ANALISIS DISKREPANSI ANTARA RATING DAN SENTIMEN TEKS PADA ULASAN APLIKASI SEABANK MENGGUNAKAN MODEL DEEP LEARNING INDOBERT

Jihan Amirah Munif, Afu Ichsan Pradana, Dwi Hartanti
Teknik Informatika, Universitas Duta Bangsa Surakarta
Jalan Bhayangkara No. 55, Surakarta 57154, Indonesia
amunifjihan03@gmail.com

ABSTRAK

Kemajuan layanan perbankan berbasis digital turut memicu lonjakan ulasan pengguna di Google Play Store, baik berupa penilaian numerik maupun teks ulasan. Akan tetapi, keduanya kerap menunjukkan ketidakkonsistenan, di mana rating yang diberikan tidak selalu merepresentasikan sentimen sesungguhnya dalam teks ulasan. Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan pengguna aplikasi SeaBank melalui model *deep learning* IndoBERT sekaligus mendeteksi diskrepansi antara rating numerik dan prediksi sentimen teks. Sebanyak 14.466 ulasan diperoleh melalui *web scraping* Google Play Store setelah melalui proses seleksi dengan mengeliminasi data kosong dan ulasan berrating 3. Pelatihan model dilakukan pada 1.500 sampel yang dipilih secara *stratified* dengan rasio 80:10:10 untuk data latih, validasi, dan uji, serta penanganan ketidakseimbangan kelas menggunakan *class weight*. Hasil evaluasi menunjukkan akurasi model sebesar 98,00% dengan *F1-score* 0,90 untuk kelas negatif dan 0,99 untuk kelas positif. Dari total 14.466 data, terdeteksi 823 ulasan (5,69%) yang mengalami diskrepansi, meliputi 655 ulasan Tipe A (rating positif, prediksi negatif) dan 168 ulasan Tipe B (rating negatif, prediksi positif). Hasil ini membuktikan bahwa penilaian numerik tidak sepenuhnya merepresentasikan sentimen yang terkandung dalam teks ulasan pengguna.

Kata kunci : *Analisis Sentimen, Diskrepansi, Rating, IndoBERT, Deep Learning, SeaBank.*

1. PENDAHULUAN

Transformasi digital di Indonesia mendorong perubahan signifikan dalam industri perbankan, terutama dengan hadirnya layanan perbankan digital berbasis aplikasi seperti SeaBank. Aplikasi ini diminati luas karena menawarkan keunggulan berupa bunga harian, bebas biaya administrasi, serta kemudahan transaksi tanpa batasan waktu dan tempat. Tingginya minat pengguna tercermin dari jumlah ulasan yang telah melampaui 3 juta di Google Play Store, menjadikan platform ini sumber data yang kaya untuk keperluan analisis layanan. Umpan balik pengguna di Google Play Store hadir dalam dua bentuk, yaitu rating numerik berskala 1–5 dan teks ulasan. Namun, penilaian bintang yang diberikan pengguna seringkali tidak secara akurat mencerminkan isi ulasan mereka [1].

Fenomena ini dikenal sebagai diskrepansi rating dan sentimen teks, yang berpotensi menimbulkan bias dalam evaluasi kualitas layanan apabila analisis hanya mengandalkan rating numerik semata. Beberapa penelitian telah mendokumentasikan fenomena diskrepansi ini. Sekitar 24,7% rating pengguna diidentifikasi bersifat bias akibat ketidaksinkronan antara penilaian kualitatif dengan nilai bintang yang diberikan [2].

Pada platform digital Indonesia, ketidaksesuaian antara rating dan sentimen teks juga ditemukan, sehingga analisis sentimen diperlukan untuk memperoleh interpretasi opini yang lebih akurat [3].

Analisis sentimen pada ulasan aplikasi perbankan digital seperti BRImo dengan *Naive Bayes* menunjukkan potensi pemanfaatan ulasan Google Play Store sebagai bahan evaluasi layanan, meskipun akurasi yang dicapai masih terbatas pada 84,52% [4].

Penelitian terdahulu yang menganalisis sentimen ulasan aplikasi SeaBank hanya berfokus pada klasifikasi sentimen dan belum menyentuh persoalan diskrepansi antara rating numerik dan isi teks ulasan [5] [6].

Di samping itu, pelabelan data pada penelitian tersebut masih menggunakan pendekatan otomatis berbasis rating yang berpotensi menghasilkan label yang tidak akurat. Kondisi ini menunjukkan adanya kesenjangan penelitian yang perlu ditangani agar evaluasi sentimen pengguna aplikasi SeaBank dapat dilakukan secara lebih komprehensif dan andal.

Analisis sentimen merupakan teknik dalam NLP yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini dalam teks ke dalam kategori positif dan negatif. BERT, model *deep learning* berbasis transformer, terbukti mampu meningkatkan performa berbagai tugas NLP secara signifikan [7].

IndoBERT dikembangkan dari arsitektur BERT dengan pelatihan khusus pada korpus bahasa Indonesia sehingga lebih mampu menangkap nuansa struktur dan konteks bahasa Indonesia. Penelitian terkini menunjukkan IndoBERT mampu mencapai akurasi 0,96 dan *F1-score makro* 0,95 dalam analisis sentimen berbahasa Indonesia, membuktikan keunggulannya dibandingkan metode konvensional [8].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengklasifikasikan sentimen ulasan pengguna aplikasi SeaBank menggunakan model IndoBERT serta mendeteksi dan menganalisis diskrepansi antara rating numerik dan hasil prediksi sentimen teks ulasan. Untuk meningkatkan keandalan

analisis, pelabelan data dilakukan secara manual guna menghindari bias yang muncul dari pendekatan pelabelan otomatis berbasis rating. Dengan demikian, penelitian ini diharapkan mampu memberikan gambaran yang lebih akurat mengenai opini pengguna terhadap layanan perbankan digital SeaBank.

2. TINJAUAN PUSTAKA

2.1. Natural Language Processing (NLP)

Pemrosesan Bahasa Alami (NLP) adalah cabang dari pemrosesan bahasa alami yang berfokus pada kemampuan komputer untuk memahami, memproses, dan menganalisis bahasa manusia secara otomatis. Perkembangan NLP sangat dipengaruhi oleh model berbasis Transformer, khususnya BERT (*Bidirectional Encoder Representations from Transformers*), yang mampu memahami konteks bahasa dalam dua arah. Menurut penelitian lain, BERT telah memajukan NLP dengan meningkatkan efisiensi berbagai tugas, seperti tokenisasi, penamaan entitas, menjawab pertanyaan, dan analisis sentimen [7].

2.2. Analisis Sentimen

Analisis sentimen merupakan teknik dalam NLP yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini dalam teks ke dalam kategori positif, negatif, dan netral. Dalam konteks ulasan aplikasi di platform digital, analisis sentimen berperan penting dalam memahami persepsi pengguna terhadap kualitas layanan. Penelitian pada ulasan aplikasi BRImo membuktikan bahwa ulasan pengguna di Google Play Store dapat dimanfaatkan sebagai sumber evaluasi layanan melalui analisis sentimen dengan *Naive Bayes* yang mencapai akurasi 84,52% [4].

Performa model juga dipengaruhi oleh karakteristik dataset, khususnya pada kondisi data tidak seimbang, sehingga pemilihan model dan strategi penanganan data menjadi faktor penting dalam analisis sentimen [9].

2.3. Deep Learning dan BERT

Sebuah cabang dari pembelajaran mesin yang disebut *deep learning* memanfaatkan jaringan saraf dalam untuk secara otomatis mengidentifikasi pola dari volume data yang sangat besar, sehingga meningkatkan kemampuannya dalam mengolah data tidak terstruktur seperti teks [10].

BERT merupakan model bahasa berbasis *deep learning* yang dirancang untuk memahami konteks kata secara dua arah dalam suatu kalimat menggunakan arsitektur Transformer. Keunggulan utama BERT adalah kemampuannya untuk dilatih terlebih dahulu pada data berukuran besar kemudian disesuaikan untuk kebutuhan tertentu, sehingga mampu menghasilkan representasi teks yang kontekstual dan meningkatkan akurasi klasifikasi dibandingkan metode berbasis fitur tradisional [7].

2.4. IndoBERT

Sebagai pengembangan dari BERT, IndoBERT lebih mampu menyesuaikan diri dengan kosakata, struktur, dan konteks linguistik bahasa Indonesia karena dilatih menggunakan korpus bahasa Indonesia. IndoBERT menunjukkan kinerja yang lebih baik daripada *Naive Bayes* dan SVM, dengan tingkat akurasi sebesar 0,96 dan skor *F1 makro* sebesar 0,95 [8].

IndoBERT menunjukkan kinerja yang baik dalam hal presisi, *recall*, dan *F1-score*, dengan tingkat akurasi mencapai 89% dalam klasifikasi sentimen ulasan Google Play Store berbahasa Indonesia [11], sehingga IndoBERT dinilai sebagai model yang efektif dan relevan untuk analisis sentimen teks berbahasa Indonesia.

2.5. Diskrepansi Rating dan Sentimen Teks

Diskrepansi rating dan sentimen teks merupakan fenomena ketidaksesuaian antara rating numerik yang diberikan pengguna dengan sentimen yang terkandung dalam teks ulasannya. Penilaian bintang yang diberikan pengguna seringkali tidak secara akurat mencerminkan isi ulasan mereka, sehingga terdapat potensi ketidaksesuaian antara kedua bentuk penilaian tersebut [1].

Sekitar 24,7% rating pengguna bersifat bias dan model berbasis transformer mampu mendeteksi ketidaksesuaian tersebut dengan lebih baik dibanding metode tradisional [2].

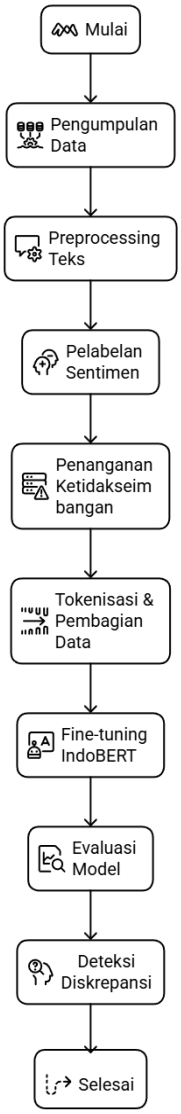
Adanya ketidaksesuaian antara penilaian bintang dan sentimen teks pada platform digital Indonesia, sehingga diperlukan analisis sentimen untuk memperoleh interpretasi opini pengguna yang lebih akurat [3].

2.6. Penelitian Terdahulu

Beberapa penelitian terdahulu telah menganalisis sentimen ulasan aplikasi SeaBank di Google Play Store. Model yang digunakan *Naive Bayes* dengan akurasi 88% dan ada yang menggunakan SVM dengan akurasi 93,99%, namun keduanya masih menggunakan metode *machine learning* tradisional dan hanya berfokus pada klasifikasi sentimen tanpa mempertimbangkan ketidaksesuaian antara rating dan isi teks ulasan. Selain itu, pelabelan data pada kedua penelitian tersebut masih menggunakan pendekatan otomatis berbasis rating yang berpotensi menghasilkan label yang tidak akurat. Berdasarkan tinjauan tersebut, belum ada penelitian yang secara khusus menggabungkan IndoBERT dengan analisis diskrepansi antara rating dan sentimen teks pada ulasan berbahasa Indonesia dari aplikasi SeaBank, sehingga penelitian ini hadir untuk mengisi kekosongan tersebut [5] [6].

3. METODE PENELITIAN

Penelitian ini dilaksanakan melalui delapan tahapan sistematis sebagaimana ditunjukkan pada Gambar 1, meliputi pengumpulan data, *preprocessing* teks, pelabelan sentimen, penanganan ketidakseimbangan data, tokenisasi dan pembagian data, *fine-tuning* IndoBERT, evaluasi model, dan deteksi diskrepansi.



Gambar 1. Tahapan penelitian

Berikut ini adalah penjelasan mengenai tahapan-tahapan penelitian:

3.1. Pengumpulan Data

Dengan menggunakan modul *google-play-scraper* berbasis Python, data primer dikumpulkan melalui *web scraping*. Parameternya diatur ke bahasa Indonesia, Indonesia sebagai negara, dan ulasan terbaru untuk periode Maret-April 2026. Teknik ini memungkinkan pengumpulan ulasan secara otomatis dalam jumlah besar dari Google Play Store. Tingginya proporsi rating positif yang diperoleh konsisten dengan temuan penelitian sebelumnya yang menyatakan bahwa ulasan pada Google Play Store dapat dikumpulkan melalui *web scraping* dan dimanfaatkan sebagai dataset untuk analisis lanjutan dalam bidang NLP [12].

3.2. Preprocessing Teks

Tahapan *preprocessing* meliputi *case folding*, penghapusan karakter non-alfabet, dan penghapusan spasi berlebih [13].

Angka dan karakter khusus dihapus karena dianggap tidak berkontribusi signifikan terhadap klasifikasi sentimen. Ulasan kosong dan ulasan berrating 3 kemudian dihapus karena bersifat netral dan berpotensi menimbulkan ambiguitas dalam proses klasifikasi sentimen [14].

3.3. Pelabelan Sentimen

Pelabelan sentimen dilakukan dalam dua tahap. Pertama, pelabelan otomatis berbasis rating sebagai analisis awal, di mana rating 1-2 dikategorikan sebagai negatif dan rating 4-5 dikategorikan sebagai positif. Kedua, pelabelan manual dilakukan terhadap data sampel yang dipilih secara *stratified* untuk memperoleh label yang lebih akurat dan mengurangi potensi bias yang dihasilkan oleh pendekatan otomatis berbasis rating.

3.4. Penanganan Ketidakseimbangan Data

Model tersebut cenderung lebih mengutamakan prediksi kelas mayoritas sambil mengabaikan kelas minoritas jika distribusi data antara kelas positif dan negatif sangat timpang. Untuk mengatasi hal ini, digunakan strategi pembobotan kelas, yang memberikan bobot lebih besar kepada kelas minoritas selama proses pelatihan [9].

Nilai bobot dihitung menggunakan persamaan berikut:

$$\omega_j = \frac{n_{samples}}{n_{classes} \times n_j} \quad (1)$$

Keterangan:

a. ω_j : Bobot untuk kelas *j*
b. $n_{samples}$: Jumlah total sampel dalam seluruh dataset
c. $n_{classes}$: Jumlah total kelas unik (2 untuk klasifikasi biner)
d. n_j : Jumlah sampel dalam kelas *j*

Penerapan teknik ini bertujuan agar model tidak bias terhadap kelas mayoritas dan tetap mampu mengenali kelas negatif dengan baik meskipun jumlahnya jauh lebih sedikit.

3.5. Tokenisasi dan Pembagian Data

Sebelum dimasukkan ke model, teks ulasan dikonversi menjadi token menggunakan tokenizer IndoBERT dengan panjang maksimum 128 token, *padding max_length*, dan truncation [6].

Dataset kemudian dibagi secara *stratified* menjadi data latih yang digunakan 80% serta untuk validasi dan data uji sama yaitu 10% dengan distribusi label yang proporsional pada setiap subset [11].

3.6. Fine-tuning IndoBERT

Proses *fine-tuning* IndoBERT dilakukan selama 5 *epoch* menggunakan data latih dengan *learning rate 2e-5*, *batch size* 16, dan *optimizer* AdamW yang dikombinasikan dengan *linear scheduler* [8] [11].

Dibandingkan dengan model bahasa umum, IndoBERT adalah sebuah model yang didasarkan pada arsitektur BERT yang lebih mampu menangkap reprsentasi sematik dan kontekstual teks berbahasa Indonesia karena model dilatih secara secara khusu menggunakan korpus bahasa Indonesia.

3.7. Evaluasi Model

Model pembelajaran mendalam berbasis transformer seperti IndoBERT mampu memahami makna kata dalam dua arah dengan menggunakan mekanisme perhatian untuk memperoleh representasi kontekstual dari teks [10].

Model tersebut dievaluasi menggunakan data uji dengan menerapkan empat metrik utama. Berikut ini adalah penjelasan mengenai metrik-metrik tersebut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Keterangan:

a. TP (*True Positive*) : jumlah titik data positif yang secara akurat diprediksi sebagai positif
b. TN (*True Negative*) : jumlah titik data negatif yang secara akurat diprediksi sebagai negatif
c. FP (*False Positive*) : jumlah titik data negatif yang secara keliru diprediksi sebagai positif
d. FN (*False Negative*) : jumlah titik data positif yang secara keliru diprediksi sebagai negatif

3.8. Deteksi Diskrepansi

Deteksi diskrepansi dilakukan dengan mengaplikasikan model IndoBERT pada keseluruhan data yang telah di*preprocessing* untuk membandingkan hasil prediksi sentimen dengan label rating asli pengguna. Ulasan dikategorikan mengalami diskrepansi Tipe A apabila pengguna memberikan rating positif (4-5) namun model memprediksi sentimen negatif, dan dikategorikan diskrepansi Tipe B apabila pengguna memberikan rating negatif (1-2) namun model memprediksi sentimen positif.

4. HASIL DAN PEMBAHASAN

Dari tahapan penelitian diatas, menghasilkan penelitian seperti dibawah ini:

4.1. Hasil Pengumpulan Data

Sebanyak 15.000 tanggapan pengguna terhadap SeaBank berhasil dihimpun dari Google Play Store melalui *web scraping*. Distribusi rating menunjukkan dominasi pada rating 5 dengan 12.528 ulasan, diikuti rating 4 sebanyak 1.120 ulasan, rating 1 sebanyak 644 ulasan, rating 3 sebanyak 534 ulasan, dan rating 2 sebanyak 174 ulasan. Tingginya proporsi rating positif mengindikasikan bahwa sebagian besar pengguna merasa puas dengan layanan aplikasi SeaBank.

4.2. Hasil Preprocessing Teks

Pemrosesan teks dilakukan secara bertahap untuk menghasilkan data yang bersih dan siap digunakan dalam pelatihan model. Penghapusan angka dan karakter khusus menghasilkan 14.828 ulasan bersih, kemudian dilanjutkan dengan penghapusan 326 ulasan berrating 3 yang bersifat netral, sehingga jumlah akhir data ulasan bersih adalah 14.466 ulasan. Contoh kalimat perbandingan dapat dilihat pada Tabel 1.

Tabel 1. Contoh Hasil Preprocessing Teks

No	Sebelum Preprocessing	Setelah Preprocessing
1.	sangat puas,mudah untuk digunakan, membantu untuk semua transaksi.trimakasih seabank □□	sangat puas mudah untuk digunakan membantu untuk semua transaksi trimakasih seabank
2.	mantab 👆	mantab
3.	gratis biaya transfer trz ya moga2..?	gratis biaya transfer trz ya moga
4.	Sangat membantu	sangat membantu
5.	sea bank mudah di gunakan, aman dan nyaman pengiriman cepat, sat set masuk.	sea bank mudah di gunakan aman dan nyaman pengiriman cepat sat set masuk

4.3. Hasil Pelabelan Sentimen

Pelabelan otomatis berbasis rating pada 14.466 ulasan menghasilkan temuan bahwa 326 ulasan (2,25%) mengalami diskrepansi pada tahap awal ini. Selanjutnya, pelabelan manual dilakukan terhadap 1.500 data sampel yang dipilih secara *stratified* dan menghasilkan 1.366 ulasan (91,1%) berlabel positif dan 134 ulasan (8,9%) berlabel negatif. Pelabelan manual ini digunakan sebagai data pelatihan model untuk memperoleh label yang lebih akurat dibandingkan pendekatan otomatis berbasis rating.

4.4. Hasil Penanganan Ketidakseimbangan Data

Distribusi data menunjukkan ketimpangan yang cukup signifikan antara kelas negatif (8,9%) dan kelas positif (91,1%). Perhitungan berdasarkan Persamaan (1) menghasilkan bobot sebesar 0,55 untuk kelas positif dan 5,60 untuk kelas negatif. Hal ini memastikan bahwa model tidak condong ke kelas mayoritas dengan memberikan bobot 10,2 kali lebih besar kepada kelas negatif selama proses pelatihan.

4.5. Hasil Tokenisasi dan Pembagian Data

Teks ulasan dikonversi menjadi token menggunakan *tokenizer* IndoBERT dengan panjang maksimum 128 token. Dataset 1.500 data kemudian dibagi secara *stratified* dengan distribusi sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Distribusi Pembagian Dataset

Subset	Jumlah Data	Positif	Negatif
Latih (80%)	1.200	1.093	107
Validasi (10%)	150	137	13
Uji (10%)	150	136	14
Total	1.500	1.366	134

4.6. Hasil Pelatihan Model

Dari 0,23 pada *epoch* pertama menjadi 0,02 pada *epoch* kelima, nilai *loss* pelatihan terus menurun, yang menunjukkan bahwa model terus belajar dan meningkatkan bobot-bobotnya selama fase pelatihan. Akurasi validasi menunjukkan tren kenaikan dari 93,33% menjadi 96,00% pada *epoch* ketiga, kemudian sedikit menurun menjadi 95,33% pada *epoch* kelima, meskipun terdapat tanda-tanda *overfitting* yang moderat, sebagaimana ditunjukkan oleh peningkatan kerugian validasi dari 0,17 pada *epoch* pertama menjadi 0,26 pada *epoch* kelima. Menerapkan pembobotan kelas pada data yang tidak seimbang secara alami mengakibatkan kenaikan kerugian validasi, yang memiliki sedikit pengaruh terhadap kapasitas generalisasi model. Tabel 3 menampilkan hasil pelatihan model.

Tabel 3. Hasil pelatihan model per *epoch*

Epoch	*Train loss*	*Train Acc*	*Val loss*	*Val Acc*
1	0,23	92,25%	0,17	93,33%
2	0,07	97,25%	0,20	93,33%
3	0,03	98,83%	0,28	96,00%
4	0,02	99,50%	0,29	94,67%
5	0,02	99,58%	0,26	95,33%

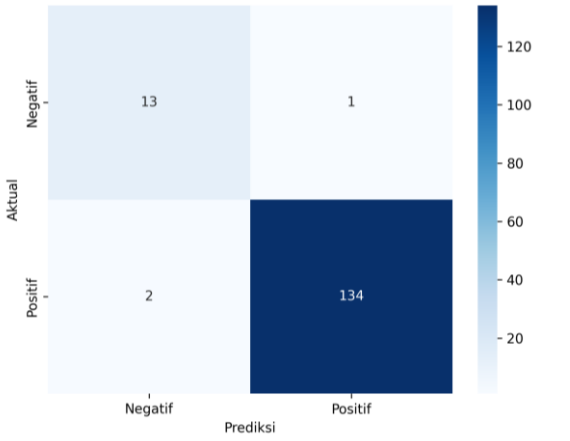
4.7. Hasil Evaluasi Model

Akurasi model yakni 98,00% dengan nilai *macro average* untuk *precision*, *recall*, dan *F1-score* secara berturut-turut sebesar 0,93, 0,96, dan 0,94. Pada kelas positif, model mencapai performa sangat tinggi dengan *precision*, *recall*, dan *F1-score* sebesar 0,99. Sementara itu, kelas negatif, memperoleh *precision* sebesar 0,87, *recall* sebesar 0,93, dan *F1-score* sebesar 0,90. Performa yang lebih rendah pada kelas negatif wajar mengingat jumlah data negatif hanya 14 dari total 150 data uji, namun penerapan *class weighting* terbukti membantu model dalam mengenali kelas minoritas dengan cukup baik. Hasil evaluasi model secara lengkap ditunjukkan pada Tabel 4.

Tabel 4. Classification report model IndoBERT

Kelas	*Precision*	*Recall*	*F1-Score*	*Support*
Negatif	0,87	0,93	0,90	14
Positif	0,99	0,99	0,99	136
Accuracy	-	-	0,98	150
Macro Average	0,93	0,96	0,94	150
Weighted Average	0,98	0,98	0,98	150

Algoritma tersebut berhasil mengidentifikasi 134 ulasan positif (*True Positive*) dan 13 ulasan negatif (*True Negative*) dari 150 titik data uji. Satu ulasan negatif keliru diklasifikasikan sebagai positif (*False Positive*), dan dua ulasan positif keliru diklasifikasikan sebagai negatif (*False Negative*). Skor True Positive dan True Negative yang tinggi menunjukkan bahwa model ini mampu membedakan dengan baik antara sentimen positif dan negatif meskipun data latihannya tidak seimbang. Hasil ini menunjukkan bahwa IndoBERT tetap menunjukkan kinerja yang memuaskan meskipun hanya menggunakan 1.200 titik data latih. Gambar 2 di bawah ini menampilkan *confusion matrix*.



Gambar 2. Confusion matrix

4.8. Hasil Deteksi Diskrepansi

Dari 14.466 ulasan yang dianalisis, sebanyak 823 ulasan (5,69%) terdeteksi mengalami diskrepansi antara rating dan sentimen teks. Diskrepansi Tipe A yang terdiri dari 655 ulasan (4,53%) merupakan kasus di mana pengguna memberikan rating positif (4-5) namun teks ulasannya mengandung sentimen negatif. Sementara diskrepansi Tipe B yang terdiri dari 168 ulasan (1,16%) merupakan kasus sebaliknya yaitu rating negatif (1-2) namun teks ulasan bersentimen positif. Hasil deteksi diskrepansi dapat dipaparkan pada Tabel 5 berikut.

Tabel 5. Hasil deteksi diskrepansi

Status	Jumlah	Persentase
Sesuai	13.643	94,31%
Diskrepansi Tipe A	655	4,53%
Diskrepansi Tipe B	168	1,16%
Total diskrepansi	823	5,69%
Total data	14.466	100%

Kasus diskrepansi Tipe A umumnya terjadi pada ulasan yang mengandung pertanyaan atau keluhan terkait fitur aplikasi seperti pinjaman, transfer, dan administrasi, meskipun pengguna memberikan rating tinggi. Hal ini mengindikasikan bahwa pengguna mungkin puas secara keseluruhan dengan aplikasi namun memiliki keluhan spesifik terhadap fitur tertentu. Sementara diskrepansi Tipe B umumnya terjadi pada ulasan yang mengandung kata-kata positif

seperti "sangat membantu", "saya suka", dan "baguss" meskipun pengguna memberikan rating rendah.

Hasil ini mendukung temuan penelitian terdahulu yang mengungkapkan bahwa sekitar 24,7% rating yang diberikan pengguna tidak mencerminkan isi ulasan teks secara konsisten [2].

Meskipun persentase diskrepansi dalam penelitian ini lebih kecil (5,69%), hal ini dapat disebabkan oleh perbedaan karakteristik dataset dan domain aplikasi yang digunakan. Fenomena diskrepansi yang ditemukan mengonfirmasi bahwa rating numerik tidak selalu dapat dijadikan satu-satunya indikator kepuasan pengguna, sehingga analisis sentimen berbasis teks diperlukan untuk memperoleh gambaran yang lebih komprehensif dan akurat mengenai persepsi pengguna terhadap layanan aplikasi SeaBank. Temuan ini relevan mengingat penelitian terkini menemukan bahwa tingginya sentimen positif pada ulasan digital banking mencerminkan efisiensi dan kemudahan penggunaan aplikasi, namun analisis yang hanya mengandalkan rating numerik berpotensi mengabaikan keluhan spesifik yang tersembunyi dalam teks ulasan [15]. Contoh kasus diskrepansi disajikan pada Tabel 6.

Tabel 6. Contoh kasus diskrepansi

Teks Ulasan	Rating	Prediksi IndoBERT	Tipe
Kenapa belum bisa mengajukan pinjaman	5	Negatif	A
Pajak agak kurangin	5	Negatif	A
Cuma bukti transfer tdk bs d cetak	5	Negatif	A
Sangat membantu	2	Positif	B
Saya suka	1	Positif	B
Bagus	1	Positif	B

5. KESIMPULAN DAN SARAN

Penelitian ini menggunakan model pembelajaran mendalam IndoBERT untuk mengklasifikasikan sentimen ulasan pengguna terhadap aplikasi SeaBank dan berhasil menghasilkan hasil yang sangat baik, di mana model yang dilatih menggunakan 1.200 data berlabel manual mencapai *accuracy* 98,00% dengan *precision*, *recall*, dan *F1-score* kelas negatif sebesar 0,87, 0,93, dan 0,90 serta kelas positif masing-masing sebesar 0,99, membuktikan bahwa meskipun dilatih menggunakan jumlah data yang relatif sedikit, IndoBERT lebih berhasil daripada metode tradisional dalam memahami konteks bahasa Indonesia. Hasil deteksi diskrepansi pada keseluruhan 14.466 data menunjukkan bahwa sebanyak 823 ulasan (5,69%) mengalami ketidaksesuaian antara rating numerik dan sentimen teks, terdiri dari 655 ulasan Tipe A (4,53%) yaitu rating positif namun prediksi negatif, dan 168 ulasan Tipe B (1,16%) yaitu rating negatif namun prediksi positif, yang mengonfirmasi bahwa rating numerik tidak selalu mencerminkan sentimen aktual yang terkandung dalam teks ulasan pengguna sehingga analisis sentimen berbasis teks menjadi pendekatan yang lebih tepat untuk memahami persepsi pengguna secara menyeluruh.

Studi berikutnya direkomendasikan untuk menambah jumlah data berlabel manual dengan cakupan yang lebih luas, mengeksplorasi model berbasis transformer lainnya seperti RoBERTa atau XLM-R untuk perbandingan performa, serta memperluas analisis diskrepansi pada aplikasi digital banking lainnya di Indonesia guna memperoleh gambaran yang lebih komprehensif mengenai perilaku pengguna dalam memberikan ulasan.

DAFTAR PUSTAKA

[1] N. Wijaya dan E. S. Panjaitan, "Analisis Sentimen Ulasan Aplikasi Instagram di Google Play Store: Pendekatan Multinomial Naive Bayes dan Berbasis Leksikon," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 2, hlm. 921–929, Sep 2024, doi: 10.47065/bits.v6i2.5615.

[2] T. Aljrees *dkk.*, "Contradiction in text review and apps rating: prediction using textual features and transfer learning," *PeerJ Computer Science*, vol. 10, hlm. e1722, Jan 2024, doi: 10.7717/peerj-cs.1722.

[3] A. A. Z. Fikri dan H. Ridho, "Identification of Inconsistent Reviews and Ratings on Apps Using Sentiment Analysis: Case Study on Indonesian Digital Media Platform," *Metris: Jurnal Sains dan Teknologi*, vol. 26, no. 01, hlm. 39–50, Jul 2025, doi: 10.25170/metris.v26i01.6779.

[4] M. K. K. Insan, U. Hayati, dan O. Nurdiawan, "ANALISIS SENTIMEN APLIKASI BRIMO PADA ULASAN PENGGUNA DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAIVE BAYES," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, hlm. 478–483, Jan 2023, doi: 10.36040/jati.v7i1.6373.

[5] N. Z. Putri, M. Martanto, A. R. Dikananda, dan A. Rifa'i, "Analisis Sentimen Aplikasi SeaBank dengan Algoritma Naive Bayes untuk Optimalisasi Pelayanan," *Jurnal Informatika Terpadu*, vol. 11, no. 1, hlm. 55–62, Apr 2025, doi: 10.54914/jit.v11i1.1721.

[6] Y. Christian, T. Wibowo, dan M. Lyawati, "Sentiment Analysis by Using Naïve Bayes Classification and Support Vector Machine, Study Case Sea Bank," *Sinkron : jurnal dan penelitian teknik informatika*, vol. 8, no. 1, hlm. 258–275, Jan 2024, doi: 10.33395/sinkron.v9i1.13141.

[7] N. M. Gardazi, A. Daud, M. K. Malik, A. Bukhari, T. Alsahfi, dan B. Alshemaimri, "BERT applications in natural language processing: a review," *Artif Intell Rev*, vol. 58, no. 6, hlm. 166, Mar 2025, doi: 10.1007/s10462-025-11162-5.

[8] A. A. Qolbu, N. Fitriyati, dan N. Inayah, "Performa Naïve Bayes, SVM, dan IndoBERT pada Analisis Sentimen Twitter IndiHome dengan Strategi Penanganan Data Tidak

Seimbang," *Jurnal Fourier*, vol. 14, no. 1, hlm. 29–44, Okt 2025, doi: 10.14421/fourier.2025.141.29-44.

[9] M. Fazri dan A. Voutama, "ANALISIS SENTIMEN PUBLIK TERHADAP DANANTARA DI MEDIA SOSIAL X MENGGUNAKAN NLP DAN PEMBELAJARAN MESIN," *JOISIE (Journal Of Information Systems And Informatics Engineering)*, vol. 9, no. 1, hlm. 197–206, Jul 2025, doi: 10.35145/joisie.v9i1.4924.

[10] M. M. Taye, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions," *Computers*, vol. 12, no. 5, hlm. 91, Apr 2023, doi: 10.3390/computers12050091.

[11] I. Mursidah, R. Sanjaya, B. Yulianto, D. Sweetania, dan P. Sularsih, "Klasifikasi Sentimen Google Play Store Aplikasi ChatGPT Berbahasa Indonesia Berbasis IndoBERT," *Jurnal Minfo Polgan*, vol. 14, no. 2, hlm. 3349–3359, Des 2025, doi: 10.33395/jmp.v14i2.15751.

[12] A. Yasin, R. Fatima, A. N. Ghazi, dan Z. Wei, "Python data odyssey: Mining user feedback from google play store," *Data in Brief*, vol. 54, hlm. 110499, Jun 2024, doi: 10.1016/j.dib.2024.110499.

[13] M. A. Palomino dan F. Aider, "Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis," *Applied Sciences*, vol. 12, no. 17, hlm. 8765, Agu 2022, doi: 10.3390/app12178765.

[14] S. F. Kadir dan A. Fairuzabadi, "Analisis Sentimen Ulasan Shopee di Google Play dengan TF-IDF dan Logistic Regression," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 2, hlm. 7940–57945, Jul 2025, doi: 10.31004/riggs.v4i2.2850.

[15] D. Tomasaputra dan M. Siallagan, "Enhancing Digital Banking Experience Through Text Mining of e-Service Reviews in Indonesia," *Ekonomis: Journal of Economics and Business*, vol. 9, no. 2, hlm. 1080–1096, Sep 2025, doi: 10.33087/ekonomis.v9i2.2653.