

**ANALISIS SENTIMEN PADA ULASAN PENGGUNA GAME MOBILE
LEGENDS DI PLAY STORE MENGGUNAKAN DEEP LEARNING
TRANSFORMERS BERT**

Muhammad Khoiril S
Universitas Satya Negara Indonesia
Khoiril001@gmail.com
Prionggo Hendradi
Universitas Satya Negara Indonesia
prionggo.hendradi@gmail.com
Riama Sibarani
Universitas Satya Negara Indonesia
riama_sarah@yahoo.com
Yusriani Saptadewi
Universitas Satya Negara Indonesia
yusrianisaptadewi@usni.ac.id

Keywords	ABSTRACT
Sentiment Analysis, Deep Learning, Mobile Legends, Transformers BERT, IndoBERT, Natural Language Processing	<i>The Mobile Legends: Bang Bang game generates millions of user reviews containing valuable sentiment for developers, but their analysis is hampered by the large data volume and the use of complex informal language. This research aims to design and implement an automated sentiment analysis system using the Transformers BERT Deep Learning model to classify user reviews into positive, negative, and neutral categories. The research method includes collecting 548,250 review data from Kaggle, pre-processing text to handle noise, and training the IndoBERT model through fine-tuning techniques. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics. The results show that the fine-tuned IndoBERT model achieved an accuracy of 83.2%, a significant improvement of +62.8% compared to the baseline (pre-trained) model which only achieved 20.4%. This success proves that the fine-tuning process effectively adapts the model to understand the unique jargon and context in game reviews. The entire research flow is implemented in an interactive web application using Streamlit as a proof of concept and a tool for visualizing the analysis results.</i>

Kata kunci	ABSTRAK
Analisis Sentimen; Deep Learning; IndoBERT; Mobile Legends; Pengolahan Bahasa Alami; Transformers BERT.	<i>Game Mobile Legends: Bang Bang menghasilkan jutaan ulasan pengguna yang mengandung sentimen berharga bagi pengembang, namun analisisnya terkendala oleh volume data yang besar serta penggunaan bahasa informal yang kompleks. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem analisis sentimen otomatis menggunakan model Deep Learning Transformers BERT untuk mengklasifikasikan ulasan pengguna ke dalam kategori positif, negatif, dan netral. Metode penelitian mencakup pengumpulan 548.250 data ulasan dari Kaggle, pra-pemrosesan teks untuk menangani noise, dan pelatihan model IndoBERT melalui teknik fine-tuning. Kinerja model dievaluasi menggunakan metrik</i>

akurasi, presisi, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa model IndoBERT yang telah disempurnakan (*fine-tuned*) mencapai akurasi 83.2%, sebuah peningkatan signifikan sebesar +62.8% dibandingkan model dasarnya (*pre-trained*) yang hanya mencapai 20.4%. Keberhasilan ini membuktikan bahwa proses *fine-tuning* secara efektif mengadaptasi model untuk memahami jargon dan konteks unik dalam ulasan game. Seluruh alur penelitian diimplementasikan dalam sebuah aplikasi web interaktif menggunakan Streamlit sebagai bukti konsep dan alat visualisasi hasil analisis.

PENDAHULUAN

Perkembangan teknologi dan internet mendorong interaksi masyarakat melalui *platform* digital, termasuk dalam hiburan *game mobile* yang sangat populer seperti *Mobile Legends: Bang Bang* dengan jutaan ulasan pengguna di *Google Play Store* (Google Play Store, 2025; Zhao et al., 2021). Ulasan tersebut mengandung sentimen penting untuk pengembang, namun volume dan bahasa informalnya menyulitkan analisis manual (Feng et al., 2023).

Permasalahan signifikan muncul dari ulasan negatif yang tidak terkelola dengan baik, berpotensi menurunkan rating dan merugikan pengembang hingga 15% pendapatan (Chen et al., 2022; r/MobileLegends, 2024). Kompleksitas bahasa Indonesia yang mengandung slang dan *code-mixing* menuntut model analisis sentimen yang mampu memahami konteks lebih mendalam (Wilie & others, 2020; Nugraha et al., 2021).

Model *Deep Learning Transformer* seperti BERT terbukti unggul dalam pemahaman konteks dua arah, memberikan hasil akurasi tinggi dalam analisis sentimen teks (Devlin et al., 2019). Namun, penerapan BERT khusus untuk ulasan game berbahasa Indonesia masih minim, sehingga penelitian ini menggunakan IndoBERT yang di-fine-tune untuk konteks tersebut.

Penelitian ini bertujuan menerapkan dan menguji efektivitas model IndoBERT *fine-tuned* pada dataset ulasan *Mobile Legends* berbahasa Indonesia, serta menyajikan hasilnya dalam aplikasi *web* interaktif untuk memudahkan visualisasi dan pengambilan keputusan.

Kontribusi utama penelitian adalah adaptasi model bahasa Indonesia untuk domain game dengan data spesifik, peningkatan performa akurasi signifikan dibanding model dasar, serta pengembangan aplikasi web sebagai alat analisis dan evaluasi model sentimen yang dapat digunakan secara praktis (Devlin et al., 2019; D. Pratama & others, 2023).

LANDASAN TEORI

Pada penelitian yang berjudul “Analisis Sentimen *Review* Hotel menggunakan Metode *Deep Learning BERT*”. Hasil penelitian tersebut menggunakan varian BERT, yaitu SmallBERT, untuk klasifikasi sentimen pada 515.000 ulasan hotel dan berhasil mencapai akurasi 91,4%. Studi ini membuktikan efektivitas tinggi dari arsitektur BERT pada data ulasan, meskipun domainnya adalah perhotelan.

Analisis Sentimen

Analisis sentimen adalah proses komputasional untuk mengkategorikan opini dalam teks ke dalam polaritas positif, negatif, atau netral, yang hasilnya sering digunakan sebagai dasar pengambilan keputusan berbasis data. Namun, penerapannya pada ulasan pengguna, khususnya dalam konteks Bahasa Indonesia, menghadapi tantangan signifikan akibat karakteristik teks yang tidak baku. Ulasan game sering kali mengandung bahasa gaul, percampuran bahasa (*code-mixing*), ambiguitas kontekstual, serta sarkasme dan ironi, yang menuntut model analisis yang

mampu memahami nuansa makna secara mendalam.

BERT

Google mengembangkan model BERT (*Bidirectional Encoder Representations from Transformers*) yang membawa peningkatan signifikan dalam representasi bahasa. Keunggulan utama BERT terletak pada kemampuannya untuk mempelajari konteks secara dua arah (*bidirectional*), yakni dengan mempertimbangkan informasi dari kata-kata sebelum dan sesudah secara simultan selama tahap pra-pelatihan. Dengan demikian, BERT mampu menangkap makna kata secara lebih mendalam dibandingkan model unidirectional seperti LSTM. Setelah melalui proses pra-pelatihan pada korpus data berskala besar, BERT dapat disesuaikan (*fine-tuned*) untuk berbagai tugas spesifik, termasuk analisis sentimen, dengan kebutuhan data pelatihan yang relatif lebih sedikit namun tetap menghasilkan performa yang unggul. (Devlin et al., 2019)

Meskipun keunggulan BERT telah terbukti, masih terdapat kesenjangan riset yang signifikan. Penelitian yang ada cenderung berfokus pada domain umum atau belum secara eksplisit menganalisis ulasan game *Mobile Legends* yang memiliki karakteristik linguistik sangat unik. Lebih penting lagi, belum ada studi yang menyajikan evaluasi komparatif secara langsung antara kinerja model IndoBERT dasar (*pre-trained*) dengan model yang telah melalui proses *fine-tuning* pada dataset spesifik ini. Oleh karena itu, penelitian ini bertujuan mengisi kesenjangan tersebut dengan menerapkan dan mengevaluasi proses *fine-tuning* untuk memberikan bukti empiris mengenai efektivitas adaptasi domain dalam meningkatkan akurasi analisis sentimen pada teks informal yang sangat menantang.

METODE PENELITIAN

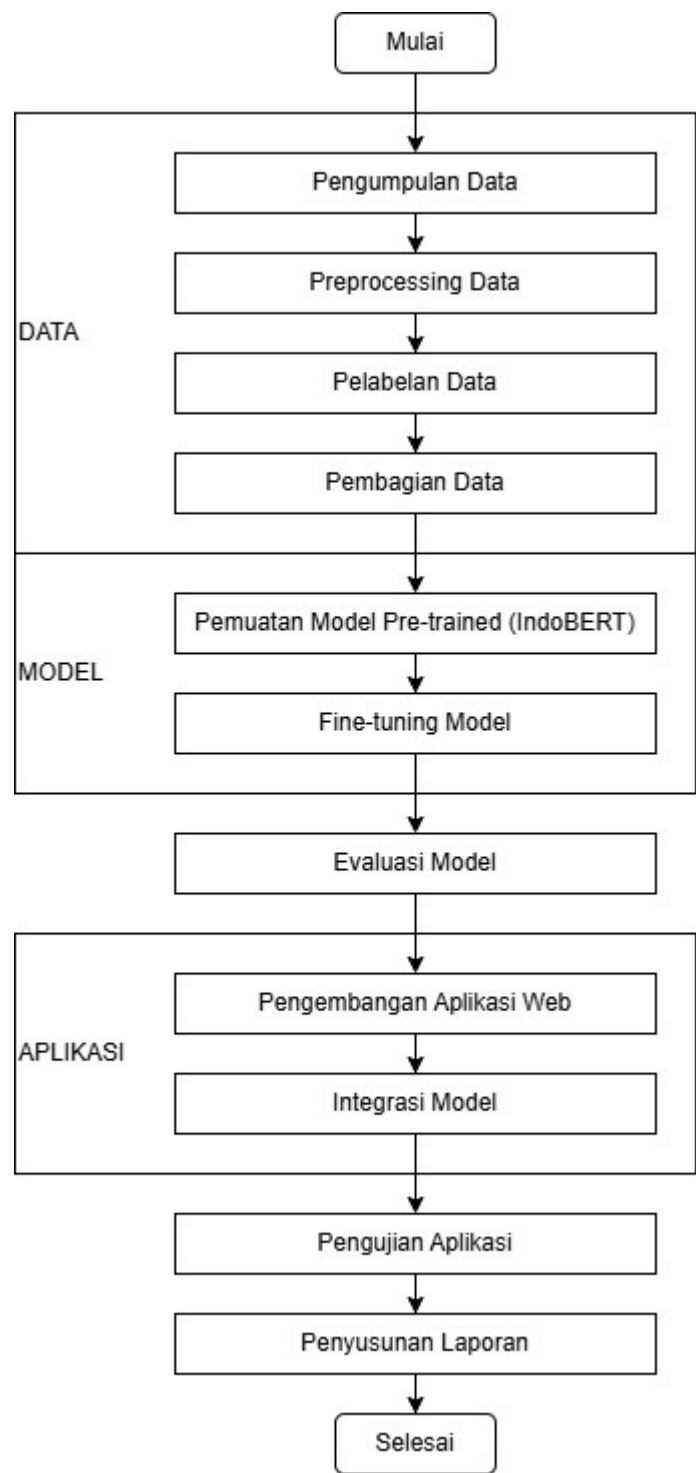
Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimental untuk membandingkan kinerja dua model analisis sentimen. Secara spesifik, penelitian ini membandingkan akurasi model IndoBERT dasar (*pre-trained*) dengan model yang sama setelah melalui proses penyempurnaan (*fine-tuning*) pada dataset spesifik. Desain ini dipilih untuk mengukur secara objektif dampak dari adaptasi domain terhadap kemampuan model dalam memahami konteks ulasan game yang unik.

Sumber data yang digunakan dalam penelitian ini adalah dataset sekunder berjudul *Mobile Legends Play Store Dataset* yang diperoleh dari platform Kaggle. Dataset ini terdiri dari 548.250 ulasan pengguna game *Mobile Legends: Bang Bang* dalam Bahasa Indonesia. Sampel data dibagi menjadi dua bagian: 480.000 ulasan sebagai data latih dan 68.250 ulasan sebagai data uji untuk evaluasi model. Pemilihan dataset ini didasarkan pada kelengkapan dan relevansinya yang tinggi terhadap objek penelitian.

Prosedur penelitian dilakukan melalui beberapa tahapan sistematis. Tahap pertama adalah pra-pemrosesan data (*text preprocessing*), yang mencakup serangkaian proses untuk membersihkan data teks mentah. Proses ini meliputi penghapusan data duplikat (*remove duplicate*), penyeragaman huruf menjadi huruf kecil (*case folding*), pembersihan karakter non-alfabetik, pemisahan teks menjadi token (*tokenization*), normalisasi kata tidak baku menggunakan kamus normalisasi, penghapusan *stopwords*, dan reduksi kata ke bentuk dasarnya (*stemming*). Selanjutnya, dilakukan pelabelan data sentimen secara otomatis berdasarkan skor rating yang diberikan pengguna, dengan kategori: negatif (rating 1-2), netral (rating 3), dan positif (rating 4-5).

Teknik analisis data berpusat pada pemodelan dan evaluasi. Penelitian ini mengimplementasikan model *indobenchmark/indobert-base-p2* sebagai baseline (*pre-trained*) dan melatihnya kembali (*fine-tuning*) pada dataset ulasan *Mobile Legends*. Proses pelatihan dikonfigurasi dengan hyperparameter yang telah teruji efektif untuk arsitektur BERT. Kinerja kedua model kemudian dievaluasi menggunakan *confusion*

matrix untuk menghitung metrik kuantitatif, yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Seluruh proses analisis, pemodelan, dan evaluasi diimplementasikan menggunakan bahasa pemrograman *Python* dalam lingkungan *Google Collaboratory*, dan hasilnya divisualisasikan melalui sebuah aplikasi web interaktif yang dibangun menggunakan framewrok Streamlit.



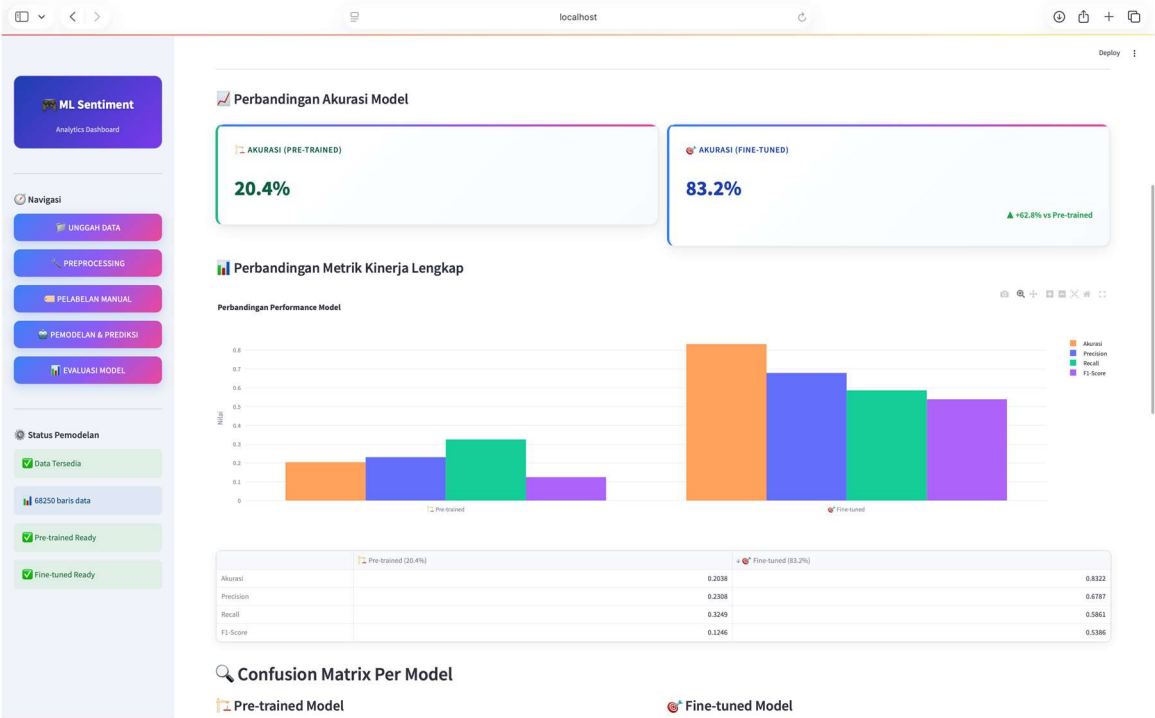
Sumber: Penulis (2025)
Gambar 1. Alur Penelitian

HASIL DAN PEMBAHASAN

Proses analisis data menghasilkan beberapa temuan yang signifikan. Pertama, melalui serangkaian tahapan pra-pemrosesan, data ulasan mentah yang tidak

terstruktur berhasil diubah menjadi format teks yang bersih dan siap untuk dianalisis. Proses ini, yang diintegrasikan ke dalam antarmuka aplikasi web, terbukti krusial untuk menghilangkan *noise* dan menyeragamkan data. Berdasarkan pelabelan otomatis berbasis skor rating, ditemukan bahwa mayoritas ulasan pengguna (sekitar 49.959 dari 66.571 data uji) memiliki sentimen positif, sementara sentimen negatif dan netral memiliki porsi yang lebih kecil.

Pada tahap evaluasi, kinerja model IndoBERT dasar (*pre-trained*) dan model yang telah disempurnakan (*fine-tuned*) dibandingkan untuk mengukur efektivitas *fine-tuning*. Hasilnya menunjukkan peningkatan performa yang signifikan. Model *pre-trained* mencapai akurasi 20.4%, sementara model *fine-tuned* berhasil mencapai akurasi 83.2%. Peningkatan ini sebesar +62.8% dan membuktikan bahwa proses penyempurnaan berhasil mengadaptasi model untuk memahami konteks ulasan game secara lebih spesifik dan akurat.

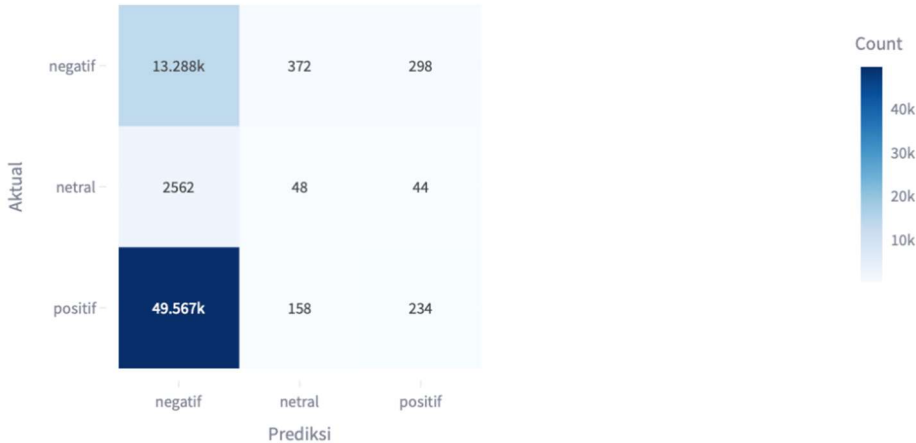


Sumber: Penulis (2025)
Gambar 2. Hasil Perbandingan *Pre-trained* dan *Fine-tuned*

Selain melakukan pelatihan model secara mandiri, penulis juga melakukan evaluasi terhadap model BERT *pre-trained* dan *fine-tuned* hasil evaluasi sebagai berikut:

 Pre-trained Model

Confusion Matrix (Akurasi: 20.4%)



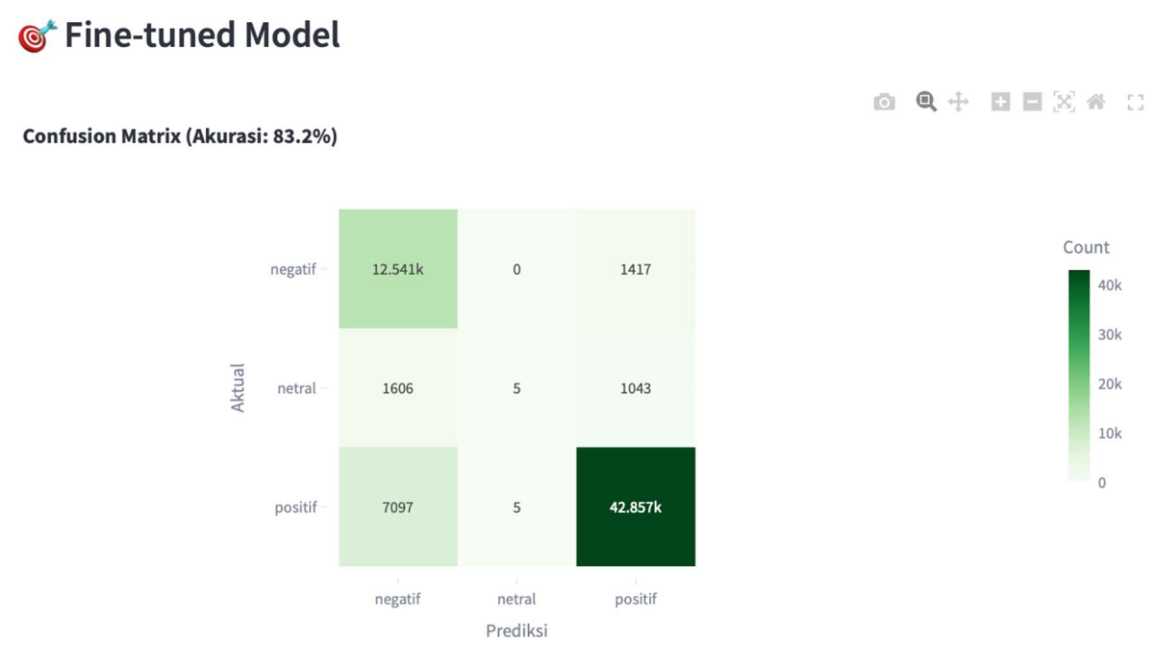
Sumber: Penulis (2025)
Gambar 3. Hasil Evaluasi Model *Pre-trained*

	 Pre-trained (20.4%)
Akurasi	0.2038
Precision	0.2308
Recall	0.3249
F1-Score	0.1246

Sumber: Penulis (2025)
Gambar 4. Hasil Ringkasan Metrik Performa Model *Pre-trained*

Pre-trained mode

1. *Accuracy* = 20,38% → performa sangat rendah, hampir mendekati random guess.
 2. *Precision (macro)* = 0,2308 → rata-rata presisi tiap kelas rendah, model cenderung salah dalam menentukan label.
 3. *Recall (macro)* = 0,3249 → model gagal mengenali sebagian besar sampel pada kelas Netral dan Positif (hanya kelas Negatif yang tinggi).
 4. *F1 (macro)* = 0,1246 → menunjukkan bahwa keseimbangan antara *precision* dan *recall* per kelas sangat buruk.
- Kesimpulan: model *pre-trained* tanpa penyesuaian tidak mampu menangani dataset ini, terutama kelas Netral dan Positif yang *recall*-nya hampir nol.



Sumber: Penulis (2025)
Gambar 5. Hasil Evaluasi Model *Fine-Trained*

Fine-tuned (83.2%)	
	0.8322
	0.6787
	0.5861
	0.5386

Sumber: Penulis (2025)
Gambar 6. Hasil Ringkasan Metrik Performa Model *Fine-Trained*

Fine-tuned model

1. Accuracy = 83,22% performa meningkat drastis, mayoritas prediksi sudah benar.
2. *Precision* (macro) = 0,6787 presisi rata-rata tiap kelas cukup baik, terutama kelas Positif (sangat tinggi).
3. *Recall* (macro) = 0,5861 model lebih mampu menangkap *instance* per kelas, meskipun kelas Netral masih hampir tidak terdeteksi.
4. F1 (macro) = 0,5386 terjadi peningkatan besar dibanding *pre-trained*, menandakan keseimbangan prediksi makin baik.
- Kesimpulan: proses *fine-tuning* berhasil meningkatkan performa model, terutama pada kelas Negatif dan Positif. Namun, model masih sangat lemah pada kelas Netral (*recall* 0,0019).
- Analisis kualitatif melalui visualisasi *word cloud* memberikan wawasan tambahan mengenai leksikon yang digunakan dalam setiap kategori sentimen.
1. Ulasan positif didominasi oleh frasa lugas seperti "sangat bagus" dan "sangat seru".



- Sumber: Penulis (2025)
Gambar 8. Hasil Ulasan Negatif

3. Ulasan netral menunjukkan ambiguitas yang tinggi, sering kali mengandung kata positif seperti "bagus" yang diimbangi dengan keluhan, mengindikasikan adanya sentimen campuran dalam satu ulasan.



Berdasarkan hasil penelitian dan implementasi sistem yang telah dilakukan, maka dapat disimpulkan bahwa:

- Adapun saran untuk pengembangan lebih lanjut dari penelitian ini adalah sebagai berikut:

1. Mengeksplorasi metode pra-pemrosesan data yang lebih canggih, seperti penanganan sarkasme dan ironi secara lebih spesifik, untuk meningkatkan akurasi model pada sentimen ambigu.
2. Menerapkan arsitektur model *Deep Learning* lain (misalnya, varian BERT yang lebih baru atau model Transformer lainnya) untuk perbandingan kinerja yang lebih komprehensif.
3. Mengembangkan fitur analisis aspek (*aspect-based sentiment analysis*) untuk mengidentifikasi sentimen pada fitur-fitur game tertentu (misalnya, matchmaking, skin, atau hero), bukan hanya sentimen secara keseluruhan.

DAFTAR PUSTAKA

- Aisyah, R., & Hidayat, R. (2023). Optimizing BERT for Indonesian Sentiment Analysis: A Case Study of E-Commerce Reviews. *Journal of Southeast Asian Language Processing*, 15(2), 45–60.
- Chen, Z., Li, X., & Wang, Y. (2022). Mobile gaming and its impact on consumer behavior in emerging markets: A case study of Mobile Legends in Indonesia. *Journal of Digital Marketing & Analytics*, 12(4), 215–227.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv. <https://arxiv.org/abs/1810.04805>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 4171–4186.
- Dewanakretarta. (2023). Mobile Legend Playstore Dataset. <https://www.kaggle.com/datasets/dewanakretarta/mobile-legend-playstore-dataset>
- Faraby, S. A., Adiwijaya, A., & Nur, M. H. (2020). Sarcasm Detection on Indonesian Restaurant Reviews using Support Vector Machine and Chi-Square Feature Selection. 1–6. <https://doi.org/10.1109/ICoICT49345.2020.9166228>
- Feng, L., Zhang, Q., & Tan, Y. (2023). Efficient sentiment analysis using *Deep Learning* models for mobile app reviews. *Proceedings of the International Conference on Data Science and Artificial Intelligence*, 123–134.
- Google Play Store. (2025). Mobile Legends: Bang Bang. <https://play.google.com/store/apps/details?id=com.mobile.legends&hl=en&gl=US>
- Koto, F., Rahmaningtyas, A., & al., et. (2020). IndoLEM and IndoBERT: A benchmark dataset and *pre-trained* language model for Indonesian NLP. ArXiv. <https://arxiv.org/abs/2011.00677>
- McKinney, W. (2021). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media.
- Nugraha, A., Budi, I., & Purwarianti, A. (2021). Analisis Sentimen pada Ulasan Aplikasi Mobile Legends: Bang Bang di Google Play Store Menggunakan Metode Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(11), 5157–5165.
- Pratama, D., & others. (2023). Comparative Analysis of *Deep Learning* Models for Indonesian Game Review Sentiment. *IEEE Access*, 11, 12345–12356.
- Pratama, Y. A., & Lhaksana, K. M. (2021). Analisis Sentimen Ulasan Pengguna Aplikasi Game Mobile Legends Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Ilmiah Teknologi Dan Rekayasa*, 26(1), 55–65. <https://doi.org/10.35760/tr.2021.v26i1.3533>
- r/MobileLegends. (2024). Project NEXT Update Feedback. <https://www.reddit.com/r/MobileLegends>

- Siregar, D., Amalia, A., & Budi, I. (2020). Sentiment Analysis on App Reviews with Negation and Contrastive Conjunction Handling. 297–302. <https://doi.org/10.1109/ICACSI51027.2020.9263155>
- Sutabri, T., Rian, H., Hendradi, P., & Febrianto. (2019). Designing the autogate pass dashboard application with Android based responsive web design technology. Proceedings of the First International Conference of Science, Engineering, and Technology, 2019.
- W3C. (2021). Web Applications. <https://www.w3.org/TR/webapps/>
- Wahyuni, E., Setyawan, G., & Nurhayati, I. (2024). FINE TUNING INDOBERT UNTUK ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI TIKET.COM DI GOOGLE PLAY STORE. Jurnal Informatika Teknologi Informasi, 7(1), 221–228. <https://doi.org/10.55338/jikoms.v7i1.2895>
- Wilie, B., & others. (2020). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. ArXiv, 2009.05387.
- Zhao, W., Liu, H., & Zhang, T. (2021). The impact of internet technology on social interactions: A comprehensive study. International Journal of Social Media and Interactive Learning, 9(2), 45–58.