Jurnal Bumigora Information Technology (BITe)
Vol. 7, No. 2, Desember 2025, Hal. 131∼140
E-ISSN: 2685-4066, DOI: 10.30812/bite.v7i2.5943

Analisis Sentimen Ulasan Mobile Banking Syariah Berbasis NLP untuk Evaluasi dan Peningkatan Kualitas Layanan Digital

NLP-Based Sentiment Analysis of Islamic Mobile Banking Reviews for Digital Service Quality Enhancement

Moh. Yushi Assani[1*], **Ari Rudiyan**[1], **Ummu Radiyah**[2]
[1]Universitas Islam Al-Azhar, Mataram, Indonesia
[2]Universitas Nusa Mandiri, Jakarta Timur, Indonesia

Informasi Artikel:
Diterima: 25 November 2025, Direvisi: 29 Desember 2025, Disetujui: 31 Desember 2025

Abstrak-
Latar Belakang: Perkembangan layanan mobile banking syariah di Indonesia menuntut peningkatan kualitas manajemen layanan digital yang berfokus pada pengalaman pengguna. Ulasan pengguna pada platform aplikasi menjadi sumber data penting untuk memahami persepsi, kepuasan, serta permasalahan yang dihadapi pengguna. Pemanfaatan Natural Language Processing (NLP) memungkinkan analisis sentimen dilakukan secara sistematis dan otomatis guna mendukung evaluasi kualitas layanan perbankan syariah.
Tujuan: Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi mobile banking syariah serta membandingkan kinerja model analisis sentimen Naïve Bayes sebagai model klasik dan IndoBERT sebagai model berbasis transformer.
Metode: Data diperoleh melalui web scraping terhadap 1.052 ulasan pengguna aplikasi Bank NTB Syariah Mobile Banking dari Google Play Store. Tahapan prapemrosesan meliputi case folding, cleaning, stopword removal, stemming, dan tokenization. Pelabelan sentimen dilakukan berdasarkan skor rating pengguna dengan tiga kategori, yaitu positif, netral, dan negatif. Kinerja model dievaluasi menggunakan metrik akurasi dan F1-score.
Hasil: Hasil pengujian menunjukkan bahwa model Naïve Bayes mencapai akurasi sebesar 78% dengan F1-score 0,74. Sementara itu, model IndoBERT memperoleh akurasi 88% dan F1-score 0,86. Perbedaan performa sebesar 10% menunjukkan bahwa model berbasis transformer lebih unggul dalam menangkap konteks semantik Bahasa Indonesia.
Kesimpulan: Penelitian ini menyimpulkan bahwa IndoBERT memiliki kinerja yang lebih baik dibandingkan Naïve Bayes dalam analisis sentimen ulasan mobile banking syariah. Implementasi NLP berbasis deep learning berpotensi meningkatkan efektivitas manajemen kualitas layanan digital melalui sistem pemantauan sentimen otomatis.

Kata Kunci: Analisis Sentimen; IndoBERT; Layanan Digital; Mobile Banking Syariah; Naïve Bayes; Natural Language Processing.

Abstract-
Background: *The development of Islamic mobile banking services in Indonesia requires improvements in the quality of digital service management that focuses on user experience. User reviews on application platforms are an important source of data for understanding user perceptions, satisfaction, and problems encountered. The use of Natural Language Processing (NLP) enables systematic and automatic sentiment analysis to support the evaluation of Islamic banking service quality.*
Objective: *This study aims to analyze the sentiment of user reviews of sharia mobile banking applications and compare the performance of the Naïve Bayes sentiment analysis model as a classic model and IndoBERT as a transformer-based model.*
Methods: *Data was obtained through web scraping of 1,052 user reviews of the Bank NTB Syariah Mobile Banking app from the Google Play Store. Preprocessing steps included case folding, cleaning, stopword removal, stemming, and tokenization. Sentiment labeling was based on user rating scores with three categories: positive, neutral, and negative. Model performance was evaluated using accuracy and F1-score metrics.*
Result: *The test results show that the Naïve Bayes model achieved an accuracy of 78% with an F1-score of 0.74. Meanwhile, the IndoBERT model achieved an accuracy of 88% and an F1-score of 0.86. The 10% difference in performance shows that transformer-based models are superior in capturing the semantic context of the Indonesian language.*

How to Cite: M. Y. Assani, A. Rudiyan, & U. Radiyah, "Analisis Sentimen Ulasan Mobile Banking Syariah Berbasis NLP untuk Evaluasi dan Peningkatan Kualitas Layanan Digital," *Jurnal Bumigora Information Technology (BITe)*, vol. 7, no. 2, pp. 131–140, Des 2025. DOI: 10.30812/bite.v7i2.5943.

Conclusion: *Penelitian ini menyimpulkan bahwa IndoBERT memiliki kinerja yang lebih baik dibandingkan Naïve Bayes dalam analisis sentimen ulasan mobile banking syariah. Implementasi NLP berbasis deep learning berpotensi meningkatkan efektivitas manajemen kualitas layanan digital melalui sistem pemantauan sentimen otomatis.*

Keywords*: Digital Service; IndoBERT; Islamic Mobile Banking; Naïve Bayes; Natural Language Processing; Sentiment Analysis.*

Penulis Korespondensi:

Moh. Yushi Assani,
Program Studi Ilmu Komputer, Universitas Islam Al-Azhar,
Email: yushi@unizar.ac.id

1. PENDAHULUAN

Dalam era digital saat ini, layanan mobile banking telah menjadi salah satu pilar utama dalam mendorong inklusi keuangan dan transformasi ekonomi digital. Kemajuan teknologi seluler memungkinkan pengguna untuk melakukan transaksi keuangan secara cepat, aman, dan efisien tanpa keterbatasan waktu maupun lokasi. Di antara berbagai inovasi yang muncul, *mobile banking* syariah memiliki peran penting dalam memperkuat sistem keuangan berbasis prinsip-prinsip Islam, dengan menyediakan layanan perbankan digital yang sesuai dengan ketentuan syariah [1], [2] Seiring meningkatnya kompetisi antarplatform perbankan digital, pemahaman terhadap kepuasan dan persepsi pengguna menjadi faktor strategis dalam peningkatan kualitas layanan dan pengelolaan manajemen digital [3].

Ulasan dan penilaian pengguna yang dipublikasikan di platform seperti *Google Play Store* menjadi sumber data yang kaya akan informasi mengenai pengalaman dan persepsi pelanggan terhadap suatu aplikasi. Namun, menganalisis ribuan teks ulasan secara manual tidak efisien dan berpotensi menimbulkan bias subjektif. Oleh karena itu, analisis sentimen —sebuah cabang dari *Natural Language Processing* (NLP)— hadir sebagai pendekatan penting untuk mengekstraksi opini pengguna dan mengklasifikasikannya secara otomatis ke dalam kategori sentimen positif, negatif, atau netral [4]. Melalui penerapan teknik NLP, organisasi dapat memperoleh wawasan berbasis data (*data driven insights*) yang berguna dalam proses pengambilan keputusan, pengembangan strategi peningkatan layanan, serta perbaikan pengalaman pengguna secara berkelanjutan [5].

Berbagai penelitian sebelumnya menunjukkan bahwa analisis sentimen dalam konteks layanan keuangan dapat dilakukan dengan pendekatan *machine learning* klasik maupun model berbasis deep learning. Pendekatan klasik seperti Naïve Bayes, *Support Vector Machine* (SVM), dan *Logistic Regression* banyak digunakan karena kesederhanaan dan interpretabilitasnya [6]. Di sisi lain, munculnya arsitektur berbasis *transformer* seperti BERT (*Bidirectional Encoder Representations from Transformers*) telah membawa peningkatan signifikan dalam akurasi klasifikasi sentimen melalui pemahaman konteks semantik yang lebih dalam [7]. Untuk bahasa Indonesia, model IndoBERT yang dilatih secara khusus pada korpus bahasa Indonesia telah terbukti memberikan kinerja unggul pada berbagai tugas NLP [3], [8], [9].

Meskipun penelitian mengenai analisis sentimen telah banyak dilakukan, kajian yang secara khusus menyoroti layanan *mobile banking* syariah di Indonesia masih terbatas. Padahal, dengan meningkatnya adopsi layanan perbankan syariah digital dan jumlah penggunanya yang terus bertambah, analisis terhadap persepsi pengguna menjadi krusial untuk memahami tingkat kepuasan, kendala, dan peluang peningkatan layanan. Oleh karena itu, penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi *mobile banking* syariah dengan memanfaatkan teknik NLP, serta membandingkan performa dua pendekatan klasifikasi sentimen, yaitu model klasik Naïve Bayes dan model berbasis *transformer* IndoBERT .

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi empiris dalam pengembangan metode analisis sentimen berbasis NLP untuk evaluasi persepsi pengguna terhadap layanan perbankan digital. Selain itu, temuan penelitian ini juga diharapkan dapat mendukung lembaga keuangan syariah dalam merumuskan

strategi peningkatan kualitas layanan, memperkuat pengelolaan digital yang berorientasi pada pengguna, dan mendorong transformasi manajemen digital yang selaras dengan prinsip-prinsip syariah.

2. METODE PENELITIAN

2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif komparatif untuk menganalisis performa dua model analisis sentimen berbasis *Natural Language Processing* (NLP) pada ulasan layanan *mobile banking* syariah. Dua model yang dibandingkan adalah Naïve Bayes sebagai representasi model klasik berbasis fitur statistik, dan IndoBERT sebagai model berbasis *Transformer* modern yang telah diadaptasi untuk Bahasa Indonesia. Tujuan utama penelitian ini adalah mengukur sejauh mana kemampuan masing-masing model dalam mengklasifikasikan sentimen pengguna terhadap aplikasi *mobile banking* syariah, yang dalam hal ini diambil dari *Google Play Store*.

2.2. Pra-pemrosesan Teks (*Text Processing*)

Tahapan *preprocessing* dilakukan untuk membersihkan dan menyeragamkan data teks agar siap diproses oleh model *Natural Language Processing* (NLP) [10]. Proses ini diawali dengan case folding, yaitu mengubah seluruh teks menjadi huruf kecil untuk menyeragamkan format penulisan. Selanjutnya dilakukan pembersihan teks (*cleaning*) dengan menghapus URL, angka, tanda baca, emoji, serta karakter non-alfabet menggunakan *regular expression* [11], [12], [13]. Tahap berikutnya adalah *stopword removal*, yaitu menghilangkan kata-kata umum yang tidak memiliki kontribusi signifikan terhadap makna sentimen, seperti “yang”, “dan”, dan “di”, dengan memanfaatkan pustaka Sastrawi. Setelah itu, dilakukan *stemming* untuk mengubah kata ke bentuk dasarnya, misalnya kata menyimpan dan penyimpanan menjadi simpan, menggunakan algoritma *stemming* Bahasa Indonesia dari Sastrawi. Selain itu, diterapkan normalisasi bahasa tidak baku guna menangani penggunaan kata tidak formal, seperti “gk” menjadi “tidak” dan “bgt” menjadi “banget”. Tahap terakhir adalah tokenisasi, yaitu memecah teks menjadi unit-unit kata (token) menggunakan pustaka NLTK. Seluruh rangkaian *preprocessing* ini menghasilkan kolom teks bersih (*clean text*) yang siap digunakan pada tahap ekstraksi fitur dan pelatihan model (lihat Gambar 1).

	clean_text
0	kapan mau ganggu transfer antar bank paraaah
1	apa apa transfer e wallet tidak qris tidak bis...
2	hmmmmmagak kecewa aku orang ntb kenapa gin sih...
3	eh gamaq pir bae aplikasi ini normal wah tahun...
4	udah tidak topup saldo linkaja transfer lain b...

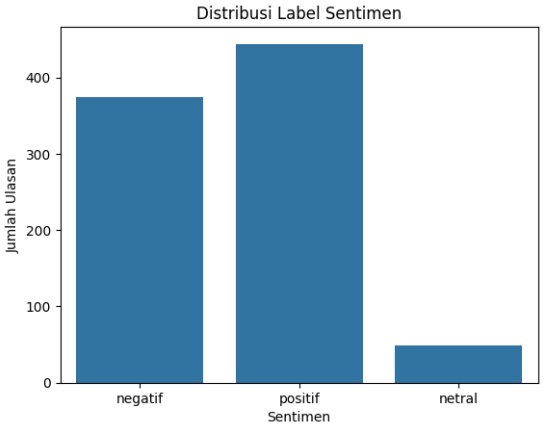
Gambar 1. Hasil *Clean Text*

2.3. *Labeling Sentiment*

Setiap ulasan diberi label sentimen berdasarkan skor rating yang diberikan pengguna pada *Google Play Store*. Proses pelabelan dilakukan dengan mengelompokkan skor 4–5 sebagai sentimen positif, skor 3 sebagai sentimen netral, serta skor 1–2 sebagai sentimen negatif. Pendekatan pelabelan ini mengacu pada metodologi yang telah digunakan dalam penelitian sebelumnya pada analisis ulasan digital, seperti yang dilakukan [14] serta [15]. Selanjutnya, distribusi label sentimen divalidasi melalui inspeksi manual dan analisis frekuensi kata untuk memastikan bahwa setiap kelas sentimen memiliki representasi data yang memadai dan tidak mengalami ketimpangan yang signifikan. Hasil dan distribusi *labeling* sentimen dapat dilihat pada Gambar 2 dan 3.

```
sentimen
positif    444
negatif    374
netral      49
Name: count, dtype: int64
```

Gambar 2. Hasil *labeling* sentimen berdasarkan rating



Gambar 3. Distribusi label sentimen

2.4. Eksplorasi Data (*Exploratory Data Analysis*)

Tahapan eksplorasi data dilakukan untuk memahami karakteristik ulasan secara statistik dan linguistik. Analisis yang dilakukan meliputi distribusi jumlah ulasan pada setiap kategori sentimen, analisis panjang ulasan berdasarkan kelas sentimen, serta frekuensi kemunculan kata-kata dominan yang muncul dalam data. Selain itu, digunakan visualisasi *word cloud* untuk masing-masing kelas sentimen guna menggambarkan pola dan kecenderungan penggunaan kata secara intuitif (Gambar 4). Hasil dari tahap eksplorasi ini digunakan sebagai dasar dalam memahami struktur data dan mendukung pemilihan metode ekstraksi fitur serta model klasifikasi yang tepat.

Gambar 4. Visualisasi sentimen positif

Hasil eksplorasi data menunjukkan bahwa mayoritas pengguna memberikan ulasan positif, dengan kata dominan seperti “mudah”, “baik”, “mantab”, “sangat bantu”, dan “bagus”. Sementara itu, ulasan negatif banyak mengandung kata seperti “tidak bisa”, “kenapa”, “transfer”, dan “buruk”. Sedangkan persentase sentimen per kelasnya adalah: Positif sebesar 51.211%, Negatif sebesar 43.137%, dan Netral sebesar 5.625%.

2.5. Pemodelan Analisis Sentimen

Penelitian ini menggunakan dua pendekatan pemodelan yang dibandingkan, yaitu model klasik dan model *transformer*. Model klasik menggunakan kombinasi TF-IDF (*Term Frequency–Inverse Document Frequency*) sebagai metode ekstraksi fitur dan *Multinomial Naïve Bayes* (MNB) sebagai algoritma klasifikasi. TF-IDF menghitung bobot setiap kata berdasarkan frekuensi kemunculannya di dokumen dan keseluruhan korpus, sedangkan Naïve Bayes mengasumsikan independensi antar fitur untuk memprediksi kelas dengan probabilitas tertinggi. Dataset dibagi menjadi data latih (80%) dan data uji (20%) menggunakan stratified sampling. Evaluasi dilakukan dengan metrik akurasi, *precision*, *recall*, dan *F1-score*. Rumus probabilistik dasar Naïve Bayes ditunjukkan pada Persamaan (1).

$$P\left(C|X\right) = \frac{P\left(C\right)\prod_{i=1}^{n} P\left(\chi_i|C\right)}{P\left(X\right)} \tag{1}$$

Di mana C merupakan kelas yang ingin diprediksi, X merupakan fitur atau data observasi, $P(C \mid X)$ merupakan probabilitas kelas C diberikan data X, $P(X \mid C)$ merupakan probabilitas munculnya data X jika kelas C benar, $P(C)$ = probabilitas awal (prior) kelas C, dan $P(X)$ merupakan probabilitas seluruh data X.

Model kedua menerapkan pendekatan *Transformer-based*, yaitu IndoBERT yang dikembangkan oleh IndoNLP Benchmark Team [8]. Model yang digunakan adalah "indobenchmark/indobert-base-p1" yang diakses melalui pustaka *Hugging Face Transformers*. Tahap pemodelan diawali dengan proses tokenisasi menggunakan *AutoTokenizer* dengan penerapan *padding* dan *truncation* hingga panjang maksimum 128 token. Selanjutnya, dilakukan *fine-tuning* terhadap model IndoBERT pra-latih dengan menambahkan lapisan klasifikasi (*dense layer*) di atas arsitektur dasar BERT untuk menyesuaikan model dengan tugas klasifikasi sentimen. Proses pelatihan dilakukan selama 3 epoch dengan *batch size* 16, *learning rate* awal sebesar 2e-5, serta menggunakan optimizer AdamW. Kinerja model kemudian dievaluasi menggunakan metrik yang sama dengan model klasik, yaitu akurasi, *precision*, *recall*, dan *F1-score*, guna memastikan perbandingan performa yang objektif.

2.6. Evaluasi dan Perbandingan Model

Evaluasi terhadap kedua model dilakukan menggunakan dataset uji yang identik untuk memastikan proses perbandingan yang adil. Setiap model kemudian dianalisis secara terpisah berdasarkan hasil prediksinya. Selanjutnya, kinerja kedua model dibandingkan menggunakan empat metrik utama, yaitu akurasi, presisi, *recall*, dan *F1-score* (lihat Persamaan (2)—(5)).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1\ score = \frac{Precision\ \times\ Recall}{Precision +\ Recall} \tag{5}$$

Di mana TP (*True Positive*) adalah jumlah data positif yang diklasifikasikan dengan benar, FP (*False Positive*) adalah jumlah data negatif yang salah diklasifikasikan sebagai positif, TN (*True Negative*) adalah jumlah data negatif yang diklasifikasikan dengan benar, dan FN (*False Negative*) adalah jumlah data positif yang salah diklasifikasikan sebagai negatif. Hasil yang diharapkan adalah bahwa model IndoBERT akan memberikan akurasi dan *F1-score* lebih tinggi dibandingkan Naïve Bayes. Keunggulan ini didasarkan pada kemampuannya dalam memahami konteks semantik secara lebih mendalam. Selain itu, IndoBERT juga lebih efektif dalam menangkap struktur sintaksis Bahasa Indonesia.

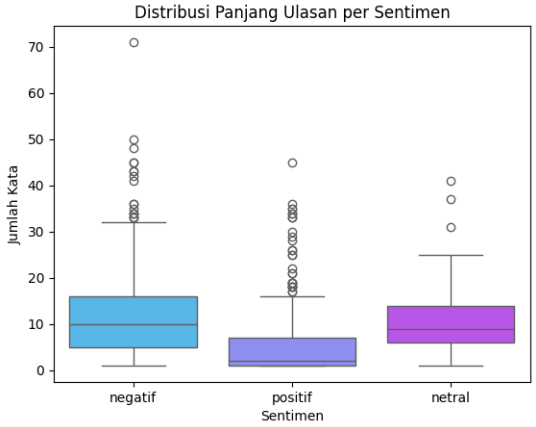
3. HASIL DAN PEMBAHASAN

Dataset yang digunakan terdiri dari 868 ulasan pengguna aplikasi Bank NTB Syariah *Mobile Banking* yang diambil dari *Google Play Store*. Berdasarkan hasil labeling berdasarkan skor rating, distribusi kelas sentimen dapat dilihat pada Tabel 1.

Tabel 1. Ulasan Pengguna

Sentimen	Jumlah	Persentase
Positif	444	51.15%
Negatif	375	43.20%
Netral	49	5.65%

Distribusi ini menunjukkan bahwa mayoritas pengguna memiliki pengalaman positif terhadap aplikasi, namun masih terdapat sejumlah keluhan dengan konteks negatif yang cukup signifikan. Hasil eksplorasi awal memperlihatkan bahwa ulasan positif didominasi oleh kata seperti “mudah”, “baik”, “mantab”, “sangat bantu”, dan “bagus”, sedangkan negatif didominasi oleh kata “tidak bisa”, “kenapa”, “transfer”, dan “buruk”. Panjang rata-rata ulasan negatif lebih tinggi dibandingkan ulasan positif, menandakan bahwa pengguna cenderung menulis ulasan lebih panjang untuk mengeluhkan masalah.



Gambar 5. Distribusi panjang ulasan

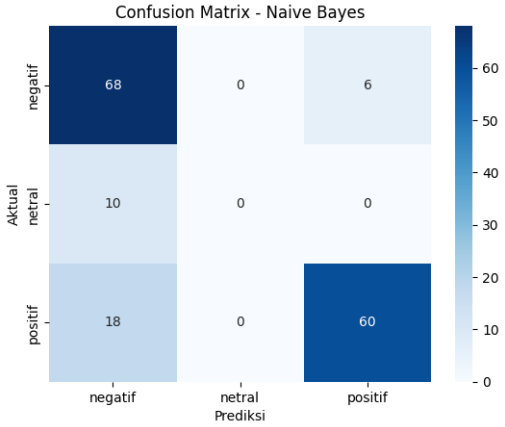
3.1. Evaluasi Model Naïve Bayes (*Baseline Model*)

Model klasik menggunakan pendekatan TF-IDF + Multinomial Naïve Bayes (MNB). Dataset dibagi menjadi 80% data latih dan 20% data uji menggunakan stratified sampling. Setelah pelatihan, hasil evaluasi pada data uji ditunjukkan pada Tabel 2.

Tabel 2. Akurasi Model Naïve Bayes

	precision	Recall	F1-score	Support
Negatif	0.71	0.92	0.80	74
Netral	0.00	0.00	0.00	10
Positif	0.91	0.77	0.83	78
accuracy			0.79	162
macro avg	0.54	0.56	0.54	162
weighted avg	0.76	0.79	0.77	162

Model Naïve Bayes mampu mengklasifikasikan ulasan dengan tingkat akurasi 79%, yang menunjukkan performa cukup baik untuk *baseline model* berbasis statistik. Namun, performa pada kelas netral relatif lebih rendah, kemungkinan karena ulasan netral sering kali memiliki struktur linguistik yang ambigu (mengandung kata positif dan negatif sekaligus). Gambar 6 memperlihatkan visualisasi *confusion matrix* bahwa kesalahan klarifikasi terutama terjadi antara kelas netral dan positif.



Gambar 6. Visualisasi *confusion matrix*

3.2. Evaluasi Model *Transformer: IndoBERT*

Model kedua menggunakan IndoBERT *(indo-benchmark/indobert-base-p1)* yang di-*fine-tune* selama tiga *epoch* dengan *batch size* 16 dan *learning rate* sebesar 2e-5. Proses *fine-tuning* dilakukan untuk menyesuaikan representasi bahasa pralatih IndoBERT dengan karakteristik dan konteks dataset penelitian. Dataset yang sama digunakan pada kedua model guna memastikan bahwa perbandingan kinerja dilakukan secara adil, sehingga perbedaan hasil yang diperoleh dapat dikaitkan secara langsung dengan perbedaan pendekatan pemodelan yang digunakan.

Tabel 3. Akurasi Model IndoBERT

	Precision	**Recall**	**F1-score**	**Support**
Negatif	0.80	0.85	0.82	74
Netral	0.00	0.00	0.00	10
Positif	0.86	0.90	0.88	78
accuracy			0.82	162
macro avg	0.54	0.58	0.57	162
weighted avg	0.78	0.82	0.80	162

Model IndoBERT menunjukkan peningkatan signifikan dibandingkan Naïve Bayes, dengan akurasi sebesar 82% dan F1-score rata-rata 0,57. Kinerja yang lebih tinggi terutama disebabkan oleh kemampuan model *transformer* dalam menangkap konteks semantik antar kata, termasuk dalam kalimat yang kompleks atau berisi ironi. Misalnya, pada ulasan seperti “sudah bagus tapi masih perlu peningkatan”, IndoBERT mampu mengenali nuansa negatif karena pemahaman kontekstual antar frasa.

Pada Tabel 3 menunjukkan bahwa baik algoritma *machine learning* klasik maupun model *transformer* IndoBERT mengalami kegagalan total dalam memprediksi kelas netral, yang ditunjukkan oleh nilai *precision*, *recall*, dan *F1-score* sebesar 0.00. Temuan ini mengungkap peran krusial ketidakseimbangan label dan ambiguitas semantik dalam konteks layanan keuangan berbahasa Indonesia, untuk memitigasi hal semacam ini terjadi di kemudian hari disarankan untuk melakukan pemberian label secara manual pada sebagian data yang memiliki ambiguitas, semisal ulasan yang memberikan bintang 1 namun isinya memberikan kata-kata positif di dalamnya.

3.3. Perbandingan Kinerja Kedua Model

Tabel 4 berikut merangkum perbandingan performa kedua model secara keseluruhan.

Tabel 4. Perbandingan Kinerja

Metrik	**Naïve Bayes**	**IndoBERT**
Akurasi	0.79	0.82
Precision	0.76	0.78
Recall	0.79	0.82

Metrik	Naïve Bayes	IndoBERT
F1-Score	0.77	0.80
Waktu Training	±3 menit	±54 menit

Dari hasil perbandingan terlihat IndoBERT unggul di seluruh matrix utama, Naïve Bayes tetap relevan untuk implementasi ringan misalnya pada sistem *real time* dengan sumber daya terbatas. Yang perlu dipertimbangkan adalah bahwa IndoBERT membutuhkan GPU dan waktu *training* yang lebih lama dibandingkan dengan Naïve Bayes, namun dapat memberikan hasil yang lebih akurat dan stabil terutama jika data yang akan dilatih memiliki kompleksitas yang tinggi.

3.4. Analisis Kesalahan

Beberapa kesalahan klasifikasi yang ditemukan umumnya disebabkan oleh karakteristik teks ulasan. Kesalahan tersebut antara lain muncul pada ulasan dengan ambiguitas bahasa campuran, seperti "lumayan walau kadang sering gangguan jaringan", di mana model tidak selalu mampu mengenali konteks utama secara tepat. Selain itu, ulasan yang sangat singkat, misalnya "mantap", "keren", "lumayan", "good", atau "ok", cenderung kurang memberikan informasi kontekstual yang memadai untuk menghasilkan prediksi yang akurat. Kesalahan juga ditemukan pada penggunaan bahasa tidak baku, seperti "gk", "tdk", dan "okey", yang meskipun telah melalui proses normalisasi, masih berpotensi menimbulkan kendala pada tahap tokenisasi awal. Dalam kondisi-kondisi tersebut, model IndoBERT menunjukkan kinerja yang sedikit lebih baik karena telah dilatih menggunakan korpus Bahasa Indonesia yang lebih beragam, termasuk teks informal dari media sosial.

4. KESIMPULAN

Secara umum, hasil penelitian ini menunjukkan bahwa pendekatan deep learning berbasis transformer (IndoBERT) lebih unggul dibandingkan pendekatan klasik Naïve Bayes dalam menganalisis sentimen Bahasa Indonesia pada domain perbankan syariah digital. Hasil pengujian menunjukkan bahwa model IndoBERT menghasilkan akurasi yang lebih tinggi yaitu 82% dibandingkan model Naïve Bayes dengan akurasi 79%. Hal ini menunjukkan kemampuan Transformer-based model dalam menangkap konteks semantik Bahasa Indonesia yang kompleks, terutama pada ulasan dengan struktur kalimat informal.

DAFTAR PUSTAKA

[1] F. Alqahtani dan D. G. Mayes, "Financial stability of Islamic banking and the global financial crisis: Evidence from the Gulf Cooperation Council," en, *Economic Systems*, vol. 42, no. 2, pp. 346–360, Jun. 2018. DOI: 10.1016/j.ecosys.2017.09.001.

[2] C. Rosanti, F. A. Artanto, dan R. E. Saputra, "Analisis Sentiment Pengguna Aplikasi Mobile Banking Pada Bank Syariah Dengan Support Vector Regression," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 4, no. 8, pp. 341–347, Jan. 5, 2025. DOI: 10.52436/1.jpti.460.

[3] W. F. Sari, R. Rahim, dan F. Adrianto, "Sentiment Analysis Using Big Data User Reviews on Mobile Banking Performance in Indonesia," *IAR Journal of Business Management*, vol. 04, no. 01, pp. 1–7, May 15, 2023. DOI: 10.47310/iarjbm.2023.v04i01.028.

[4] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, en. Cambridge University Press, Oct. 15, 2020, 451 pp. Google Books: PdX7DwAAQBAJ.

[5] E. Cambria et al., "New Avenues in Opinion Mining and Sentiment Analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15–21, Mar. 2013. DOI: 10.1109/MIS.2013.30.

[6] W. Medhat, A. Hassan, dan H. Korashy, "Sentiment analysis algorithms and applications: A survey," en, *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014. DOI: 10.1016/j.asej.2014.04.011.

[7] J. Devlin et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." version 2, pre-published.

[8] F. Koto et al. "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP." version 1, pre-published.

[9] A. A. P. Simarmata dan T. B. Sasongko, "Sentiment Analysis on BRImo Application Reviews Using IndoBERT," *Journal of Applied Informatics and Computing*, vol. 9, no. 3, pp. 851–862, Jun. 3, 2025. DOI: 10.30871/jaic.v9i3.8162.

[10] Y. Setiawan dan L. A. Wulandhari, "Comparative Analysis of IndoBERT and LSTM for Multi-Label Text Classification of Indonesian Motivation Letter," *Jurnal Online Informatika*, vol. 10, no. 2, pp. 260–269, Aug. 17, 2025. DOI: 10.15575/join.v10i2.1499.

[11] R. Z. Suchrady dan A. Purwarianti. "Indo LEGO-ABSA: A Multitask Generative Aspect Based Sentiment Analysis for Indonesian Language." arXiv: 2311.01757 [cs], pre-published.

[12] T. A. Bawana, F. Mansor, dan K. Noordin, "Gauging Customer Sentiment Regarding Indonesian Islamic Digital Banks," *AL-IKTISAB: Journal of Islamic Economic Law*, vol. 8, no. 1, pp. 101–118, Nov. 2, 2024. DOI: 10.21111/aliktisab.v8i2.12838.

[13] E. Nurmiati, M. Qomarul Huda, dan S. Mitsalina, "Sentiment Analysis and Topics Modeling on Mobile Banking Reviews of Sharia Bank in Indonesia Using Naive Bayes and Latent Dirichlet Allocation," in *2024 12th International Conference on Cyber and IT Service Management (CITSM)*, Batam, Indonesia: IEEE, Oct. 3, 2024, pp. 1–6. DOI: 10.1109/CITSM64103.2024.10775521.

[14] M. M. Dakwah et al., "Sentiment Analysis on Marketplace in Indonesia using Support Vector Machine and Naïve Bayes Method," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 10, no. 1, p. 39, Feb. 16, 2024. DOI: 10.26555/jiteki.v10i1.28070.

[15] F. Indriani et al., "Comparative Evaluation of IndoBERT, IndoBERTweet, and mBERT for Multilabel Student Feedback Classification," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 6, pp. 748–757, Dec. 27, 2024. DOI: 10.29207/resti.v8i6.6100.

[Halaman ini sengaja dikosongkan.]