JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9779
Hal 880-888
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Analisis Sentimen Ulasan Mobile Banking Bank Kalbar pada Google Play Store Menggunakan IndoBert

Hasrul Rahman*, Hanafi

Fakultas Ilmu Komputer, PJJ Informatika, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia
Email: [1*]hasrul@students.amikom.ac.id, [2]hanafi@amikom.ac.id
Email Penulis Korespondensi: hasrul@students.amikom.ac.id
Submitted **07-05-2026;** Accepted **19-05-2026;** Published **30-06-2026**

Abstrak

Ulasan pengguna pada Google Play Store dapat menjadi sumber data penting untuk memahami persepsi masyarakat terhadap layanan mobile banking. Saat tulisan ini dibuat nilai rata-rata review pengguna aplikasi adalah 3,2 artinya nilai tersebut tergolong jelek sehingga perlu dikaji lagi oleh pihak bank hal-hal apa saja nanti yang dapat meningkatkan angka review tersebut untuk meminimalkan risiko reputasi karena kebanyakan pengguna yang akan menginstall aplikasi tentu melihat rating terlebih dahulu sebelum memutuskan mengintall aplikasi, dimana secara umum rating yang dianggap sangat baik adalah minimal 4,5. Penelitian ini bertujuan menganalisis sentimen ulasan pengguna aplikasi Mobile Banking Bank Kalbar dengan pendekatan pemrosesan bahasa alami menggunakan IndoBERT serta dua model pembanding, yaitu TF-IDF dengan Logistic Regression dan RNN BiLSTM. Dataset yang digunakan berjumlah 2.465 ulasan dengan tiga kelas sentimen, yaitu positif, negatif, dan netral. Distribusi data menunjukkan ketidakseimbangan kelas, dengan 1.445 ulasan positif, 894 ulasan negatif, dan 126 ulasan netral. Data dibagi secara stratified menjadi 70% data latih dan 30% data uji. Tahapan penelitian meliputi pembersihan teks, pembagian data, pelatihan model, evaluasi menggunakan accuracy, macro-F1, precision, recall, dan confusion matrix. Hasil eksperimen menunjukkan bahwa model TF-IDF + Logistic Regression memperoleh performa terbaik dengan accuracy 0,8581 dan macro-F1 0,6777. Model RNN BiLSTM menghasilkan accuracy 0,8311 dan macro-F1 0,6777, sedangkan IndoBERT memperoleh accuracy 0,8041 dan macro-F1 0,6774. Walaupun IndoBERT tidak menghasilkan accuracy tertinggi, model ini menunjukkan kemampuan lebih baik dalam mengenali kelas netral berdasarkan nilai recall 0,6316. Hasil ini mengindikasikan bahwa pada dataset yang relatif kecil dan tidak seimbang, model klasik berbasis TF-IDF masih dapat memberikan performa kompetitif dibandingkan model deep learning dan transformer.

Kata Kunci: Analisis Sentimen; Mobile Banking; Bank Kalbar; Google Play Store; IndoBERT

Abstract

User reviews on the Google Play Store can serve as an important data source for understanding public perceptions of mobile banking services. At the time this paper was written, the application's average user review score was 3.2, indicating a poor rating. Therefore, the bank needs to further examine the factors that could improve this rating in order to minimize reputational risk, since most users who intend to install an application tend to check its rating first before deciding whether to install it. In general, a rating considered very good is at least 4.5. This study aims to analyze user review sentiment toward the Bank Kalbar Mobile Banking application using a natural language processing approach with IndoBERT, along with two comparison models: TF-IDF with Logistic Regression and RNN BiLSTM. The dataset consists of 2,465 reviews classified into three sentiment classes: positive, negative, and neutral. The data distribution shows class imbalance, with 1,445 positive reviews, 894 negative reviews, and 126 neutral reviews. The data were split using a stratified method into 70% training data and 30% testing data. The research stages included text cleaning, data splitting, model training, and evaluation using accuracy, macro-F1, precision, recall, and a confusion matrix. The experimental results show that the TF-IDF + Logistic Regression model achieved the best performance, with an accuracy of 0.8581 and a macro-F1 score of 0.6777. The RNN BiLSTM model obtained an accuracy of 0.8311 and a macro-F1 score of 0.6777, while IndoBERT achieved an accuracy of 0.8041 and a macro-F1 score of 0.6774. Although IndoBERT did not achieve the highest accuracy, it demonstrated better capability in identifying the neutral class, as indicated by a recall score of 0.6316. These findings indicate that, for a relatively small and imbalanced dataset, a classical TF-IDF-based model can still deliver competitive performance compared with deep learning and transformer-based models.

Keywords: Sentiment Analysis; Mobile Banking; Bank Kalbar; Google Play Store; IndoBERT

1. PENDAHULUAN

Transformasi layanan perbankan ke kanal digital telah mengubah pola interaksi nasabah dengan bank, terutama melalui aplikasi mobile banking yang digunakan untuk cek saldo, transfer, pembayaran, pembelian, dan layanan transaksi lainnya. Pada konteks Bank Kalbar, aplikasi Mobile Banking menjadi salah satu kanal penting untuk mendukung transaksi yang lebih mudah, cepat, aman, dan efisien [1], [2]. Ulasan pada Google Play Store menjadi salah satu sumber data yang kaya karena memuat penilaian spontan pengguna terhadap masalah login, keterlambatan notifikasi, kegagalan transaksi, kestabilan aplikasi, serta persepsi terhadap kemudahan penggunaan. Bagi bank pembangunan daerah seperti Bank Kalbar, sumber data ini bernilai strategis karena dapat memperlihatkan suara nasabah yang sering kali tidak tertangkap secara utuh melalui survei formal. Dalam perspektif pengalaman pengguna layanan finansial digital, ulasan daring juga semakin sering dipakai untuk membaca determinan pengalaman pengguna dan kepuasan layanan perbankan digital berbasis AI maupun aplikasi pembayaran [3], [4]. Urgensi penelitian ini semakin kuat karena pada saat pengumpulan data, aplikasi Mobile Banking Bank Kalbar memperoleh rating rata-rata 3,2, yang menunjukkan perlunya evaluasi berbasis data terhadap persepsi pengguna untuk meminimalkan risiko reputasi dan meningkatkan kualitas layanan digital

Analisis sentimen menawarkan pendekatan sistematis untuk mengubah ulasan bebas menjadi informasi terstruktur yang dapat ditindaklanjuti. Berbagai studi sebelumnya menunjukkan bahwa teks ulasan aplikasi dapat digunakan untuk

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9779
Hal 880-888
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

memetakan kualitas layanan, kebutuhan pengguna, dan prioritas pengembangan sistem [5], [6], [7]. Pada layanan mobile banking Indonesia, sentimen negatif umumnya muncul pada isu verifikasi, pengiriman kode OTP, gangguan jaringan, kegagalan transaksi, dan keandalan aplikasi, sedangkan sentimen positif lebih sering berkaitan dengan kemudahan, kecepatan, dan kepraktisan layanan [5], [6], [8]. Karena itu, analisis sentimen tidak hanya relevan sebagai tugas klasifikasi teks, tetapi juga sebagai instrumen evaluasi pengalaman pengguna dan pengambilan keputusan bagi pengelola layanan digital.

Penelitian terkait pada domain perbankan dan aplikasi seluler memperlihatkan spektrum metode yang cukup luas. Permana dkk. menggabungkan klasifikasi sentimen dan pemodelan topik pada ulasan aplikasi mobile banking dan menunjukkan bahwa kendala kode OTP, login, serta koneksi jaringan merupakan topik dominan dalam sentimen negatif [6]. Adiningtyas dan Auliani mengukur kualitas layanan tiga aplikasi mobile banking besar di Indonesia melalui sentimen berbasis ulasan Google Play dan menemukan bahwa dimensi efisiensi, kepatuhan, ketersediaan sistem, serta privasi perlu mendapatkan perhatian serius [5]. Prasetyo dan Agastya menunjukkan bahwa pendekatan Support Vector Machine pada ulasan aplikasi perbankan di Google Play dapat memberikan akurasi yang tinggi ketika data dibersihkan dan distribusi latih-uji dijaga stabil [8]. Safarah dkk. juga menegaskan bahwa kombinasi sentimen dan topic modeling efektif untuk membaca kepuasan pengguna Livin Mandiri serta mengidentifikasi keluhan teknis yang berulang [7]. Pada kasus BRImo, Bimantara dan Zufria memperlihatkan bahwa kombinasi TF-IDF, NLP, dan SVM masih kompetitif untuk teks ulasan aplikasi perbankan [9].

Di sisi lain, perkembangan model bahasa berbasis transformer telah mendorong pemanfaatan representasi kontekstual yang lebih kuat. BERT menjadi fondasi penting bagi banyak tugas klasifikasi teks modern [10]. Untuk bahasa Indonesia, Koto dkk. memperkenalkan IndoBERT melalui IndoLEM [11], sedangkan Wilie dkk. menghadirkan sumber daya IndoNLU yang memperkuat ekosistem evaluasi untuk tugas pemahaman bahasa Indonesia [12]. Pengembangan berikutnya juga melahirkan IndoBERTweet untuk domain media sosial yang lebih informal [11], sementara inisiatif sumber daya terbuka seperti NusaCrowd menunjukkan bahwa ketersediaan korpus dan dataset Indonesia terus berkembang [13]. Sejumlah penelitian terbaru menunjukkan bahwa IndoBERT efektif pada analisis sentimen berbahasa Indonesia, baik untuk ulasan hotel [14], analisis berbasis aspek [15], ulasan e-commerce Google Play [16], maupun ulasan aplikasi BRImo [17]. Temuan-temuan tersebut menegaskan bahwa model pralatih monolingual cenderung lebih sesuai untuk teks Indonesia yang kaya variasi leksikal, bentuk tidak baku, dan campuran istilah layanan.

Meski demikian, masih terdapat celah penelitian yang relevan. Pertama, sebagian besar studi terdahulu berfokus pada bank nasional atau aplikasi digital dengan volume pengguna sangat besar, sedangkan aplikasi milik bank pembangunan daerah relatif jarang diteliti secara mendalam [5], [8], [7]. Kedua, beberapa penelitian memang telah memakai IndoBERT, tetapi objeknya bukan Bank Kalbar sehingga karakter keluhan pengguna, stabilitas aplikasi, dan konteks layanan lokal belum tercermin secara spesifik [17], [16], [14]. Ketiga, banyak penelitian menonjolkan hasil klasifikasi akhir tanpa membandingkan secara terbuka perilaku metode klasik, recurrent network, dan transformer pada korpus yang sama. Padahal, perbandingan seperti ini penting agar pemilihan model tidak hanya mengikuti tren, tetapi juga mempertimbangkan ukuran data, sifat teks, serta biaya komputasi.

Pemilihan metode dalam penelitian ini didasarkan pada kebutuhan untuk memperoleh perbandingan yang adil antara pendekatan klasik, deep learning, dan transformer. TF-IDF dengan Logistic Regression digunakan sebagai baseline klasik karena efisien, mudah diinterpretasikan, dan sesuai untuk dataset teks berukuran terbatas. RNN BiLSTM digunakan karena mampu mempelajari pola sekuensial dalam ulasan pengguna. IndoBERT digunakan karena memiliki representasi bahasa Indonesia yang lebih kontekstual dan relevan untuk menangani teks ulasan yang cenderung informal, tidak baku, serta mengandung variasi istilah teknis layanan perbankan. Dengan membandingkan ketiga pendekatan tersebut, penelitian ini dapat menunjukkan model yang paling sesuai untuk kondisi data ulasan Mobile Banking Bank Kalbar.

Kontribusi utama penelitian ini adalah menyediakan kajian empiris analisis sentimen terhadap ulasan pengguna Mobile Banking Bank Kalbar berbasis data Google Play Store dengan membandingkan tiga pendekatan klasifikasi, yaitu TF-IDF dengan Logistic Regression, RNN BiLSTM, dan IndoBERT. Penelitian ini juga memberikan kontribusi pada konteks objek kajian, karena berfokus pada aplikasi mobile banking bank pembangunan daerah yang masih jarang diteliti. Selain itu, penelitian ini menyajikan analisis performa model pada dataset tiga kelas sentimen yang tidak seimbang, sehingga dapat memberikan gambaran lebih realistis mengenai kelebihan dan keterbatasan masing-masing pendekatan dalam klasifikasi sentimen teks ulasan berbahasa Indonesia.

Berdasarkan uraian tersebut, tujuan penelitian ini adalah untuk: 1) menganalisis distribusi sentimen ulasan pengguna Mobile Banking Bank Kalbar pada Google Play Store; 2) membandingkan performa TF-IDF dengan Logistic Regression, RNN BiLSTM, dan IndoBERT berdasarkan accuracy, precision, recall, macro-F1, dan confusion matrix; 3) mengidentifikasi kecenderungan kata atau istilah dominan yang merepresentasikan sentimen positif, negatif, dan netral; serta 4) merumuskan implikasi hasil analisis sebagai dasar evaluasi kualitas layanan digital Mobile Banking Bank Kalbar. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan kontribusi akademik dalam pengembangan analisis sentimen berbahasa Indonesia serta kontribusi praktis bagi pengelola aplikasi dalam memprioritaskan perbaikan layanan berbasis persepsi pengguna..

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9779
Hal 880-888
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan sumber data berupa ulasan pengguna aplikasi Mobile Banking Bank Kalbar pada Google Play Store. Alur penelitian ditunjukkan pada Gambar 1 dan terdiri atas lima tahap utama, yaitu akuisisi data, pra-pemrosesan, pembagian data, pemodelan, serta evaluasi. Dalam konteks rekayasa data, alur tersebut sejalan dengan praktik umum analisis sentimen yang menuntut keterlacakan sejak sumber teks mentah sampai keluaran klasifikasi [18], [19]. Pendekatan eksperimental digunakan karena penelitian ini tidak hanya mendeskripsikan data ulasan, tetapi juga menguji dan membandingkan performa beberapa model klasifikasi sentimen pada dataset yang sama. Model yang dibandingkan terdiri atas TF-IDF dengan Logistic Regression, RNN BiLSTM, dan IndoBERT.

Dataset yang digunakan berisi 2.465 ulasan yang telah memiliki label sentimen tiga kelas, yaitu positif, negatif, dan netral. Struktur datanya memuat nama pengguna, waktu ulasan, isi ulasan, label, dan skor. Fokus penelitian ini diarahkan pada pengolahan teks dan evaluasi model, bukan pada desain skema anotasi dari awal. Distribusi kelas yang tidak seimbang tetap dicatat sebagai variabel penting karena jumlah kelas positif jauh lebih dominan dibanding kelas netral. Ketidakseimbangan seperti ini lazim pada data ulasan aplikasi dan dapat memengaruhi sensitivitas model terhadap kelas minoritas [16], [14].

Tabel 1. Tahapan penelitian Analisis Sentimen

Tahap	Input	Proses/Metode	Output
Akuisisi data	Ulasan pengguna Google Play Store	Pengumpulan data ulasan aplikasi Mobile Banking Bank Kalbar	Dataset awal ulasan pengguna
Pra-pemrosesan teks	Teks ulasan mentah	Case folding, pembersihan karakter, tokenisasi, stopword removal, dan lemmatisasi	Teks hasil pra-pemrosesan
Pembagian data	Dataset berlabel	Stratified split dengan proporsi 70% data latih dan 30% data uji	Data latih dan data uji
Pemodelan	Data latih	Pelatihan TF-IDF + Logistic Regression, RNN BiLSTM, dan IndoBERT	Model klasifikasi sentimen
Evaluasi	Data uji dan hasil prediksi	Accuracy, precision, recall, F1-score, macro-F1, weighted average, dan confusion matrix	Perbandingan performa model

2.2 Pra-pemrosesan dan Representasi Teks

Tahap pra-pemrosesan dilakukan secara berurutan. Pertama, seluruh teks diubah menjadi huruf kecil untuk menurunkan variasi token yang sebenarnya sama secara semantik. Kedua, spasi berlebih, tanda baca, dan karakter non-alfanumerik dihapus menggunakan ekspresi reguler. Ketiga, teks dipotong menjadi token kata. Keempat, stopword bahasa Indonesia dari pustaka NLTK dihapus agar model lebih fokus pada kata informatif. Kelima, peneliti menerapkan lemmatisasi untuk mengembalikan token ke bentuk dasar. Meskipun pendekatan ini sederhana, hasilnya cukup efektif untuk baseline berbasis TF-IDF. Akan tetapi, untuk transformer berbahasa Indonesia, normalisasi berlebihan perlu dipertimbangkan ulang karena dapat menghilangkan sinyal kontekstual, slang, atau variasi morfologi yang justru penting bagi IndoBERT [11], [12].

Setelah teks dibersihkan, data dipisah menjadi data latih dan data uji dengan proporsi 70:30 menggunakan strategi stratified split agar komposisi tiap kelas tetap relatif konsisten. Proporsi ini dipilih karena memberikan keseimbangan antara jumlah data untuk pembelajaran dan jumlah data yang cukup untuk evaluasi. Pada jalur TF-IDF, representasi teks dibentuk menggunakan unigram dan bigram dengan batas kosakata 15.000 fitur. Model klasifikasi yang digunakan adalah Logistic Regression. Jalur ini dipilih sebagai baseline klasik karena terkenal stabil, efisien, dan sangat kompetitif pada teks pendek [8], [9].

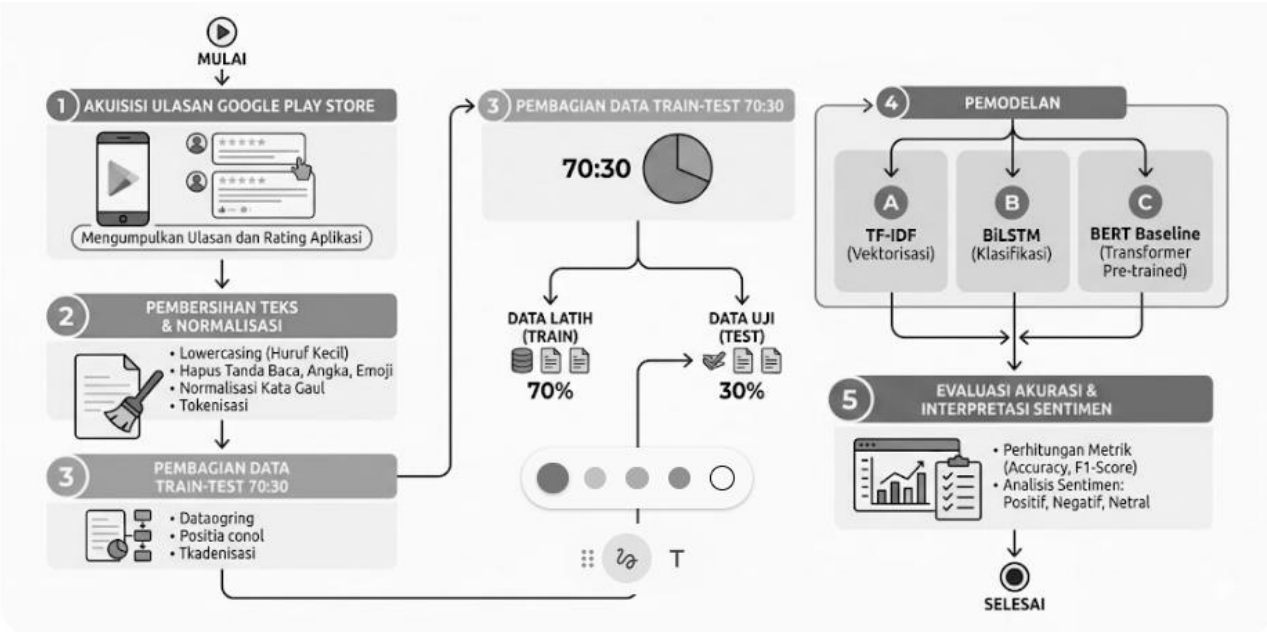
2.3 Pemodelan dan Evaluasi

Pertama, model TF-IDF + Logistic Regression digunakan sebagai baseline. TF-IDF berfungsi mengubah teks menjadi representasi numerik berdasarkan bobot kata, sedangkan Logistic Regression digunakan sebagai algoritma klasifikasi. Pada eksperimen ini, model menggunakan class_weight="balanced" untuk mengurangi dampak ketidakseimbangan kelas. Kedua, model RNN BiLSTM digunakan sebagai pendekatan deep learning. Model ini memanfaatkan embedding layer, Bidirectional LSTM, dropout, dense layer, dan output softmax untuk mengklasifikasikan tiga kelas sentimen. Tokenisasi dilakukan dengan jumlah kata maksimum 30.000 dan panjang sekuens maksimum 120 token. Model dilatih selama 10 epoch dengan batch size 32. Pilihan ini dimaksudkan untuk melihat apakah representasi sekuensial mampu menangkap pola sentimen lebih baik daripada model leksikal pada data ulasan yang singkat dan bising [20].

Ketiga, model IndoBERT digunakan sebagai pendekatan transformer. Model yang digunakan adalah indobenchmark/indobert-base-p1. Pada eksperimen, IndoBERT digunakan dengan panjang input maksimum 100 token dan classification head berbasis dense layer. IndoBERT dipilih karena dirancang untuk representasi bahasa Indonesia dan telah digunakan dalam berbagai tugas natural language understanding bahasa Indonesia. Dalam penelitian lanjutan, arsitektur IndoBERT dapat diganti dengan model IndoBERT-lite untuk waktu komputasi yang lebih singkat namun akurasi yang seharusnya hanya sedikit saja lebih rendah namun dalam mendapatkan kesesuaian antara bahasa sumber dan tokenizer tetap terjaga [11], [12].

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9779
Hal 880-888
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, F1-score, macro-F1, weighted average, dan confusion matrix. Macro-F1 digunakan sebagai indikator penting karena dataset memiliki distribusi kelas yang tidak seimbang. Accuracy saja tidak cukup karena model dapat terlihat tinggi performanya jika dominan benar pada kelas mayoritas, tetapi gagal mengenali kelas minoritas..



Gambar 1. Alur penelitian analisis sentimen ulasan Mobile Banking Bank Kalbar

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Dataset

Hasil deskriptif dataset menunjukkan bahwa sentimen positif mendominasi korpus, sedangkan sentimen netral menjadi kelas paling sedikit. Dominasi kelas positif mengindikasikan bahwa sebagian pengguna masih menilai aplikasi ini membantu aktivitas transaksi. Akan tetapi, kehadiran 721 ulasan negatif juga memperlihatkan bahwa masalah layanan digital belum dapat diabaikan. Kelas netral yang kecil memperbesar kemungkinan model kesulitan membedakan ulasan yang informatif tetapi tidak secara eksplisit menunjukkan emosi kuat. Pola seperti ini umum muncul pada ulasan aplikasi, terutama ketika pengguna menulis komentar singkat berbentuk saran atau deskripsi masalah tanpa kata evaluatif yang tegas [5], [8], [7].

Tabel 2 memperlihatkan distribusi kelas pada dataset. Persentase positif mencapai sekitar 58,62%, negatif 36,27%, dan netral 5,11%. Komposisi ini penting karena berpengaruh langsung terhadap perilaku model. Baseline yang sederhana sering kali lebih stabil pada data tidak seimbang apabila fitur leksikal masih cukup informatif, sementara model yang lebih kompleks memerlukan data yang lebih besar agar dapat menggeneralisasi dengan baik terhadap kelas minoritas [16], [14]. Oleh karena itu, hasil perbandingan berikut tidak boleh dibaca semata-mata sebagai kemenangan satu algoritma atas algoritma lain, tetapi sebagai interaksi antara sifat data, teknik representasi, dan kapasitas model.

Tabel 2. Distribusi sentimen pada dataset ulasan

Kelas Sentimen	Jumlah Ulasan	Persentase
Positif	1508	58,62%
Negatif	721	36,27%
Netral	236	5,11%

Distribusi tersebut menunjukkan adanya ketidakseimbangan kelas, terutama pada kelas netral yang jumlahnya jauh lebih sedikit dibandingkan kelas positif dan negatif. Oleh karena itu, penelitian ini tidak hanya menggunakan accuracy, tetapi juga macro-F1, precision, recall, dan confusion matrix agar performa model terhadap setiap kelas dapat dianalisis secara lebih objektif.

3.2 Perbandingan Model

Pada skenario pertama, model TF-IDF dengan Logistic Regression menghasilkan sebesar 0,8581 dan macro-F1 sebesar 0,6777. Pada kelas negatif, model memperoleh precision 0,8121, recall 0,9515, dan F1-score 0,8763. Pada kelas positif, model memperoleh precision 0,9610, recall 0,8525, dan F1-score 0,9035. Namun, performa pada kelas netral masih rendah, dengan precision 0,2439, recall 0,2632, dan F1-score 0,2532. Performa ini merupakan yang terbaik di antara

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9779
Hal 880-888
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

seluruh model. Secara metodologis, hasil tersebut cukup masuk akal. Ulasan aplikasi perbankan cenderung pendek, langsung menyebut gangguan tertentu, dan berisi kata kunci spesifik seperti login, gagal, SMS, pulsa, daftar, atau mutasi. Fitur n-gram mampu menangkap pola tersebut dengan sangat baik tanpa harus mempelajari ketergantungan kalimat yang panjang. Selain itu, Logistic Regression relatif tahan terhadap ukuran data menengah dan lebih mudah dioptimalkan dibanding model neural yang lebih berat [8], [9], [20].

Keunggulan baseline TF-IDF juga diperjelas oleh daftar kata berbobot tinggi pada tiap kelas. Untuk kelas negatif, token dominan yang muncul adalah "ribet", "aplikasi", "gangguan", "gagal", "login", "daftar", "pulsa", "sm", "susah", dan "gak". Kumpulan token ini menegaskan bahwa inti ketidakpuasan pengguna terletak pada hambatan akses awal, kegagalan autentikasi, konsumsi pulsa atau SMS, serta persepsi bahwa proses penggunaan aplikasi terlalu rumit. Pada kelas positif, kata-kata yang dominan adalah "mempermudah", "terbaik", "oke", "good", "memudahkan", "ok", "mantap", "bagus", "mudah", dan "membantu". Dengan kata lain, sentimen positif terutama berporos pada pengalaman fungsional yang lancar dan rasa terbantu saat melakukan transaksi. Kelas netral justru menampilkan token seperti "mutasi rekening", "gangguan jaringan", "rekening", "error", "mutasi", dan beberapa token yang tampak bising. Hal ini menunjukkan bahwa kelas netral bercampur antara laporan fitur, status gangguan, dan noise hasil pra-pemrosesan.

Skenario kedua menggunakan RNN BiLSTM. Accuracy sebesar 0,8311 dan macro-F1 sebesar 0,6777. Pada kelas negatif, model memperoleh precision 0,8992, recall 0,8657, dan F1-score 0,8821. Pada kelas positif, model memperoleh precision 0,9361, recall 0,8433, dan F1-score 0,8873. Pada kelas netral, model memperoleh precision 0,1868, recall 0,4474, dan F1-score 0,2636.

Dibandingkan TF-IDF + Logistic Regression, RNN BiLSTM memiliki recall kelas netral yang lebih tinggi. Artinya, model ini lebih banyak menangkap data netral yang sebelumnya sulit dikenali. Namun, precision kelas netral masih rendah, yang menunjukkan bahwa sebagian prediksi netral berasal dari kelas lain. Kondisi ini menyebabkan performa keseluruhan RNN BiLSTM belum mampu melampaui model TF-IDF + Logistic Regression. .

Dalam penelitian lanjutan, kondisi seperti ini dapat diperbaiki dengan menambah jumlah data, melakukan regularisasi yang lebih kuat, menggunakan embedding Indonesia yang relevan, atau langsung berpindah ke model transformer yang telah dipralatih untuk bahasa Indonesia [14], [15], [21].

Skenario ketiga adalah dengan Indobert. Hasilnya menunjukkan accuracy sebesar 0,8041 dan macro-F1 sebesar 0,6774. Pada kelas negatif, IndoBERT memperoleh precision 0,9064, recall 0,9030, dan F1-score 0,9047. Pada kelas positif, model memperoleh precision 0,9880, recall 0,7581, dan F1-score 0,8579. Pada kelas netral, model memperoleh precision 0,1714, recall 0,6316, dan F1-score 0,2697. Hasil ini merupakan yang terendah, sekaligus yang paling mahal secara komputasi. Bila hanya membaca angka, dapat timbul kesan bahwa transformer tidak cocok untuk kasus ini. Akan tetapi, kesimpulan demikian terlalu tergesa-gesa. Kinerja yang rendah lebih tepat dipahami sebagai akibat dari ketidakselarasan antara model pralatih dan karakter data. Kosakata Indobert yang digunakan bersifat formal, sedangkan korpus penelitian berupa ulasan Indonesia dengan kosakata perbankan lokal, campuran singkatan, dan bentuk informal. Dengan kata lain, baseline ini belum merepresentasikan potensi sebenarnya dari IndoBERT [10], [11], [12].

Perbedaan inilah yang menjadi landasan utama mengapa IndoBERT tetap relevan sebagai tema penelitian. IndoBERT dibangun dari korpus bahasa Indonesia dan dievaluasi pada beragam tugas pemahaman bahasa Indonesia [11], [12]. Untuk domain teks informal, keluarga model seperti IndoBERTweet juga menunjukkan keuntungan adaptasi kosakata terhadap variasi bahasa pengguna [21]. Penelitian lain yang secara eksplisit memakai IndoBERT pada ulasan Indonesia juga melaporkan hasil yang kompetitif, baik pada ulasan e-commerce [16], ulasan hotel [14], analisis berbasis aspek [15], maupun ulasan BRImo [17]. Oleh sebab itu, baseline transformer seharusnya dilihat sebagai bukti bahwa arsitektur BERT membutuhkan pasangan tokenizer dan pretraining yang tepat. Jika encoder dan tokenizer diganti ke IndoBERT, serta dilakukan fine-tuning penuh atau parsial, sangat mungkin performa transformer meningkat dan bahkan melampaui baseline TF-IDF. Argumen ini juga diperkuat oleh studi aplikasi layanan publik digital seperti SIGNAL yang memanfaatkan BERT untuk ulasan pengguna berbahasa Indonesia [22].

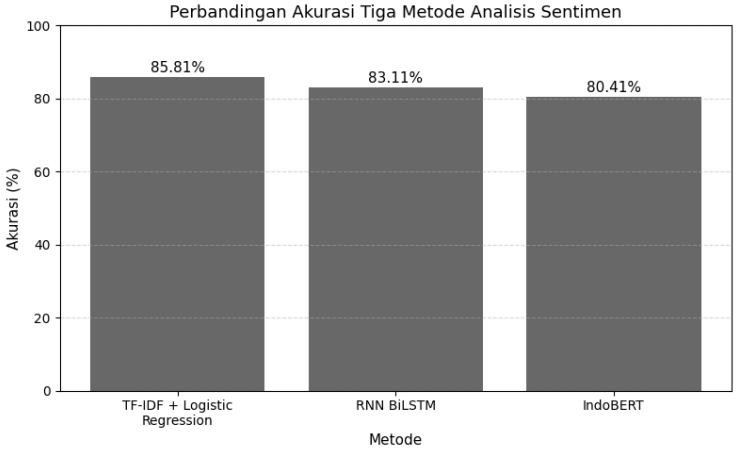
Tabel 3 merangkum perbandingan ketiga model. Dari sisi efisiensi, TF-IDF unggul mutlak karena melatih model dalam hitungan detik. Dari sisi akurasi, TF-IDF juga menempati posisi terbaik pada eksperimen saat ini. BiLSTM berada di tengah: lebih lambat dan kurang akurat daripada TF-IDF, tetapi masih lebih baik daripada baseline BERT yang tidak sesuai bahasa..

Tabel 3. Perbandingan hasil eksperimen

Metode	Akurasi Uji	Macro F1
TF-IDF + Logistic Regression	85,8%	0.677665
BiLSTM	83,1%	0.677656
IndoBERT	80,4%	0.677408

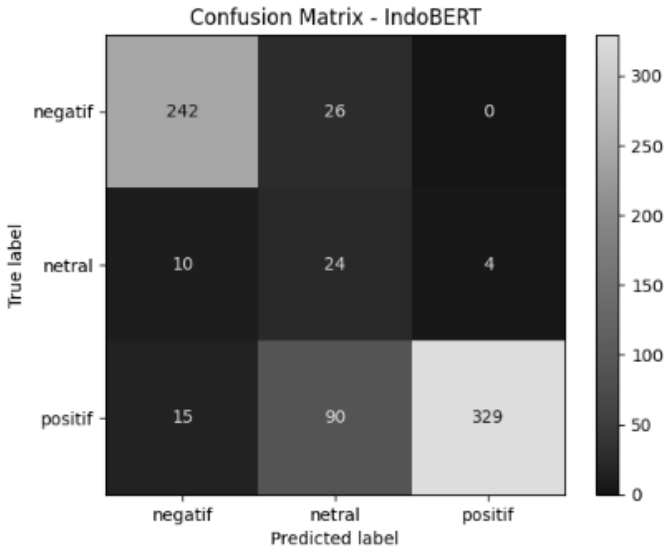
Gambar 2 menunjukkan perbandingan akurasi untuk ketiga model. baseline BERT menjadi pengingat bahwa model canggih tidak selalu lebih baik bila konfigurasi dasarnya tidak cocok dengan domain dan bahasa data. Dalam banyak riset terapan, pemilihan model harus mempertimbangkan biaya komputasi, ketersediaan data, kebutuhan interpretasi, dan kesesuaian model pralatih dengan bahasa target [18], [20].

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9779
Hal 880-888
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 2. Perbandingan akurasi ketiga metode

Secara substantif, temuan token dominan dari kelas negatif memperlihatkan tiga kelompok masalah utama pada Mobile Banking Bank Kalbar. Kelompok pertama adalah hambatan autentikasi dan akses, yang tercermin dari kata "login", "daftar", "gagal", dan "susah". Kelompok kedua adalah biaya atau mekanisme verifikasi berbasis SMS, yang tercermin dari kata "pulsa" dan "sm". Kelompok ketiga adalah persepsi kualitas sistem secara umum, misalnya "gangguan", "ribet", dan "aplikasi". Pola ini selaras dengan studi pada aplikasi perbankan lain di Indonesia yang juga menempatkan login, OTP, verifikasi, dan stabilitas aplikasi sebagai sumber utama sentimen negatif [5], [6], [17], [7]. Dengan demikian, masalah yang muncul pada Bank Kalbar bukan kasus yang terisolasi, melainkan bagian dari tantangan struktural mobile banking di Indonesia



Gambar 3.Confusion Matrix Indobert

3.3 Interpretasi Sentimen dan Implikasi

Di sisi positif, kata-kata seperti "mempermudah", "mudah", "membantu", dan "bagus" menunjukkan bahwa pengguna menghargai manfaat inti layanan ketika aplikasi berjalan sebagaimana mestinya. Artinya, nilai bisnis aplikasi tetap kuat: nasabah merasakan efisiensi dan kepraktisan bila proses transaksi lancar. Temuan ini penting bagi pengambilan keputusan. Fokus pengembangan tidak harus dimulai dari penambahan fitur baru, tetapi dari pengurangan friksi pada fitur yang sudah ada. Bila aspek aktivasi, login, notifikasi, dan transaksi dasar diperkuat, maka sentimen positif berpeluang meningkat tanpa investasi yang terlalu spekulatif pada fitur tambahan [5], [1], [[2]].

Dari perspektif akademik, penelitian ini memberi tiga implikasi. Pertama, pada dataset ulasan aplikasi berukuran menengah, baseline leksikal masih sangat relevan dan tidak boleh diabaikan. Kedua, interpretabilitas model sederhana justru menjadi keunggulan ketika tujuan penelitian mencakup identifikasi kata kunci keluhan. Ketiga, transformasi menuju IndoBERT tetap penting, tetapi harus dilakukan dengan desain eksperimen yang benar, yakni memakai tokenizer dan checkpoint Indonesia, menyesuaikan strategi fine-tuning, serta menilai performa bukan hanya dengan akurasi, tetapi juga precision, recall, macro F1-score, dan confusion matrix untuk melihat perilaku kelas netral [17], [16], [15].

Penelitian ini juga memiliki keterbatasan yang perlu dinyatakan secara eksplisit. Hasil transformer belum dapat diklaim sebagai hasil final IndoBERT karena evaluasi yang tersedia masih terfokus pada akurasi dan belum menyajikan

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9779
Hal 880-888
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

metrik per kelas. Tahap pelabelan juga tidak dibahas secara rinci di sehingga validitas anotasi perlu diperkuat pada studi berikutnya, misalnya melalui pedoman anotasi, pemeriksaan inter-annotator agreement, atau validasi manual sampel acak. Keterbatasan lain adalah belum adanya analisis aspek, padahal ulasan mobile banking sangat kaya dengan tema spesifik seperti login, transfer, notifikasi, mutasi, dan layanan pelanggan [6], [15].

Walaupun demikian, penelitian ini tetap memberikan dasar yang kokoh untuk studi lanjut. Secara praktis, Bank Kalbar dapat menggunakan hasil ini untuk memprioritaskan perbaikan pada proses aktivasi, autentikasi, dan stabilitas layanan berbasis SMS atau notifikasi. Secara metodologis, penelitian lanjutan dapat memulai dari dataset yang sama, lalu mengganti baseline transformer dengan IndoBERT, melakukan fine-tuning terkontrol, dan menambahkan analisis aspek agar keluhan pengguna dapat dipetakan lebih operasional. Dengan pendekatan tersebut, analisis sentimen tidak lagi berhenti pada klasifikasi positif-negatif-netral, tetapi berubah menjadi alat diagnosis mutu layanan digital yang lebih presisi.

Jika dilihat dari sudut pandang rekayasa sistem, keunggulan TF-IDF pada eksperimen ini juga memiliki nilai praktis yang besar. Model sederhana lebih mudah diintegrasikan ke dalam dashboard pemantauan ulasan karena proses pelatihannya cepat, pembaruan model dapat dilakukan berkala tanpa kebutuhan komputasi tinggi, dan hasil koefisiennya mudah dijelaskan kepada pemangku kepentingan nonteknis. Dalam konteks institusi keuangan daerah, karakteristik tersebut sangat penting. Tim pengembang dan pengelola layanan biasanya memerlukan sistem analitik yang tidak hanya akurat, tetapi juga mudah diaudit, mudah dipelihara, dan transparan ketika dipakai untuk menetapkan prioritas perbaikan. Karena itu, baseline TF-IDF dapat dipandang sebagai solusi operasional jangka pendek yang realistis sambil menunggu pengembangan transformer yang lebih matang.

Meskipun demikian, transformasi menuju IndoBERT tetap penting untuk sasaran jangka menengah dan jangka panjang. Ulasan pengguna pada Google Play Store sering memuat kalimat yang ambigu, ironi ringan, campuran kata baku dan nonbaku, serta konteks yang baru dapat dipahami bila model menangkap relasi antarfrasa secara kontekstual. Contoh seperti “sudah update tapi tetap gagal login”, “fiturnya bagus cuma SMS verifikasi tidak masuk”, atau “lumayan membantu, tapi mutasi sering error” memperlihatkan bahwa satu ulasan bisa memuat sinyal positif dan negatif sekaligus. Pada kondisi semacam ini, model berbasis bag-of-words dapat kehilangan nuansa kontras, sedangkan IndoBERT berpotensi membacanya lebih baik karena mempertimbangkan konteks token di sekitarnya [11], [12], [21]. Oleh sebab itu, hasil eksperimen saat ini sebaiknya tidak menghentikan langkah pada model klasik, melainkan menjadi pijakan untuk eksperimen transformer yang lebih sesuai bahasa.

Roadmap eksperimen IndoBERT yang disarankan dapat dilakukan dalam beberapa tahap. Tahap pertama adalah mempertahankan dataset yang sama agar pembandingan tetap adil, lalu mengganti tokenizer dan checkpoint baseline transformer ke IndoBERT atau IndoBERT-lite. Tahap kedua adalah melakukan fine-tuning penuh terhadap seluruh parameter atau setidaknya membuka sebagian layer encoder, karena pendekatan encoder beku seperti sering membatasi kemampuan adaptasi domain. Tahap ketiga adalah menambahkan validasi silang, macro F1-score, precision, recall, dan confusion matrix agar performa kelas netral dapat dinilai lebih akurat [17], [16], [15]. Tahap keempat adalah melakukan analisis error untuk melihat jenis keluhan apa yang paling sering salah diklasifikasikan, misalnya antara netral dan negatif, atau antara pujian parsial dan keluhan teknis.

Selain pengembangan model, penelitian lanjutan juga perlu memperhatikan rekayasa label dan kualitas pra-pemrosesan. Beberapa token netral yang muncul pada koefisien TF-IDF menunjukkan adanya noise, misalnya token yang terpotong, salah eja, atau tidak bermakna jelas. Masalah ini dapat diperkecil melalui normalisasi kamus, koreksi ejaan, penanganan singkatan domain perbankan, dan pengayaan leksikon istilah seperti OTP, PIN, QRIS, mutasi, registrasi, atau verifikasi. Untuk IndoBERT, langkah tersebut harus dirancang hati-hati agar tidak menghapus terlalu banyak informasi konteks. Di sinilah keunggulan model bahasa Indonesia menjadi penting: alih-alih melakukan normalisasi agresif, peneliti dapat memanfaatkan tokenizer subword dan konteks kalimat untuk menangkap variasi penulisan pengguna [11]-[21].

Tabel 4. Kata-kata indikatif teratas dari model TF-IDF

Kelas	Kata/Frasa Indikatif
Negatif	ribet, aplikasi, gangguan, gagal, login, daftar, pulsa, sm, susah, gak
Netral	mutasi rekening, gangguan jaringan, rekening, error, kasiy, vintage, menolong, binggung, mutasi, fiturnya
Positif	mempermudah, terbaik, oke, good, memudahkan, ok, mantap, bagus, mudah, membantu

3.4 Rekomendasi Fine-Tuning IndoBERT

Agar tema penelitian benar-benar berpusat pada IndoBERT, skenario eksperimen lanjutan perlu dirancang secara eksplisit. Pertama, tokenizer harus diganti ke tokenizer IndoBERT agar pemenggalan subword sesuai dengan pola morfologi bahasa Indonesia. Kedua, data sebaiknya dipisahkan menjadi train, validation, dan test yang konsisten, misalnya 70:15:15 atau melalui validasi silang terstratifikasi. Ketiga, karena kelas netral berukuran kecil, peneliti dapat mencoba class weighting, focal loss, atau oversampling yang terkontrol agar model tidak terlalu condong ke kelas positif. Keempat, evaluasi wajib memasukkan confusion matrix dan macro F1-score sehingga kualitas klasifikasi kelas minoritas dapat dinilai lebih adil [17], [16], [15].

Selain itu, fine-tuning IndoBERT perlu disesuaikan dengan karakter ulasan aplikasi yang sangat pendek dan informal. Learning rate rendah, early stopping, serta pemeriksaan error per kelompok keluhan dapat membantu menjaga generalisasi model. Penelitian lanjutan juga dapat membandingkan dua skenario: sentimen tiga kelas umum dan sentimen

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9779
Hal 880-888
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

berbasis aspek, misalnya untuk aspek login, transaksi, notifikasi, mutasi, dan layanan bantuan. Dengan strategi tersebut, hasil analisis tidak hanya berhenti pada penilaian global, tetapi juga dapat langsung diterjemahkan menjadi daftar prioritas pengembangan aplikasi yang lebih rinci dan berbasis data.

Tabel 5. Rancangan eksperimen lanjutan berbasis IndoBERT

Komponen	Rekomendasi
Model pralatih	IndoBERT atau IndoBERT-lite sesuai kapasitas komputasi
Pembagian data	Train-validation-test terstratifikasi atau validasi silang
Strategi pelatihan	Fine-tuning penuh/parsial, early stopping, learning rate rendah
Metrik evaluasi	Accuracy, precision, recall, macro F1-score, confusion matrix
Analisis lanjutan	Error analysis dan sentiment aspect mapping untuk login, transaksi, notifikasi, mutasi

4. KESIMPULAN

Penelitian ini menunjukkan bahwa analisis sentimen terhadap 2.465 ulasan pengguna Mobile Banking Bank Kalbar pada Google Play Store mampu menggambarkan persepsi pengguna terhadap layanan mobile banking, dengan distribusi sentimen yang didominasi oleh ulasan positif, diikuti ulasan negatif, dan sebagian kecil ulasan netral. Hasil eksperimen menunjukkan bahwa TF-IDF + Logistic Regression memperoleh performa terbaik dengan accuracy 0,8581 dan macro-F1 0,6777. RNN BiLSTM memperoleh accuracy 0,8311 dan macro-F1 0,6777, sedangkan IndoBERT memperoleh accuracy 0,8041 dan macro-F1 0,6774. Dengan demikian, TF-IDF + Logistic Regression menjadi model yang paling sesuai untuk klasifikasi sentimen ulasan Mobile Banking Bank Kalbar pada dataset penelitian ini karena menghasilkan akurasi tertinggi dengan pendekatan yang lebih sederhana dan efisien, sementara IndoBERT menunjukkan keunggulan khusus dalam mengenali kelas netral meskipun belum menjadi model dengan performa keseluruhan terbaik.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung eksperimen , serta proses penyusunan artikel ilmiah ini, antara lain dari karyawan Bank Kalbar, Pusat Teknologi Informasi dan Data IAIN Pontianak, serta dosen-dosen pengajar dari Universitas Amikom Yogyakarta.

REFERENCES

[1] Bank Kalbar, “Mobile Banking.” Accessed: Apr. 20, 2026. [Online]. Available: https://bankkalbar.co.id/Mobile_Banking.php
[2] Google Play Store, “Mobile Banking Bank Kalbar.” Accessed: Apr. 01, 2026. [Online]. Available: https://play.google.com/store/apps/details?id=com.xlink.xmobilebankingadkalbar
[3] D. Perea-Khalifi, A. I. Irimia-Diéguez, and P. Palos-Sánchez, “Exploring the determinants of the user experience in P2P payment systems in Spain: a text mining approach,” *Financ. Innov.*, vol. 10, no. 1, p. 2, Jan. 2024, doi: 10.1186/s40854-023-00496-0.
[4] F. Mi Alnaser, S. Rahi, M. Alghizzawi, and A. H. Ngah, “Does artificial intelligence (AI) boost digital banking user satisfaction? Integration of expectation confirmation model and antecedents of artificial intelligence enabled digital banking,” *Heliyon*, vol. 9, no. 8, p. e18930, Aug. 2023, doi: 10.1016/j.heliyon.2023.e18930.
[5] H. Adiningtyas and A. S. Auliani, “Sentiment analysis for mobile banking service quality measurement,” 2024, *Elsevier BV*. doi: 10.1016/j.procs.2024.02.150.
[6] M. Eksa Permana, H. Ramadhan, I. Budi, A. Budi Santoso, and P. Kresna Putra, “Sentiment analysis and topic detection of mobile banking application review,” *2020 5th Int. Conf. Informatics Comput. ICIC 2020*, 2020, doi: 10.1109/ICIC50835.2020.9288616.
[7] K. Andini Safarah, D. I. Inan, R. Juita, and V. L. Arie Sirait, “ANALYSING SERVICE QUALITY USING SENTIMENT ANALYSIS AND TOPIC MODELING: A CASE STUDY OF THE LIVIN MANDIRI APPLICATION,” *J. Inf. Syst. Informatics Eng.*, vol. 8, no. 2, pp. 209–220, 2024, [Online]. Available: https://doi.org/10.35145/joisie.v8i2.4517
[8] M. J. Prasetyo and I. M. A. Agastya, “Sentiment Analysis of Banking Application Reviews on Google Play Store using Support Vector Machine Algorithm,” *SISTEMASI*, vol. 13, no. 6, p. 2386, Nov. 2024, doi: 10.32520/stmsi.v13i6.4536.
[9] M. D. Bimantara and I. Zufria, “Text Mining Sentiment Analysis on Mobile Banking Application Reviews using TF-IDF Method with Natural Language Processing Approach,” *JINAV J. Inf. Vis.*, vol. 5, no. 1, pp. 115–123, 2024, doi: 10.35877/454ri.jinav2772.
[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proceedings of the 2019 Conference of the North*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.
[11] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” in *Proceedings of the 28th International Conference on Computational Linguistics*, Stroudsburg, PA, USA: International Committee on Computational Linguistics, 2020, pp. 757–770. doi: 10.18653/v1/2020.coling-main.66.
[12] S. Cahyawijaya *et al.*, “IndoNLG: Benchmark and Resources for Evaluating Indonesian Natural Language Generation,” in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021, pp. 8875–8898. doi: 10.18653/v1/2021.emnlp-main.699.
[13] S. Cahyawijaya *et al.*, “NusaCrowd: Open Source Initiative for Indonesian NLP Resources,” in *Findings of the Association for Computational Linguistics: ACL 2023*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2023, pp. 13745–

13818. doi: 10.18653/v1/2023.findings-acl.868.

[14] Y. A. Singgalen, "Performance Analysis of IndoBERT for Sentiment Classification in Indonesian Hotel Review Data," *J. Inf. Syst. Res.*, vol. 6, no. 2, pp. 976–986, Jan. 2025, doi: 10.47065/josh.v6i2.6505.

[15] S. Apriliani, A. Erfina, and C. Warman, "Fine-Tuned IndoBERT for Aspect-Based Sentiment Analysis of Indonesian Five-Star Hotel Reviews," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 14, no. 4, pp. 437–445, Oct. 2025, doi: 10.32736/sisfokom.v14i4.2491.

[16] K. C. Pradhisa and R. Fajriyah, "Analisis Sentimen Ulasan Pengguna E-commerce di Google Play Store Menggunakan Metode IndoBERT," *Build. Informatics, Technol. Sci.*, vol. 6, no. 1, Jun. 2024, doi: 10.47065/bits.v6i1.5247.

[17] A. A. P. Simarmata and T. B. Sasongko, "Sentiment Analysis on BRImo Application Reviews Using IndoBERT," *J. Appl. Informatics Comput.*, vol. 9, no. 3, pp. 851–862, 2025, doi: 10.30871/jaic.v9i3.8162.

[18] Natan Kharisma A, Dewi Lestari, and Gatot T Pranoto, "Sentiment Analysis Review Threads Google Play Store with RoBERTa Modelx`," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 14, no. 4, pp. 272–280, 2025, doi: 10.22146/jnteti.v14i4.22038.

[19] M. Ullah, J. Li, and B. Wadood, "Analysis of Urban Expansion and its Impacts on Land Surface Temperature and Vegetation Using RS and GIS, A Case Study in Xi'an City, China," 2020, *Springer Science and Business Media LLC*. doi: 10.1007/s41748-020-00166-6.

[20] W. Ullah, Z. Zhang, and K. Stefanidis, "Sentiment Analysis of Mobile Apps Using BERT," H. Fujita, Y. Wang, Y. Xiao, and A. Moonis, Eds., Cham: Springer Nature Switzerland, 2023, pp. 66–78. doi: 10.1007/978-3-031-36822-6_6.

[21] F. Koto, J. H. Lau, and T. Baldwin, "IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021, pp. 10660–10668. doi: 10.18653/v1/2021.emnlp-main.833.

[22] R. Savitri, F. Rizki, and A. Sobri, "Implementation of BERT in Sentiment Analysis of National Digital Samsat (SIGNAL) User Reviews Based on Machine Learning," *MATICS J. Ilmu Komput. dan Teknol. Inf. (Journal Comput. Sci. Inf. Technol.*, vol. 17, no. 2, pp. 67–75, 2025, doi: 10.18860/mat.v17i2.32059.