yang melakukan fine-tuning IndoBERT pada ulasan aplikasi Tiket.com di Google Play Store menghasilkan akurasi keseluruhan sebesar 91%, dengan F1-score tinggi pada kelas positif dan negatif. Hasil ini semakin menegaskan bahwa IndoBERT

efektif dalam menangani teks berbahasa Indonesia yang bersumber dari platform digital dengan karakteristik bahasa yang beragam (Nuryadi, 2025).

Meskipun beberapa penelitian telah membahas analisis sentimen pada game atau platform Roblox secara umum, belum banyak penelitian yang secara khusus menganalisis satu game tertentu di dalam platform. Penelitian-penelitian terdahulu cenderung berfokus pada topik politik, aplikasi non-game, atau game populer lain seperti Genshin Impact, masih kurang penelitian yang menggunakan model IndoBERT berbasis Transformer pada data yang bersumber dari media sosial dengan karakteristik bahasa informal seperti TikTok. Kesenjangan penelitian ini menjadi peluang penting untuk studi lebih lanjut. Penelitian ini melengkapi kekurangan penelitian terdahulu dengan mengusulkan penerapan fine-tuning IndoBERT untuk menganalisis sentimen ulasan pengguna terhadap game Violence District pada platform TikTok.

Oleh karena itu, penelitian ini memusatkan analisis pada satu game tertentu, penelitian ini memberikan pemahaman yang lebih mendalam tentang aspek yang paling disukai atau dikritik oleh pemain, serta potensi perbaikan yang dapat dilakukan pengembang. Hasil penelitian diharapkan dapat menjadi rujukan bagi pengembang Violence District, peneliti analisis sentimen, serta akademisi yang mempelajari interaksi pengguna pada platform berbasis user-generated content.

2. KAJIAN TEORITIS

A. Analisis Sentimen

Analisis sentimen merupakan bidang kajian dalam text mining dan natural language processing yang bertujuan untuk mengidentifikasi, menganalisis, serta mengklasifikasikan pendapat, emosi, sikap, atau evaluasi yang terkandung dalam suatu teks ke dalam kategori tertentu, seperti sentimen positif atau negatif. Analisis sentimen juga dikenal sebagai opinion mining, yaitu studi komputasional yang berfokus pada pengolahan informasi subjektif yang diekspresikan oleh pengguna dalam bentuk bahasa tertulis (Untuk et al., n.d.). Dalam penelitian ini, analisis sentimen diterapkan untuk memahami persepsi pemain terhadap game *Violence District* berdasarkan ulasan yang tersebar di platform digital Tiktok (Lacy, n.d.).

B. BERT (*Bidirectional Encoder Representations from Transformers*)

BERT (*Bidirectional Encoder Representations from Transformers*) merupakan model bahasa berbasis arsitektur Transformer yang dikembangkan untuk memahami

representasi bahasa secara dua arah (*bidirectional*). Berbeda dengan model sebelumnya yang cenderung memproses teks secara berurutan, BERT mampu mempertimbangkan konteks kata dari sisi kiri dan kanan secara bersamaan melalui setiap lapisan *encoder.* Dalam proses pelatihannya (*pretraining*), BERT menggunakan dua metode utama, yaitu *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP). Kedua metode tersebut membantu model memahami hubungan antar kata serta keterkaitan antar kalimat berdasarkan konteksnya. Kemampuan dalam menangkap informasi kontekstual secara lebih mendalam membuat BERT mampu memberikan performa yang lebih baik dibandingkan beberapa model sebelumnya, seperti LSTM dan CNN, terutama pada berbagai tugas NLP seperti klasifikasi teks dan analisis sentimen (Devlin, Chang, Lee, & Toutanova, 2019).

Secara arsitektur, BERT terdiri dari beberapa lapisan bidirectional encoder yang disusun secara bertingkat. Pada tahap pretraining, BERT dilatih menggunakan pendekatan unsupervised learning agar mampu mempelajari pola dan representasi bahasa secara umum sebelum diterapkan pada tugas tertentu melalui proses fine-tuning. Setiap lapisan encoder menerima masukan berupa urutan kata atau kalimat, kemudian menggunakan mekanisme self-attention untuk memahami hubungan antar kata berdasarkan konteksnya. Hasil dari proses tersebut selanjutnya diproses melalui jaringan feed-forward sebelum diteruskan ke lapisan encoder berikutnya. Pada model IndoBERT-base, proses pemodelan ini dilakukan sebanyak 12 lapisan encoder sesuai dengan jumlah lapisan yang dimiliki, sehingga model mampu menghasilkan representasi teks yang lebih kontekstual. Pada varian IndoBERT-base, proses tersebut dilakukan sebanyak dua belas kali sesuai jumlah lapisan encoder yang dimilikinya (Wijaya, Seputra, & Dewi, 2025).

C. IndoBERT (Indonesian BERT)

IndoBERT merupakan implementasi BERT untuk bahasa Indonesia yang dikembangkan dalam kerangka IndoNLU (*Indonesian Natural Language Understanding*), yaitu benchmark dan sumber daya evaluasi untuk tugas-tugas NLP berbahasa Indonesia (Wilie, n.d.). IndoBERT mengadopsi arsitektur multi-layer bidirectional Transformer encoder dengan mekanisme self-attention sehingga mampu menghasilkan representasi teks yang kontekstual melalui pelatihan dua arah (*bidirectional training*). Model ini dilatih menggunakan korpus besar berbahasa Indonesia (Indo4B) yang mencakup teks formal dan informal, sehingga mampu

menangkap variasi linguistik bahasa Indonesia secara lebih komprehensif. Dalam pengembangannya, IndoBERT tersedia dalam dua varian arsitektur, yaitu base (12 layer, 12 attention heads, 768 hidden size, ±124 juta parameter) dan large (24 layer, 16 attention heads, 1024 hidden size, ±335 juta parameter), yang memungkinkan fleksibilitas dalam pemilihan kapasitas model sesuai kompleksitas tugas. Selain digunakan melalui fine-tuning langsung pada tugas downstream seperti klasifikasi sentimen, IndoBERT juga dapat ditingkatkan performanya melalui pendekatan continual pre-training atau *domain-specific post-training*, di mana model generik dilatih kembali menggunakan korpus domain tertentu untuk memperkuat pemahaman konteks spesifik sebelum dilakukan fine-tuning. IndoNLU sendiri menyediakan berbagai dataset evaluasi, termasuk tugas klasifikasi sentimen (SmSA), klasifikasi topik, dan tugas pemahaman bahasa lainnya, yang membuktikan bahwa IndoBERT secara konsisten mengungguli model machine learning klasik maupun embedding statis dalam berbagai skenario NLP berbahasa Indonesia (Maharani, Yustiawan, Rochim, & Purwarianti, 2023).

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif. Desain penelitian yang digunakan adalah desain eksperimen komputasional berbasis *deep learning*. Secara umum, alur penelitian meliputi pengumpulan 6.970 komentar melalui *web scraping*, *preprocessing* teks hingga diperoleh 5.370 data bersih, pelabelan sentimen semi-otomatis berbasis *lexicon* yang diverifikasi manual, tokenisasi menggunakan tokenizer IndoBERT (*WordPiece tokenizer*), pembagian data menjadi 4.296 data latih, 537 data validasi, dan 537 data uji, proses *fine-tuning* model IndoBERT untuk klasifikasi sentimen ke dalam tiga kelas (positif, negatif, netral), serta evaluasi performa model menggunakan metrik klasifikasi *accuracy*, *precision*, *recall*, dan *F1-score* Pengumpulan data dilakukan menggunakan teknik *web scraping* pada platform digital yang menjadi media utama pemain dalam memberikan ulasan dan komentar terkait game Violence District. Platform yang digunakan adalah TikTok. Data yang dikumpulkan berupa ulasan atau komentar pengguna dalam bentuk teks yang membahas pengalaman bermain, kualitas *gameplay*, performa teknis, serta permasalahan yang dialami pemain. Proses pengambilan data dilakukan menggunakan kata kunci yang relevan, seperti “Violence District Roblox”, “Violence District gameplay”, “Violence District bug”, dan variasi kata kunci lainnya

yang berkaitan dengan game tersebut. Pengumpulan data dibatasi pada rentang waktu Januari 2025 hingga Maret 2026 untuk menjaga konsistensi dan relevansi data.

4. HASIL DAN PEMBAHASAN

A. Implementasi Fine-Tuning Model IndoBERT

Proses pelatihan model dilakukan menggunakan pendekatan full fine-tuning, yaitu seluruh parameter model IndoBERT diperbarui selama proses pelatihan. Model dilatih dengan konfigurasi learning rate = 2e-5, batch size = 16, optimizer AdamW dengan weight decay 0,01, maksimum epoch = 10, dan patience = 3 untuk mekanisme Early Stopping. Model terbaik ditentukan berdasarkan nilai validation loss terendah yang dicapai selama proses pelatihan.

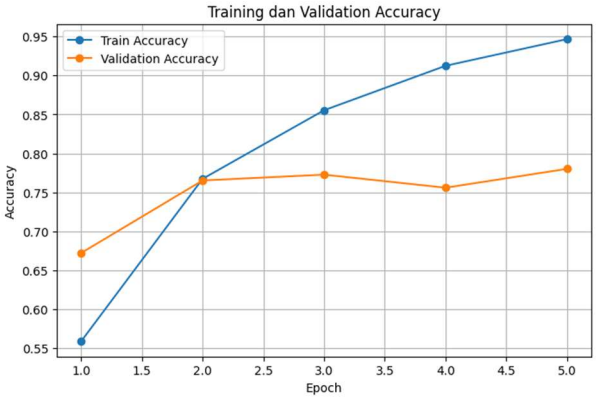
Model IndoBERT (indobenchmark/indobert-base-p1) dimuat dengan layer classifier yang diinisialisasi ulang agar sesuai dengan tiga kelas sentimen pada penelitian ini (POSITIF, NEGATIF, NETRAL). Proses pelatihan dipantau pada setiap epoch melalui nilai training loss, validation loss, training accuracy, dan validation accuracy. Proses pelatihan berhenti pada Epoch 5 karena mekanisme Early Stopping terpicu setelah validation loss tidak membaik selama 3 epoch berturut-turut, dan model terbaik tersimpan pada Epoch 2 dengan validation loss terendah sebesar 0,4785.

Tabel 1 Hasil Pelatihan Model IndoBERT Per Epoch

Epoch	Train Loss	Validation Loss	Train Accuracy	Validation Accuracy
1	0,9046	0, 7566	55,89 %	67,23%
2	0,5786	0,6145	76,75%	76,54%
3	0,3701	0,6548	85,54%	77,28%
4	0,2333	0,8071	91,25%	75,61%
5	0,1632	0,9284	94,67%	78,03%

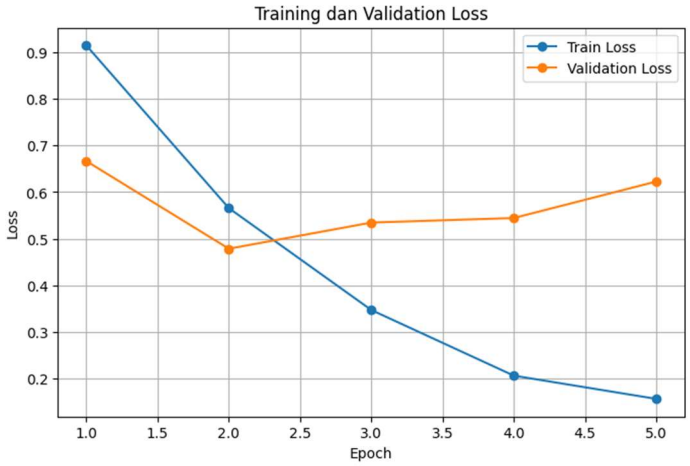
Grafik Accuracy History

Untuk mengevaluasi konsistensi kinerja model yang dibangun, grafik Accuracy History pada Gambar di abwah menunjukkan perbandingan antara akurasi data training dan data validation pada setiap epoch. Grafik ini digunakan untuk melihat seberapa besar perbedaan antara akurasi training dan validation sebagai indikator ada atau tidaknya overfitting pada model, dengan jumlah epoch yang ditentukan sebanyak 5 epoch dan penerapan fungsi early stopping.



Gambar 1 Grafik Accuracy History Training dan Validation

Berdasarkan Gambar di atas, kurva training accuracy (garis biru) menunjukkan peningkatan yang konsisten dan tajam dari 55,89% pada epoch pertama hingga 94,67% pada epoch kelima. Kurva validation accuracy (garis oranye) juga menunjukkan tren peningkatan secara keseluruhan, meski dengan pola yang lebih landai dan sedikit fluktuatif, yaitu meningkat signifikan pada dua epoch pertama dari 67,23% menjadi 76,54%, kemudian meningkat sedikit pada epoch ketiga menjadi 77,28%, sebelum menurun pada epoch keempat menjadi 75,61%, dan kembali meningkat pada epoch kelima menjadi 78,03%.



Gambar 2 Grafik Loss History Training dan Validation

Berdasarkan Gambar di atas, kurva training loss (garis biru) menunjukkan penurunan yang konsisten dari 0,9046 pada epoch pertama menjadi 0,1632 pada epoch kelima, menandakan model terus mempelajari pola dari data latih dengan baik. Kurva validation loss (garis oranye) menunjukkan pola yang berbeda, yaitu menurun dari 0,7566 pada epoch pertama menjadi titik terendah 0,6145 pada epoch kedua, kemudian berbalik

meningkat secara konsisten pada epoch ketiga (0,6548), epoch keempat (0,8071), hingga epoch kelima (0,9284).

Kedua grafik tersebut saling mengonfirmasi satu sama lain bahwa model mengalami overfitting sejak epoch ketiga, di mana model semakin menghafal pola spesifik pada data latih namun tidak diiringi peningkatan kemampuan generalisasi pada data validasi. Validation loss yang tidak mengalami perbaikan selama tiga epoch berturut-turut (epoch 3, 4, dan 5) berhasil terdeteksi oleh mekanisme early stopping dengan patience=3, sehingga proses pelatihan dihentikan secara otomatis pada epoch kelima. Model dengan validation loss terbaik, yaitu pada epoch kedua dengan nilai loss 0,6145 dan akurasi validasi 76,54%, kemudian disimpan dan dimuat kembali sebagai model final yang digunakan pada tahap evaluasi. Hasil ini menunjukkan bahwa mekanisme early stopping yang diterapkan berhasil bekerja sebagaimana mestinya, yaitu mencegah model dipertahankan pada kondisi yang semakin overfitting dan memastikan model akhir yang dipilih memiliki kemampuan generalisasi terbaik yang dapat dicapai selama proses pelatihan.

B. Evaluasi Model

Evaluasi performa model IndoBERT yang telah di-fine-tune dilakukan menggunakan data uji yang sepenuhnya terpisah dan tidak terlibat dalam proses pelatihan maupun validasi. Pengukuran performa menggunakan empat metrik utama yaitu accuracy, precision, recall, dan F1-score, serta analisis confusion matrix untuk memahami distribusi kesalahan klasifikasi pada masing-masing kelas sentimen secara mendetail.

Hasil Metrik Evaluasi

Tabel 2 Hasil Metrik Evaluasi Model IndoBERT

Metrik Evaluasi	Nilai
Accuracy	75,61%
Precision (macro)	76,36%
Recall (macro)	72,51%
F1-Score (macro)	73,33%

Berdasarkan pada Tabel di atas, model IndoBERT yang telah di-fine-tune berhasil mencapai akurasi sebesar 75,61% pada data uji. Nilai precision macro sebesar 76,36% menunjukkan rata-rata ketepatan model dalam memprediksi setiap kelas sentimen tanpa mempertimbangkan proporsi jumlah data pada masing-masing kelas. Nilai recall macro sebesar 72,51% mencerminkan kemampuan rata-rata model dalam menemukan kembali

data yang termasuk dalam masing-masing kelas secara menyeluruh. F1-score macro sebesar 73,33% sebagai rata-rata harmonis antara precision dan recall memberikan penilaian yang komprehensif dan seimbang terhadap performa model secara keseluruhan pada ketiga kelas, dan menunjukkan bahwa model tidak condong hanya pada kelas dengan jumlah data yang lebih dominan.

Klasifikasi Report Per Kelas

Selain evaluasi secara keseluruhan, dilakukan pula analisis performa per kelas sentimen untuk mengetahui kemampuan model dalam mengklasifikasikan masing-masing kategori secara individual.

Tabel 3 Klasifikasi Report Per Kelas Sentimen

Kelas	Precision	Recall	F1-Score	Support
POSITIF	72,32%	93,72%	81,64%	223
NEGATIF	87,68%	69,14%	77,32%	175
NETRAL	69,09%	54,68%	61,04%	139
Macro Avg	76,36%	72,51%	73,33%	537
Weighted Avg	76,49%	75,61%	74,90%	537
Accuracy	-	-	75,61%	537

Berdasarkan Tabel di atas, model menghasilkan accuracy keseluruhan sebesar 75,61%. Kelas POSITIF mencatat recall tertinggi di antara ketiga kelas, yaitu sebesar 93,72%, dengan F1-score sebesar 81,64%, menunjukkan bahwa model sangat baik dalam mengenali komentar yang mengandung sentimen positif dan mampu menemukan sebagian besar komentar aktual POSITIF dengan tepat. Namun, precision POSITIF hanya sebesar 72,32%, menunjukkan bahwa cukup banyak prediksi POSITIF yang sebenarnya berasal dari kelas lain, terutama dari kelas NETRAL.

Kelas NEGATIF menunjukkan precision tertinggi di antara ketiga kelas, yaitu sebesar 87,68%, yang berarti prediksi model terhadap kelas NEGATIF cenderung akurat. Namun, recall-nya hanya sebesar 69,14%, sehingga masih terdapat cukup banyak komentar NEGATIF yang tidak berhasil dikenali oleh model dan justru diklasifikasikan ke kelas lain. F1-score kelas NEGATIF sebesar 77,32% dengan jumlah sampel sebesar 175 dari 537 data uji.

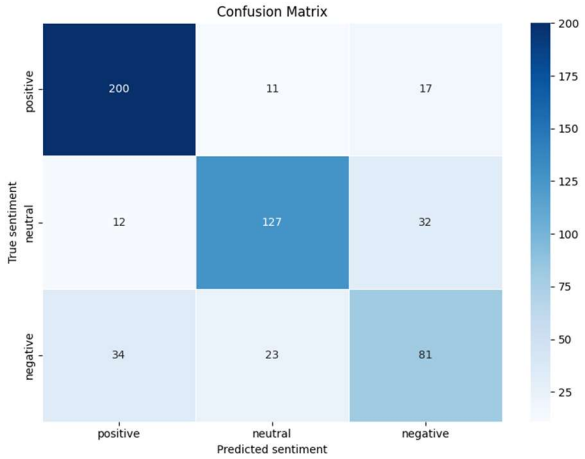
Kelas NETRAL memiliki performa terendah di antara ketiga kelas, dengan precision 69,09%, recall 54,68%, dan F1-score 61,04%. Hal ini menunjukkan bahwa model memiliki tingkat kesalahan yang relatif lebih tinggi dalam mengidentifikasi komentar netral dibandingkan dua kelas lainnya, kemungkinan disebabkan oleh

karakteristik komentar netral yang sering bercampur dengan ekspresi sentimen lemah sehingga sulit dibedakan secara konsisten.

Secara keseluruhan, nilai Macro Average Precision sebesar 76,36%, Recall sebesar 72,51%, dan F1-Score sebesar 73,33%, sementara Weighted Average Precision sebesar 76,49%, Recall sebesar 75,61%, dan F1-Score sebesar 74,90%. Hasil ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang cukup baik secara keseluruhan, meskipun masih terdapat ruang perbaikan, khususnya pada kelas NETRAL yang performanya jauh di bawah dua kelas lainnya.

Analisis Confusion Matrix

Confusion matrix digunakan untuk menganalisis distribusi prediksi model secara lebih mendetail dengan membandingkan label aktual dan label yang diprediksi oleh model pada setiap kelas. Gambar 5 menunjukkan confusion matrix hasil prediksi model IndoBERT pada 537 data uji, di mana sumbu vertikal merepresentasikan label aktual (true sentiment) dan sumbu horizontal merepresentasikan label yang diprediksi oleh model (predicted sentiment).



Gambar 3 Confusion Matrix Model IndoBERT pada Data Uji

Berdasarkan pada Gambar di atas, hasil confusion matrix menunjukkan bahwa dari 228 data dengan label aktual POSITIF, model berhasil memprediksi dengan benar sebanyak 200 data, sedangkan 11 data salah diprediksi sebagai NETRAL dan 17 data salah diprediksi sebagai NEGATIF. Untuk kelas NETRAL, dari total 171 data aktual, model berhasil mengklasifikasikan dengan benar sebanyak 127 data, dengan 12 data yang salah diprediksi sebagai POSITIF dan 32 data salah diprediksi sebagai NEGATIF.

Sementara itu, untuk kelas NEGATIF, dari total 138 data aktual, model berhasil memprediksi dengan benar sebanyak 81 data, dengan 34 data salah diprediksi sebagai POSITIF dan 23 data salah diprediksi sebagai NETRAL.

Kesalahan klasifikasi terbesar terjadi pada kelas NEGATIF yang salah diprediksi sebagai POSITIF sebanyak 34 kasus, diikuti kesalahan kelas NETRAL yang salah diprediksi sebagai NEGATIF sebanyak 32 kasus, dan kesalahan kelas NEGATIF yang salah diprediksi sebagai NETRAL sebanyak 23 kasus. Hal ini mengindikasikan bahwa model masih mengalami kesulitan dalam membedakan komentar negatif dari komentar yang memiliki kecenderungan sentimen positif maupun netral, sehingga cukup banyak komentar negatif yang justru diklasifikasikan sebagai positif atau netral. Kelas NEGATIF menunjukkan recall sebesar 58,70% dengan 81 prediksi benar dari 138 data aktual, menjadikannya kelas dengan tingkat kesalahan klasifikasi tertinggi di antara ketiga kelas.

Tabel 4 Detail Nilai Confusion Matrix

Aktual \ Prediksi	***Positive***	***Negative***	***Neutral***	**Total Aktual**
Positif	200	11	17	228
Negatif	12	127	32	171
Netral	34	23	81	138
Total Prediksi	246	161	130	537

C. Pembahasan

Hasil implementasi dan pengujian menunjukkan bahwa model IndoBERT yang di-*fine-tune* pada dataset komentar TikTok mengenai game Violence District berhasil mencapai performa klasifikasi sentimen dengan akurasi 75,61% dan F1-score *weighted* 74,90%. Capaian ini membuktikan bahwa pendekatan *fine-tuning* model Transformer pra-latih berbahasa Indonesia dapat diterapkan pada data komentar media sosial yang bersifat informal dan beragam, meskipun komentar TikTok umumnya lebih singkat, mengandung lebih banyak bahasa gaul, singkatan, dan ekspresi informal yang lebih sulit diproses dibandingkan teks yang lebih formal dan terstruktur.

Penerapan mekanisme *early stopping* pada penelitian ini terbukti berperan penting dalam menghasilkan model yang memiliki kemampuan generalisasi yang baik. Berdasarkan grafik Accuracy History, kurva *training accuracy* meningkat secara konsisten dari 55,89% pada epoch pertama hingga 94,67% pada epoch kelima. *Validation accuracy* juga menunjukkan tren peningkatan meski dengan pola yang lebih landai dan sedikit fluktuatif, yaitu meningkat signifikan pada dua epoch pertama dari 67,23%

menjadi 76,54%, sedikit meningkat pada epoch ketiga menjadi 77,28%, kemudian menurun pada epoch keempat menjadi 75,61%, dan kembali meningkat pada epoch kelima menjadi 78,03%. Gap antara kedua kurva yang semula hanya sekitar 0,21 poin persentase pada epoch kedua melebar menjadi sekitar 8,26 poin persentase pada epoch ketiga, dan terus melebar hingga sekitar 16,64 poin persentase pada epoch kelima, yang mengindikasikan model mulai mengalami *overfitting* sejak epoch ketiga. Mekanisme *early stopping* dengan *patience* = 3 berhasil mendeteksi kondisi ini dan menghentikan pelatihan secara otomatis pada epoch kelima, kemudian memuat kembali model dengan *validation loss* terbaik pada epoch kedua (0,6145) sebagai model final yang digunakan pada tahap evaluasi.

Analisis per kelas mengungkapkan bahwa model memiliki performa yang bervariasi antar kelas. Kelas POSITIF mencatat *recall* tertinggi di antara ketiga kelas sebesar 93,72% dan F1-score sebesar 81,64%, sejalan dengan jumlah sampel terbanyak pada data uji sebesar 223 data, sehingga model memiliki lebih banyak data representatif untuk mempelajari karakteristik teks positif. Namun, *precision* POSITIF hanya sebesar 72,32%, yang menunjukkan bahwa cukup banyak prediksi POSITIF sebenarnya berasal dari kelas lain, terutama NETRAL. Kelas NEGATIF berada di posisi tengah dengan *precision* tertinggi di antara ketiga kelas sebesar 87,68%, namun *recall* hanya sebesar 69,14% dan F1-score sebesar 77,32%. Kelas NETRAL menunjukkan performa terendah di antara ketiga kelas dengan *precision* 69,09%, *recall* 54,68%, dan F1-score 61,04%, yang kemungkinan disebabkan oleh karakteristik komentar netral yang sering bercampur dengan ekspresi sentimen lemah sehingga sulit dibedakan secara konsisten oleh model.

Analisis confusion matrix lebih lanjut mengungkapkan pola kesalahan yang informatif. Dari 228 data aktual POSITIF, model berhasil memprediksi dengan benar sebanyak 200 data, dengan 11 data salah diklasifikasikan sebagai NETRAL dan 17 data salah diklasifikasikan sebagai NEGATIF. Dari 171 data aktual NETRAL, model berhasil mengklasifikasikan dengan benar sebanyak 127 data, sementara 12 data salah diprediksi sebagai POSITIF dan 32 data salah diprediksi sebagai NEGATIF. Dari 138 data aktual NEGATIF, model berhasil memprediksi dengan benar sebanyak 81 data, dengan 34 data salah diklasifikasikan sebagai POSITIF dan 23 data sebagai NETRAL. Kesalahan terbesar secara keseluruhan terjadi pada kelas NEGATIF yang salah diprediksi sebagai POSITIF sebanyak 34 kasus, diikuti kelas NETRAL yang salah diprediksi sebagai

NEGATIF sebanyak 32 kasus, dan kelas NEGATIF yang salah diprediksi sebagai NETRAL sebanyak 23 kasus. Pola ini mencerminkan bahwa model masih mengalami kesulitan dalam membedakan komentar negatif dari komentar yang memiliki kecenderungan sentimen positif maupun netral, sehingga cukup banyak komentar negatif justru diklasifikasikan sebagai positif, terutama pada ekspresi informal dan ambigu yang umum dijumpai pada komentar media sosial berbahasa Indonesia.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, penelitian ini berhasil menunjukkan bahwa pendekatan fine-tuning pada model IndoBERT efektif dalam mengklasifikasikan sentimen komentar pengguna TikTok terhadap game Violence District ke dalam tiga kelas sentimen, yaitu POSITIF, NEGATIF, dan NETRAL. Model IndoBERT yang di-fine-tune dengan konfigurasi learning rate 2e-5, batch size 16, optimizer AdamW (weight decay 0,01), maksimum 10 epoch, dan early stopping (patience 3) menghasilkan akurasi 75,61%, precision macro 76,36%, recall macro 72,51%, dan F1-score macro 73,33% pada data uji. Training loss menurun konsisten dari 0,9046 pada epoch pertama menjadi 0,1632 pada epoch kelima, sementara validation loss mencapai titik terbaik pada epoch kedua (0,6145) sebelum kembali meningkat pada epoch-epoch berikutnya. Gap antara training accuracy dan validation accuracy yang semula hanya sekitar 0,21 poin persentase pada epoch kedua melebar menjadi 8,26 poin persentase pada epoch ketiga dan terus melebar hingga 16,64 poin persentase pada epoch kelima, mengindikasikan model mulai overfitting sejak epoch ketiga. Mekanisme early stopping berhasil mendeteksi kondisi ini dan menghentikan pelatihan pada epoch kelima, dengan model pada epoch kedua (validation loss terendah) dimuat kembali sebagai model final.

Analisis per kelas menunjukkan kelas POSITIF memiliki performa tertinggi (F1-score 81,64%, recall 93,72%) sejalan dengan jumlah sampel terbanyak pada data uji, diikuti NEGATIF (F1-score 77,32%, precision tertinggi 87,68%), sedangkan NETRAL menunjukkan performa terendah (precision 69,09%, recall 54,68%, F1-score 61,04%) akibat karakteristik komentar netral yang sering bercampur dengan ekspresi sentimen lemah sehingga sulit dibedakan secara konsisten oleh model. Analisis confusion matrix mengungkapkan kesalahan klasifikasi terbesar terjadi pada kelas NETRAL yang salah diprediksi sebagai POSITIF (50 kasus), diikuti kelas NEGATIF yang salah diprediksi

sebagai POSITIF (30 kasus), dan kelas NEGATIF yang salah diprediksi sebagai NETRAL (24 kasus), menunjukkan model masih kesulitan membedakan komentar netral dari ekspresi sentimen positif atau negatif yang lemah, terutama pada bahasa informal dan ambigu yang umum dijumpai pada komentar media sosial berbahasa Indonesia.

DAFTAR REFERENSI

Alkindi, A. F., & Nasution, N. (2024). ANALISIS SENTIMEN ULASAN PENGGUNA PADA GAME ROBLOX DENGAN METODE SUPPORT VECTOR MACHINE DAN NAIVE BAYES. *J-Com (Journal of Computer)*, *4*(2), 164–177. https://doi.org/10.33330/j-com.v4i2.3319

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. Stroudsburg, PA, USA: Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Dewi, N. K. F. P., Sudipa, I. G. I., Sunarya, I. W., Kusuma Dewi, N. W. J., & Kusuma, A. S. (2025). Sentiment Analysis of Roblox Game Reviews Using Support Vector Machine Method. *Sinkron*, *9*(4), 1863–1876. https://doi.org/10.33395/sinkron.v9i4.15272

Haris, M., Suharso, A., & Nurkifli, E. H. (2024). *Analisis Sentimen Pada Game Efootball Di Google Play Store Menggunakan Algoritma Indobert*.

Lacy, F. (n.d.). *International Journal Of Electrical And Computer Engineering*. Retrieved from https://digitalcommons.subr.edu/Osp_Facpubs/13

Maharani, N. P. I., Yustiawan, Y., Rochim, F. C., & Purwarianti, A. (2023). *Domain-Specific Language Model Post-Training For Indonesian Financial Nlp*. Retrieved from http://arxiv.org/Abs/2310.09736

Manalu, P. D. (n.d.). Implementasi Algoritma Klasifikasi Untuk Analisis Sentimen Media Sosial Tiktok Tahun 20. *Jurnal Teknik Informatika Dan Teknologi Informasi*. https://doi.org/10.55606/Jutiti.V5i1.5644

Nuryadi, D. (2025). *Fine Tuning Indobert Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Tiket.Com Di Google Play Store*.

Suryono, R. R. (2025). *Sentiment Classification Of Indonesian-Language Roblox Reviews Using Indobert With Smote Optimization*. Retrieved from http://jurnal.polibatam.ac.id/Index.Php/Jaic

Untuk, D., Persyaratan, M., Gelar, M., Komputer, S., Oleh, D., & Fatihin, A. (n.d.). *Analisis Sentimen Terhadap Ulasan Aplikasi Mobile Menggunakan Metode Support Vector Machine (Svm) Dan Pendekatan Lexicon Based*.

Wijaya, I. N. S. W., Seputra, K. A., & Dewi, N. P. N. P. (2025). Fine Tunning Model Indobert Untuk Analisis Sentimen Berita Pariwisata Indonesia. *Jurnal Pendidikan Teknologi Dan Kejuruan*, *22*(2). Retrieved from https://www.detik.com/Search/Searchall?Query=Wisata&Siteid=3&Sortby=Time&Fromdatex=01/01/2022&

Wilie, B. (n.d.). *Indonlu: Benchmark And Resources For Evaluating Indonesian Natural Language Understanding*. Retrieved from https://github.com/Annisanurulazhar/Absa-Playground

Yuningsih, S. H., & Jaizul, A. (2025). Roblox As A Digital Cultural Space: Generation Z's Identity, Creativity, And Social Interaction. *International Journal Of Humanities, Law, And Politics*, *3*(4), 164–168.