

Analisis Sentimen Ulasan Pengguna Livin' by Mandiri Menggunakan Model Deep Transformer IndoBERT

Duwi Purnama Sidik^{*1}, Fatqurhohman²

¹Universitas Brawijaya, Indonesia

²Universitas Muhammadiyah Jember, Indonesia

e-mail: duwi763@gmail.com

Received: 23-08-2026

Accepted: 03-09-2026

Published: 05-09-2026

Abstrak

Perkembangan layanan perbankan digital menghasilkan sejumlah besar ulasan pengguna yang dapat dimanfaatkan untuk memahami persepsi terhadap kualitas layanan aplikasi. Penelitian ini bertujuan menganalisis sentimen ulasan pengguna aplikasi Livin' by Mandiri menggunakan model IndoBERT melalui proses *fine-tuning*. Data penelitian berasal dari ulasan pengguna yang diklasifikasikan ke dalam dua kelas sentimen, yaitu positif dan negatif, dengan rating 3 dikeluarkan sebagai kategori netral. Sebanyak 15.000 ulasan dipilih menggunakan *stratified sampling* dan dibagi menjadi 70% data latih, 20% validasi, dan 10% data uji. Proses pemodelan menggunakan IndoBERT Base P1 dengan panjang maksimum 128 token dan pelatihan selama tiga epoch. Evaluasi dilakukan menggunakan accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model terbaik diperoleh pada epoch pertama dengan validation accuracy sebesar 92,77% dan F1-score sebesar 94,49%. Pada data uji, model menghasilkan accuracy sebesar 92%, macro F1-score 91%, dan weighted F1-score 92%. F1-score kelas negatif mencapai 0,88, sedangkan kelas positif sebesar 0,94. Hasil tersebut menunjukkan bahwa IndoBERT mampu melakukan klasifikasi sentimen ulasan Livin' by Mandiri dengan kinerja yang baik serta berpotensi digunakan untuk mendukung analisis opini pengguna pada layanan perbankan digital.

Kata Kunci: IndoBERT; Analisis Sentimen; Livin' by Mandiri; Transformer; Ulasan Pengguna.

How to Cite:

Sidik, D. P., & Fatqurhohman, F. (2026). Analisis Sentimen Ulasan Pengguna Livin' by Mandiri Menggunakan Model Deep Transformer IndoBERT. *JINTEN: Journal of Intelligent Systems and Digital Science*, 1(2), 107-120.

Copyright ©2026 to the Author. Published by CV. Ihsan Cahaya Pustaka

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



PENDAHULUAN

Transformasi layanan perbankan digital telah mengubah pola interaksi masyarakat dengan layanan keuangan. Aplikasi *mobile banking* berkembang dari sekadar sarana transaksi menjadi kanal layanan yang menyediakan berbagai kebutuhan perbankan dalam satu platform. Kualitas aplikasi menjadi semakin penting karena pengalaman pengguna tidak hanya dipengaruhi oleh kelengkapan fitur, tetapi juga oleh kemudahan penggunaan, stabilitas sistem, kecepatan transaksi, dan keandalan layanan. Ulasan pengguna pada platform digital dapat merepresentasikan pengalaman tersebut secara langsung dan menyediakan informasi yang bernilai untuk mengevaluasi kualitas layanan. Mahmutovic (2026) menunjukkan bahwa ulasan pengguna *mobile banking* dapat digunakan untuk mengidentifikasi dimensi pengalaman dan faktor yang berkaitan dengan kepuasan pengguna.

Analisis sentimen menjadi salah satu pendekatan yang dapat digunakan untuk mengolah informasi dari ulasan pengguna dalam jumlah besar. Metode ini memungkinkan teks yang tidak

terstruktur diklasifikasikan berdasarkan kecenderungan sentimennya sehingga pola persepsi pengguna dapat diidentifikasi secara lebih sistematis. Basri (2024) menunjukkan bahwa data ulasan aplikasi perbankan seluler di Indonesia dapat dimanfaatkan untuk pemodelan topik dan analisis sentimen. Pemanfaatan analisis sentimen pada layanan *mobile banking* juga berkembang karena ulasan pengguna sering mengandung informasi mengenai kelebihan maupun permasalahan yang dialami selama menggunakan aplikasi. Informasi tersebut dapat menjadi dasar untuk memahami respons pengguna terhadap layanan digital.

Aplikasi Livin' by Mandiri merupakan salah satu objek yang relevan untuk dikaji melalui pendekatan analisis sentimen karena memiliki jumlah ulasan pengguna yang besar pada Google Play Store. Penelitian sebelumnya telah menerapkan beberapa metode pembelajaran mesin untuk menganalisis ulasan aplikasi tersebut. Sutrisno dan Putri (2026) menggunakan TF-IDF dan Naive Bayes untuk melakukan klasifikasi sentimen pada aplikasi Livin' by Mandiri. Ali dan Ismail (2025) membandingkan Neural Network, Random Forest, dan Support Vector Machine dalam menganalisis sentimen aplikasi Livin. Kurnia et al. (2025) juga mengevaluasi kinerja beberapa algoritma *machine learning* pada ulasan pengguna Livin' by Mandiri dari Google Play Store.

Pendekatan pembelajaran mesin berbasis fitur memiliki keterbatasan dalam merepresentasikan hubungan kontekstual antarkata. TF-IDF, misalnya, memberikan bobot berdasarkan kemunculan kata tanpa secara langsung memahami hubungan makna antarkata dalam suatu konteks kalimat. Karakteristik ulasan aplikasi yang singkat, informal, dan beragam dapat menyebabkan makna sentimen tidak selalu dapat ditentukan hanya berdasarkan kata-kata tertentu. Perkembangan model Transformer menawarkan pendekatan yang lebih kontekstual dalam merepresentasikan bahasa. Devlin et al. (2019) memperkenalkan BERT sebagai model berbasis Transformer yang memanfaatkan representasi dua arah untuk memahami hubungan kontekstual dalam teks melalui proses *pre-training* dan *fine-tuning*. Kokab et al. (2022) menunjukkan bahwa model Transformer memiliki kemampuan yang kuat dalam tugas analisis sentimen karena dapat menangkap representasi semantik dan konteks teks.

Kebutuhan terhadap model yang sesuai dengan karakteristik Bahasa Indonesia mendorong pengembangan berbagai sumber daya pemrosesan bahasa alami khusus Bahasa Indonesia. Wilie et al. (2020) memperkenalkan IndoNLU sebagai *benchmark* dan sumber daya untuk mengevaluasi pemahaman bahasa alami Bahasa Indonesia, termasuk penggunaan IndoBERT sebagai model bahasa berbasis Transformer. IndoBERT memungkinkan proses *fine-tuning* untuk berbagai tugas klasifikasi teks dengan memanfaatkan representasi bahasa yang telah dipelajari pada tahap *pre-training*. Pendekatan tersebut telah digunakan dalam berbagai penelitian analisis sentimen berbahasa Indonesia. Jayadianti et al. (2022) menunjukkan penerapan *fine-tuning* IndoBERT pada ulasan

berbahasa Indonesia, sedangkan Geni et al. (2023), Sayarizki et al. (2024), dan Putri et al. (2024) menerapkannya pada data media sosial dengan karakteristik bahasa yang beragam.

Penelitian terbaru memperlihatkan bahwa IndoBERT semakin banyak digunakan untuk menganalisis teks yang berasal dari berbagai platform digital. A'la (2025) menerapkan IndoBERT untuk klasifikasi sentimen ulasan permainan berbahasa Indonesia dengan mempertimbangkan ketidakseimbangan kelas. Apriliani et al. (2025) menggunakan *fine-tuned* IndoBERT untuk analisis sentimen berbasis aspek pada ulasan hotel berbahasa Indonesia. Supriyadi dan Makatita (2025) memanfaatkan IndoBERT untuk menganalisis komentar pengguna mengenai adopsi QRIS di Indonesia. Aqilah et al. (2026) menerapkan pendekatan berbasis IndoBERT untuk analisis sentimen pada ulasan aplikasi berbahasa Indonesia, sedangkan Triawan et al. (2026) mengembangkan kerangka IndoBERT yang diperkaya fitur untuk analisis sentimen ulasan layanan *ride-hailing*. Perkembangan tersebut menunjukkan bahwa kemampuan representasi kontekstual IndoBERT relevan untuk mengolah teks pengguna yang memiliki variasi bahasa dan konteks.

Domain perbankan digital juga mulai menjadi objek penerapan IndoBERT dalam analisis sentimen. Simarmata dan Sasongko (2025) menerapkan IndoBERT pada ulasan aplikasi BRImo dan menunjukkan penggunaan model Transformer untuk mengklasifikasikan sentimen pengguna aplikasi perbankan. Assani et al. (2025) juga memanfaatkan pendekatan NLP untuk menganalisis ulasan *mobile banking* dalam konteks peningkatan kualitas layanan digital. Penelitian pada aplikasi perbankan seluler di negara lain turut menunjukkan bahwa *user-generated reviews* dapat digunakan untuk mengevaluasi pengalaman pengguna dan kualitas layanan (Mahmutovic, 2026). Kajian *online review* pada aplikasi *mobile banking* Korea juga memperlihatkan pemanfaatan analisis sentimen untuk mengidentifikasi persepsi pengguna terhadap berbagai aspek layanan (Online Review Based IPA and IPCA, 2024).

Kesenjangan penelitian terlihat pada penerapan model Transformer terhadap ulasan pengguna Livin' by Mandiri. Penelitian yang secara khusus membahas aplikasi Livin' by Mandiri masih didominasi pendekatan TF-IDF dan algoritma *machine learning*, seperti Naive Bayes, Support Vector Machine, Random Forest, dan Neural Network (Sutrisno & Putri, 2026; Ali & Ismail, 2025; Kurnia et al., 2025). Penelitian IndoBERT pada domain perbankan digital telah dilakukan pada aplikasi lain, seperti BRImo, tetapi belum secara khusus menguji IndoBERT pada ulasan pengguna Livin' by Mandiri dalam penelitian yang menjadi fokus artikel ini (Simarmata & Sasongko, 2025). Kondisi tersebut menunjukkan adanya ruang untuk menguji kemampuan model Transformer berbahasa Indonesia dalam memahami konteks sentimen pada ulasan aplikasi Livin' by Mandiri.

Penelitian ini menerapkan *fine-tuning* IndoBERT untuk melakukan klasifikasi sentimen biner pada ulasan pengguna aplikasi Livin' by Mandiri. Data penelitian bersumber dari kumpulan ulasan

pengguna aplikasi yang tersedia pada Google Play Store dan diklasifikasikan ke dalam sentimen positif dan negatif berdasarkan skema pelabelan yang digunakan dalam penelitian. Model IndoBERT digunakan untuk mempelajari representasi kontekstual ulasan sehingga proses klasifikasi tidak hanya bergantung pada kemunculan kata secara individual. Kinerja model dievaluasi menggunakan accuracy, precision, recall, dan F1-score untuk memperoleh gambaran menyeluruh mengenai kemampuan klasifikasi. Penelitian ini diharapkan memberikan kontribusi empiris terhadap penerapan model Transformer berbahasa Indonesia pada analisis sentimen layanan perbankan digital serta memberikan informasi yang dapat mendukung pemahaman terhadap persepsi pengguna Livin' by Mandiri.

METODE

Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimental untuk membangun model klasifikasi sentimen ulasan pengguna aplikasi Livin' by Mandiri. Model yang digunakan adalah IndoBERT (Indonesian Bidirectional Encoder Representations from Transformers) melalui proses fine-tuning untuk mengklasifikasikan ulasan ke dalam dua kelas, yaitu positif dan negatif. Tahapan penelitian mencakup persiapan data, pembentukan label, analisis eksploratif, pengambilan sampel, pembagian dataset, tokenisasi, fine-tuning, dan evaluasi model. Penggunaan IndoBERT didasarkan pada kemampuan model Transformer dalam membangun representasi bahasa secara kontekstual serta ketersediaannya sebagai model yang dikembangkan untuk pemrosesan bahasa Indonesia (Devlin et al., 2019; Wilie et al., 2020).

Dataset dan Preprocessing

Data penelitian menggunakan dataset Livin' by Mandiri App Reviews yang tersedia pada Kaggle dan berasal dari ulasan pengguna aplikasi. Tahap preprocessing dilakukan dengan menghapus data yang tidak memiliki isi pada kolom ulasan, mengubah ulasan ke format teks, serta menghapus spasi berlebih dan ulasan kosong. Rating 3 dikeluarkan karena penelitian menggunakan klasifikasi biner, sedangkan rating 1–2 dikategorikan sebagai negatif (0) dan rating 4–5 sebagai positif (1). Tahapan tersebut dilakukan untuk menghasilkan data ulasan yang dapat digunakan dalam proses klasifikasi sentimen.

Analisis Eksploratif Data

Analisis eksploratif (Exploratory Data Analysis/EDA) dilakukan untuk mengetahui karakteristik data sebelum pemodelan. Analisis mencakup distribusi rating, distribusi kelas sentimen, panjang ulasan berdasarkan jumlah kata, serta rata-rata jumlah thumbs up pada masing-masing kelas.

WordCloud juga digunakan untuk memvisualisasikan kata-kata yang paling sering muncul pada ulasan positif dan negatif sehingga karakteristik awal data dapat diamati sebelum proses fine-tuning.

Sampling dan Pembagian Dataset

Sebanyak 15.000 ulasan dipilih dari data yang telah diberi label menggunakan stratified sampling dengan random state 42. Penggunaan stratified sampling bertujuan mempertahankan proporsi kelas sentimen dalam sampel penelitian. Dataset kemudian dibagi secara terstratifikasi menjadi 70% data latih (10.500 ulasan), 20% data validasi (3.000 ulasan), dan 10% data uji (1.500 ulasan). Data validasi digunakan untuk memantau kinerja selama pelatihan dan menentukan model terbaik, sedangkan data uji digunakan untuk evaluasi akhir.

Tokenisasi IndoBERT

Model yang digunakan adalah IndoBERT Base P1 dengan identifier indobenchmark/indobert-base-p1. Tokenisasi dilakukan menggunakan tokenizer dari Hugging Face Transformers dengan panjang maksimum 128 token. Setiap ulasan dikonversi menjadi `input_ids` dan `attention_mask`, sedangkan `truncation` digunakan untuk teks yang melebihi panjang maksimum dan `padding` diterapkan untuk menyamakan panjang input. Tokenisasi tersebut menghasilkan representasi yang sesuai dengan format masukan model IndoBERT untuk proses klasifikasi.

Fine-Tuning IndoBERT

Model IndoBERT yang telah melalui tahap pre-training digunakan sebagai model dasar dan disesuaikan dengan data ulasan *Living by Mandiri* melalui proses fine-tuning. Pendekatan fine-tuning memungkinkan model bahasa yang telah memperoleh representasi umum Bahasa Indonesia disesuaikan dengan tugas klasifikasi sentimen tertentu, sebagaimana telah diterapkan dalam penelitian analisis sentimen berbasis IndoBERT (Jayadianti et al., 2022). Pelatihan dilakukan selama 3 epoch, dengan batch size 16 untuk data latih dan batch size 32 untuk evaluasi, warm-up steps 100, serta weight decay 0,01. Model terbaik ditentukan berdasarkan F1-score validasi.

Evaluasi Model

Evaluasi model dilakukan menggunakan 1.500 data uji yang tidak digunakan selama proses pelatihan. Kinerja klasifikasi diukur menggunakan accuracy, precision, recall, dan F1-score untuk memperoleh gambaran performa model secara keseluruhan dan pada masing-masing kelas. Confusion matrix digunakan untuk mengidentifikasi jumlah klasifikasi yang benar dan salah pada kelas positif dan negatif. Perkembangan model selama pelatihan juga dianalisis melalui validation loss, validation accuracy, dan validation F1-score.

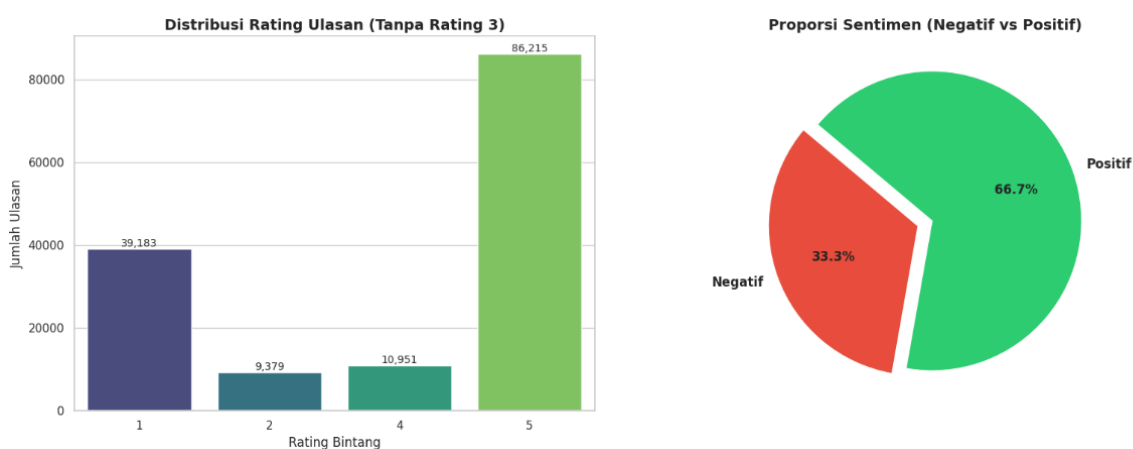
Pengujian Prediksi

Pengujian tambahan dilakukan dengan memberikan beberapa ulasan baru kepada model terbaik yang telah diperoleh melalui proses fine-tuning. Teks terlebih dahulu ditokenisasi menggunakan konfigurasi yang sama dengan tahap pelatihan, kemudian diproses oleh model untuk memperoleh probabilitas masing-masing kelas menggunakan softmax. Kelas dengan probabilitas tertinggi ditetapkan sebagai hasil prediksi sentimen. Tahap ini digunakan untuk menunjukkan kemampuan model dalam mengklasifikasikan teks baru di luar data uji.

HASIL DAN PEMBAHASAN

Distribusi Data Sentimen

Hasil preprocessing menunjukkan bahwa setelah ulasan dengan rating 3 dikeluarkan, data terbagi menjadi dua kelas sentimen, yaitu positif dan negatif. Kelas positif berjumlah 97.166 ulasan, sedangkan kelas negatif berjumlah 48.562 ulasan. Distribusi tersebut menunjukkan bahwa ulasan positif lebih dominan dibandingkan ulasan negatif dengan proporsi sekitar 66,7% dan 33,3%. Kondisi ini memperlihatkan bahwa data yang digunakan memiliki distribusi kelas yang tidak sepenuhnya seimbang, sehingga penggunaan *stratified sampling* menjadi penting agar proporsi kedua kelas tetap terjaga pada data eksperimen.



Gambar 1. Distribusi Rating dan Proporsi Sentimen Ulasan Livin' by Mandiri

Gambar 1 memperlihatkan 2 karakteristik utama data ulasan. Distribusi rating menunjukkan bahwa rating 5 memiliki jumlah ulasan paling tinggi dibandingkan rating lainnya, sedangkan rating 2 menjadi kategori dengan jumlah paling rendah. Pola tersebut menunjukkan bahwa distribusi rating tidak merata dan cenderung terkonsentrasi pada nilai ekstrem, khususnya rating 5 dan rating 1. Gambar 1(b) memperlihatkan bahwa kelas positif memiliki proporsi lebih besar daripada kelas negatif, yaitu 66,7% berbanding 33,3%. Ketidakseimbangan ini dipertahankan secara proporsional pada tahap pengambilan sampel menggunakan *stratified sampling* agar distribusi kedua kelas tetap terwakili dalam data latih, validasi, dan uji.



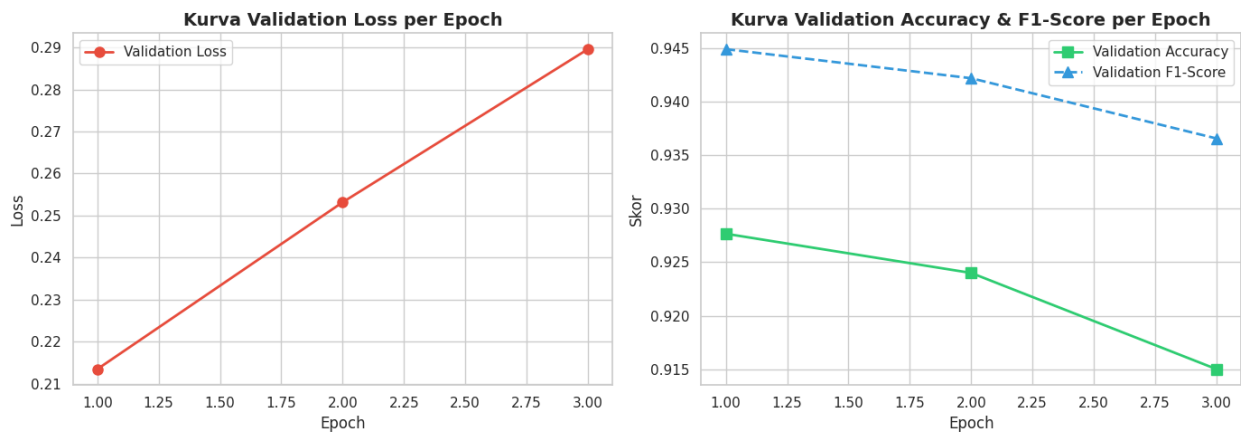
Pembagian Dataset

Hasil Fine-Tuning IndoBERT

Hasil pelatihan menunjukkan bahwa performa terbaik IndoBERT pada data validasi diperoleh pada epoch pertama. Pada tahap tersebut, validation loss mencapai 0,213454, validation accuracy sebesar 0,927667, F1-score sebesar 0,944938, precision sebesar 0,959299, dan recall sebesar 0,931000. Pada epoch kedua, validation loss meningkat menjadi 0,253169, sedangkan accuracy dan F1-score masing-masing menurun menjadi 0,924000 dan 0,942220. Penurunan kembali terjadi pada epoch ketiga dengan validation loss sebesar 0,289600, accuracy 0,915000, dan F1-score 0,936583.

Tabel 2. Kinerja IndoBERT pada Setiap Epoch

Epoch	Training Loss	Validation Loss	Accuracy	F1-Score	Precision	Recall
1	0,269890	0,213454	0,927667	0,944938	0,959299	0,931000
2	0,153684	0,253169	0,924000	0,942220	0,955293	0,929500
3	0,156319	0,289600	0,915000	0,936583	0,931717	0,941500



Gambar 4. Kurva Validation Loss, Validation Accuracy, dan Validation F1-Score

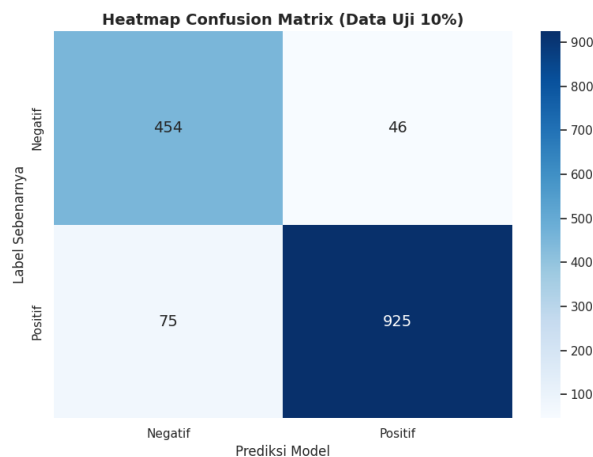
Gambar 4 memperlihatkan bahwa validation loss meningkat secara bertahap dari epoch pertama hingga epoch ketiga, sedangkan validation accuracy dan F1-score menunjukkan kecenderungan menurun. Pola tersebut menunjukkan bahwa epoch pertama memberikan kinerja validasi terbaik, sementara penambahan epoch tidak memberikan peningkatan performa pada data validasi. Kondisi ini mengindikasikan gejala *overfitting* ringan pada tahap pelatihan berikutnya. Oleh karena itu, pemilihan *checkpoint* terbaik berdasarkan F1-score validasi menjadi penting untuk mempertahankan model dengan kemampuan generalisasi terbaik. Hasil ini sejalan dengan penelitian sebelumnya yang menunjukkan pentingnya strategi *fine-tuning* dan pemilihan model dalam penerapan IndoBERT untuk klasifikasi sentimen (Jayadianti et al., 2022; A'la, 2025).

Evaluasi Model pada Data Uji

Evaluasi akhir dilakukan menggunakan 1.500 data uji. IndoBERT menghasilkan accuracy sebesar 0,92, dengan macro F1-score sebesar 0,91 dan weighted F1-score sebesar 0,92. Performa masing-masing kelas menunjukkan bahwa kelas negatif memperoleh precision 0,86, recall 0,91, dan F1-score 0,88, sedangkan kelas positif memperoleh precision 0,95, recall 0,93, dan F1-score 0,94. Hasil tersebut menunjukkan bahwa model mampu mengklasifikasikan kedua kelas dengan kinerja yang baik, dengan performa kelas positif sedikit lebih tinggi dibandingkan kelas negatif.

Tabel 3. Hasil Evaluasi IndoBERT pada Data Uji

Kelas	Precision	Recall	F1-Score	Support
Negatif	0,86	0,91	0,88	500
Positif	0,95	0,93	0,94	1.000
Accuracy				0,92
Macro Average	0,91	0,92	0,91	1.500
Weighted Average	0,92	0,92	0,92	1.500



Gambar 5. Confusion Matrix

Gambar 5 menunjukkan pola prediksi model pada data uji. Sebanyak 454 ulasan negatif berhasil diklasifikasikan sebagai negatif, sedangkan 46 ulasan negatif diprediksi sebagai positif. Pada kelas positif, 925 ulasan berhasil diklasifikasikan sebagai positif dan 75 ulasan diprediksi sebagai negatif. Kesalahan klasifikasi lebih banyak terjadi pada ulasan positif yang diprediksi sebagai negatif. Pola tersebut sesuai dengan hasil evaluasi per kelas, terutama precision negatif yang lebih rendah daripada precision positif. Secara keseluruhan, nilai F1-score 0,88 pada kelas negatif dan 0,94 pada kelas positif menunjukkan bahwa model mampu mempertahankan keseimbangan antara precision dan recall pada kedua kelas.

Pengujian pada Ulasan Baru

Pengujian terhadap ulasan baru dilakukan untuk melihat kemampuan model dalam memberikan prediksi pada teks yang berada di luar data uji. Model memprediksi ulasan “Aplikasi Livin yang baru sangat bagus dan mempermudah transaksi sehari-hari!” sebagai positif dengan probabilitas 98,26%, sedangkan ulasan mengenai aplikasi yang sering *crash* ketika melakukan transfer diprediksi sebagai negatif dengan probabilitas 95,98%. Ulasan yang memuji tampilan antarmuka memperoleh probabilitas positif 98,44%, sedangkan keluhan terhadap kegagalan fitur login sidik jari diprediksi negatif dengan probabilitas 95,04%.

Hasil pengujian tersebut memperlihatkan bahwa IndoBERT mampu memberikan klasifikasi yang konsisten pada contoh ulasan dengan ekspresi sentimen yang eksplisit. Ulasan yang mengandung penilaian positif seperti “bagus”, “mempermudah”, dan “keren” dikategorikan positif, sedangkan ekspresi keluhan seperti “crash”, “mengecewakan”, dan “gagal” dikategorikan negatif. Meskipun pengujian ini bersifat ilustratif dan tidak menggantikan evaluasi pada data uji, hasil probabilitas yang tinggi menunjukkan bahwa model memiliki tingkat keyakinan yang kuat pada contoh teks yang diberikan.

Pembahasan

Hasil penelitian menunjukkan bahwa IndoBERT mampu menghasilkan performa yang baik untuk klasifikasi sentimen ulasan pengguna Livin' by Mandiri. Accuracy sebesar 92% dan weighted F1-score sebesar 92% menunjukkan bahwa model mampu mempertahankan kinerja yang relatif konsisten pada data uji, sedangkan macro F1-score sebesar 91% menunjukkan bahwa performa kedua kelas tetap cukup seimbang. Temuan ini mendukung penelitian Jayadianti et al. (2022) yang menunjukkan efektivitas *fine-tuning* IndoBERT untuk analisis sentimen berbahasa Indonesia serta penelitian yang lebih baru pada berbagai jenis ulasan digital (Maysara, 2026; Aqiilah et al., 2026). Hasil tersebut juga sejalan dengan penelitian Simarmata dan Sasongko (2025) yang menerapkan IndoBERT pada ulasan aplikasi BRImo dalam konteks perbankan digital.

Perbedaan penelitian ini dengan penelitian terdahulu pada Livin' by Mandiri terletak pada pendekatan representasi bahasa yang digunakan. Sutrisno dan Putri (2026) menggunakan TF-IDF dan Naive Bayes, sedangkan Ali dan Ismail (2025) serta Kurnia et al. (2025) mengevaluasi beberapa algoritma *machine learning* untuk analisis sentimen Livin' by Mandiri. Pendekatan tersebut bergantung pada representasi fitur dan algoritma klasifikasi yang dibangun secara terpisah. IndoBERT menawarkan representasi kontekstual melalui arsitektur Transformer sehingga hubungan antarkata dalam teks dapat dipertimbangkan secara lebih komprehensif (Devlin et al., 2019; Wilie et al., 2020). Hasil penelitian ini memperlihatkan bahwa pendekatan tersebut dapat diterapkan secara efektif pada ulasan pengguna aplikasi perbankan digital.

Distribusi data yang didominasi oleh kelas positif menjadi aspek penting dalam interpretasi hasil. Jumlah ulasan positif yang lebih besar berpotensi membuat model memperoleh lebih banyak contoh pola bahasa positif selama proses pembelajaran. Kondisi tersebut tercermin pada precision dan F1-score kelas positif yang lebih tinggi daripada kelas negatif. Meskipun demikian, recall kelas negatif mencapai 0,91 sehingga sebagian besar ulasan negatif tetap berhasil dikenali. Temuan ini menunjukkan bahwa model mampu menangani kedua kelas dengan cukup baik, walaupun masih terdapat ruang untuk meningkatkan kemampuan dalam membedakan pola ulasan negatif yang lebih ambigu.

Secara keseluruhan, penelitian membuktikan bahwa IndoBERT dapat digunakan sebagai pendekatan Transformer untuk klasifikasi sentimen biner pada ulasan pengguna Livin' by Mandiri. Model memperoleh accuracy 92% pada data uji dan menunjukkan kemampuan prediksi dengan tingkat keyakinan tinggi pada beberapa ulasan baru. Hasil tersebut memperkuat posisi IndoBERT sebagai salah satu pendekatan yang potensial untuk analisis teks berbahasa Indonesia, khususnya pada data *user-generated content* dalam layanan perbankan digital. Pengembangan berikutnya dapat diarahkan pada data yang lebih besar, penanganan ketidakseimbangan kelas secara lebih eksplisit, serta analisis sentimen berbasis aspek untuk mengidentifikasi komponen layanan yang paling banyak memengaruhi pengalaman pengguna.

KESIMPULAN

Penelitian ini berhasil menerapkan model IndoBERT melalui proses *fine-tuning* untuk melakukan klasifikasi sentimen biner pada ulasan pengguna aplikasi Livin' by Mandiri. Data penelitian diproses dengan mengeluarkan rating 3 sebagai kategori netral, kemudian rating 1–2 dikategorikan sebagai negatif dan rating 4–5 sebagai positif. Sebanyak 15.000 ulasan dipilih secara *stratified sampling* dan dibagi menjadi 70% data latih, 20% data validasi, dan 10% data uji. Pendekatan tersebut memungkinkan model mempelajari karakteristik bahasa pada data ulasan sekaligus mempertahankan distribusi kelas pada setiap subset penelitian.

Hasil evaluasi menunjukkan bahwa IndoBERT mampu memberikan performa klasifikasi yang baik pada data uji. Model memperoleh accuracy sebesar 92%, macro F1-score sebesar 91%, dan weighted F1-score sebesar 92%. Performa kelas positif mencapai precision 0,95, recall 0,93, dan F1-score 0,94, sedangkan kelas negatif memperoleh precision 0,86, recall 0,91, dan F1-score 0,88. Proses pelatihan menunjukkan bahwa epoch pertama menghasilkan performa validasi terbaik dengan accuracy 0,927667 dan F1-score 0,944938, sementara validation loss meningkat dan performa validasi menurun pada epoch berikutnya. Kondisi tersebut menunjukkan adanya indikasi *overfitting* setelah epoch pertama dan memperkuat pentingnya pemilihan model berdasarkan performa validasi.

Temuan penelitian menunjukkan bahwa IndoBERT memiliki kemampuan yang baik dalam mengenali pola sentimen pada ulasan pengguna Livin' by Mandiri dan menghasilkan prediksi yang konsisten pada contoh ulasan baru. Model mampu memberikan probabilitas prediksi yang tinggi pada ulasan yang secara eksplisit menyampaikan pengalaman positif maupun keluhan terhadap aplikasi. Hasil tersebut menunjukkan bahwa pendekatan berbasis Transformer dapat menjadi alternatif yang efektif untuk analisis opini pengguna pada layanan perbankan digital. Penelitian selanjutnya dapat mengembangkan pendekatan ini melalui pemanfaatan data yang lebih besar, perbandingan dengan model lain, serta analisis sentimen berbasis aspek untuk mengidentifikasi fitur atau layanan yang secara spesifik menjadi sumber kepuasan dan keluhan pengguna.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada seluruh pihak yang telah memberikan dukungan dalam pelaksanaan penelitian dan penyusunan artikel ini, khususnya pihak yang mendukung penyediaan dan pemanfaatan dataset ulasan pengguna aplikasi Livin' by Mandiri serta pengembangan dan penggunaan sumber daya komputasi, perangkat lunak, dan model IndoBERT yang digunakan dalam penelitian. Apresiasi juga disampaikan kepada institusi dan pihak-pihak yang memberikan dukungan akademik, arahan, serta masukan selama proses penelitian hingga penyelesaian artikel ini. Dukungan tersebut berperan penting dalam membantu pelaksanaan pengolahan data, pengembangan model, analisis hasil, dan penyempurnaan naskah penelitian.

DAFTAR PUSTAKA

- A'la, F. Y. (2025). Optimasi klasifikasi sentimen ulasan game berbahasa Indonesia: IndoBERT dan SMOTE untuk menangani ketidakseimbangan kelas. *Edumatic: Jurnal Pendidikan Informatika*, 9(1), 256–265. <https://doi.org/10.29408/edumatic.v9i1.29666>
- Ali, M. H., & Ismail, J. (2025). Analisis sentimen aplikasi Livin dengan algoritma Neural Network, Random Forest, dan SVM. *Jurnal Penelitian Ilmu Komputer*, 3(2), 58–66. <https://jpilkom.org/index.php/journal/article/view/84>
- Aqiilah, I. N., Saptono, R., & Syaifuddin, A. (2026). Sentence-level sentiment analysis of Indonesian app reviews using IndoBERTweet. *Journal of Computing Theories and Applications*, 4(1), 74–91. <https://doi.org/10.62411/jcta.16240>
- Assani, M. Y., Rudiyan, A., & Radiyah, U. (2025). NLP-based sentiment analysis of Islamic mobile banking reviews for digital service quality enhancement. *Jurnal Bumigora Information Technology (BITe)*, 7(2), 131–140. <https://doi.org/10.30812/bite.v7i2.5943>
- Basri, M. H. (2024). Pemodelan topik dan analisis sentimen pada teks ulasan pengguna aplikasi perbankan seluler di Indonesia. *The Indonesian Journal of Computer Science*, 13(4). <https://doi.org/10.33022/ijcs.v13i4.4200>
- 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*, 1, 4171–4186. <https://arxiv.org/abs/1810.04805>
- Geni, L., Yulianti, E., & Sensuse, D. I. (2023). Sentiment analysis of tweets before the 2024 elections in Indonesia using BERT language models. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, 9(3), 746–757. <https://doi.org/10.26555/jiteki.v9i3.26490>

- 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>
- Kokab, S. T., Asghar, S., & Naz, S. (2022). Transformer-based deep learning models for the sentiment analysis of social media data. *Array*, 14, 100157. <https://doi.org/10.1016/j.array.2022.100157>
- Kurnia, O. D., Daniati, E., & Ristyawan, A. (2025). Kinerja algoritma machine learning dalam analisis sentimen ulasan pengguna aplikasi Livin' by Mandiri pada Google Play Store. *Jurnal Mahasiswa Sistem Informasi (JMSI)*, 6(2), 223–231. <https://doi.org/10.24127/jmsi.v6i2.8993>
- Mahmutovic, K. (2026). Analyzing user-generated reviews to identify experience dimensions and their impact on satisfaction with mobile banking applications. *Ekonomski Vjesnik/Econviews – Review of Contemporary Business, Entrepreneurship and Economic Issues*, 38(2). <https://doi.org/10.51680/ev.38.2.6>
- Maysara, A. L. (2026). Optimalisasi akurasi dan stabilitas analisis sentimen ulasan e-commerce Indonesia melalui fine-tuning Transformer IndoBERT. *Infotekmesin*, 17(1), 9–17. <https://doi.org/10.35970/infotekmesin.v17i1.3037>
- Online review based IPA and IPCA: The case of Korean mobile banking apps. (2024). *International Journal of Bank Marketing*, 43(4), 731–756. <https://doi.org/10.1108/IJBM-03-2024-0136>
- Simarmata, A. A. P., & Sasongko, T. B. (2025). Sentiment analysis on BRImo application reviews using IndoBERT. *Journal of Applied Informatics and Computing*, 9(3), 851–862. <https://doi.org/10.30871/jaic.v9i3.8162>
- Sutrisno, J., & Putri, J. G. E. (2026). Analisis sentimen pada aplikasi Livin' by Mandiri menggunakan metode TF-IDF dan Naive Bayes. *Jiems (Journal of Industrial Engineering and Management Systems)*, 17(1). <https://doi.org/10.30813/jiems.v17i1.8931>
- Triawan, P., Tahyudin, I., & Purwadi. (2026). A hybrid feature-enriched IndoBERT framework for sentiment analysis of ride-hailing service reviews in Indonesia. *Journal of Information Systems and Informatics*, 8(2). <https://doi.org/10.63158/journalisi.v8i2.1587>
- 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>

Sumber Dataset

Ahmadijaya, G. (2023). *Livin' by Mandiri app reviews* [Data set]. Kaggle.
<https://www.kaggle.com/datasets/itanium/livin-by-mandiri-app-reviews>