

PERBANDINGAN KINERJA *SUPPORT VECTOR MACHINE*, *LOGISTIC REGRESSION*, DAN *INDOBERT* PADA ANALISIS SENTIMEN *WONDR* BY BNI

Melisya Sesy Amelia¹, Shafa Sabrina Almas², Siti Mukaromah³, Eka Dyar Wahyuni⁴

^{1,2,3,4}Universitas Pembangunan Nasional “Veteran” Jawa Timur, Indonesia

¹23082010068@student.upnjatim.ac.id

²23082010055@student.upnjatim.ac.id

³sitimukaromah.si@upnjatim.ac.id

⁴ekawahyuni.si@upnjatim.ac.id

Received: 07-06- 2026

Revised: 21-06-2026

Approved: 28-06-2026

ABSTRAK

Perkembangan layanan perbankan digital mendorong institusi perbankan untuk terus meningkatkan kualitas aplikasi perbankan mobile banking untuk memenuhi kebutuhan dan kepuasan nasabah. Penelitian ini dilakukan untuk menganalisis serta membandingkan tingkat efektivitas algoritma klasifikasi sentimen yang meliputi SVM, Logistic Regression, dan IndoBERT untuk menganalisis opini pengguna yang disampaikan melalui ulasan aplikasi *WONDR* by BNI. Data penelitian diperoleh melalui proses ekstraksi umpan balik pengguna yang tersedia pada Google Play Store, yang memiliki jumlah ulasan 10.001. Setelah melalui tahap data cleaning, dataset berkurang menjadi 9.999 data karena penghapusan data duplikat dan data yang tidak valid, kemudian sebanyak 9.998 data digunakan pada tahap analisis setelah menghapus satu data yang tidak dapat diproses akibat karakter non-teks. Proses penelitian meliputi pemrosesan awal teks berbahasa Indonesia, pelabelan sentimen menggunakan Groq API berbasis Large Language Model (LLM), pembagian dataset menjadi data latih dan data uji dengan rasio 80:20, ekstraksi fitur TF-IDF untuk model SVM dan Logistic Regression, serta proses fine-tuning IndoBERT sebagai model berbasis transformer. Hasil dari pengujian menunjukkan bahwa IndoBERT yang telah disempurnakan memberikan hasil terbaik dengan accuracy sebesar 88,9%, presicion 89,3%, recall 88,9%, dan F1-score 89,1%. Sementara itu, SVM memperoleh accuracy 85,5%, precision 83,6%, recall 85,5%, dan F1-score 84,0%, sedangkan Logistic Regression memperoleh accuracy 85,5%, precision 83,5%, recall 85,5%, serta nilai F1-score sebesar 83,0%. Berdasarkan hasil penelitian, diketahui bahwa algoritma berbasis transformer menunjukkan kemampuan yang lebih unggul dalam konteks bahasa Indonesia jika dibandingkan dengan model machine learning berbasis TF-IDF. Hasil penelitian ini diharapkan dapat memberikan rujukan bagi peneliti maupun praktisi guna menentukan metode klasifikasi sentimen yang paling sesuai untuk mengevaluasi ulasan pengguna aplikasi perbankan digital berdasarkan data ulasan yang tersedia.

Kata Kunci: Analisis sentimen, *WONDR* by BNI, Support Vector Machine, Logistic Regression, IndoBERT

PENDAHULUAN

Adopsi teknologi digital yang semakin luas telah mengubah cara industri perbankan beroperasi, khususnya dalam menghadirkan layanan perbankan digital berbasis aplikasi *mobile* kepada nasabah. Kehadiran *mobile banking* memberikan kemudahan bagi nasabah untuk melakukan beragam transaksi keuangan dengan lebih efisien, cepat dan tanpa perlu datang ke kantor cabang. Menurut berbagai penelitian terkini, mutu layanan digital dianggap sebagai salah satu elemen penting yang berpengaruh pada kepuasan serta kesetiaan pengguna terhadap

layanan perbankan [1]. Dengan demikian, bank terus berusaha untuk meningkatkan mutu aplikasi yang mereka sediakan agar dapat memenuhi kebutuhan dan ekspektasi nasabah.

Salah satu aplikasi perbankan digital yang dikembangkan untuk mendukung transformasi layanan perbankan adalah WONDR by BNI. Aplikasi ini menyediakan berbagai fitur transaksi dan pengelolaan keuangan yang dapat diakses melalui perangkat seluler. Sebagai aplikasi yang digunakan oleh banyak nasabah, WONDR by BNI mendapat banyak umpan balik dari pengguna pada *platform Google Play Store*. Ulasan tersebut menggambarkan pengalaman pengguna saat memakai aplikasi ini, dimulai dari bentuk apresiasi, kritik, maupun saran perbaikan. Data yang terkandung dalam umpan balik pengguna memiliki nilai yang penting karena dapat dimanfaatkan sebagai bahan evaluasi untuk meningkatkan kualitas layanan digital yang diberikan.

Pentingnya analisis terhadap ulasan tersebut semakin meningkat karena WONDR by BNI merupakan aplikasi *mobile banking* generasi terbaru yang diperkenalkan oleh BNI sebagai pengganti aplikasi BNI *Mobile Banking*. Pada masa awal peluncurannya, aplikasi ini memperoleh perhatian yang tinggi dari pengguna karena menghadirkan perubahan antarmuka serta berbagai fitur baru. Kondisi tersebut menyebabkan banyak pengguna memberikan ulasan mengenai pengalaman migrasi, stabilitas sistem, maupun kemudahan penggunaan aplikasi. Oleh karena itu, analisis sentimen terhadap ulasan pada periode awal peluncuran menjadi penting sebagai bahan evaluasi terhadap kualitas layanan digital yang diberikan.

Namun, jumlah ulasan yang terus bertambah setiap hari menyebabkan proses analisis secara manual menjadi kurang efisien dan membutuhkan waktu yang relatif lama. Untuk mengatasi hal tersebut, diperlukan metode otomatis yang dapat mengenali serta mengelompokkan persepsi pengguna berdasarkan sentimen yang diekspresikan. Salah satu pendekatan yang banyak diterapkan adalah klasifikasi sentimen, yaitu teknik dalam bidang *Natural Language Processing* (NLP) yang berfungsi untuk mengidentifikasi dan mengelompokkan umpan balik ke dalam klasifikasi negatif, positif, atau netral [2]. Analisis sentimen telah digunakan di beragam bidang, seperti media *social*, *e-commerce*, layanan publik, hingga aplikasi digital, untuk membantu organisasi memahami persepsi pengguna terhadap produk atau layanan tertentu.

Penelitian analisis sentimen telah mengadopsi berbagai metode klasifikasi teks. Kemampuan dalam mengolah representasi teks berbasis TF-IDF menjadikan SVM dan *Logistic Regression* sebagai dua metode *machine learning* yang banyak digunakan untuk tugas klasifikasi sentimen. Pada penelitian yang dilakukan oleh Satya dkk. [3] memperlihatkan bahwa SVM dan *Logistic Regression* dapat digunakan secara efektif untuk mengelompokkan sentimen yang terkandung dalam ulasan pengguna aplikasi. SVM menawarkan sejumlah keunggulan dalam membentuk *hyperplane* optimal untuk memisahkan kelas data, sedangkan *Logistic Regression* dikenal sebagai algoritma yang sederhana, efisien, dan mudah diimplementasikan pada berbagai kasus klasifikasi teks.

Seiring perkembangan teknologi NLP, model berbasis *Transformer* semakin

populer karena mereka bisa memahami konteks bahasa dengan lebih efektif daripada metode *machine learning* tradisional. Di antara berbagai model NLP yang tersedia, *IndoBERT* banyak dimanfaatkan untuk analisis teks berbahasa Indonesia. Model ini dirancang dengan mengacu pada arsitektur BERT dan telah dilatih menggunakan korpus besar bahasa Indonesia sehingga mampu menangkap hubungan dan makna kata berdasarkan konteks penggunaannya. [4]. Beberapa penelitian mengungkapkan bahwa *IndoBERT* mampu menghasilkan kinerja yang lebih unggul dibandingkan metode yang menggunakan TF-IDF pada berbagai tugas NLP, termasuk analisis sentimen [5]. Keunggulan tersebut diperoleh karena *IndoBERT* tidak hanya mempertimbangkan frekuensi kemunculan kata, tetapi juga memahami keterkaitan dan konteks antara kata-kata dalam suatu kalimat.

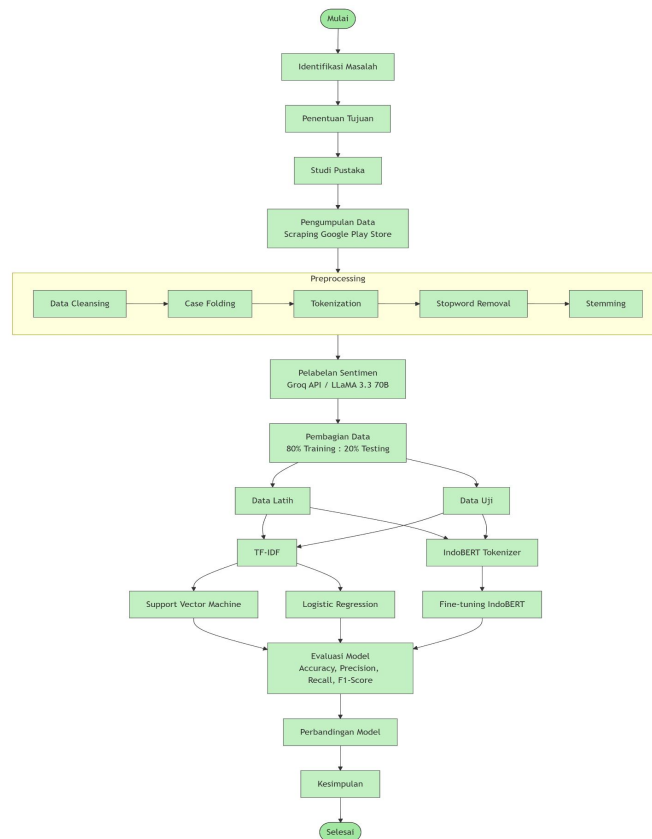
Keunggulan tersebut menjadi penting untuk diterapkan pada analisis ulasan aplikasi perbankan digital. Karakteristik ulasan pengguna umumnya mengandung istilah layanan keuangan, singkatan, bahasa tidak baku, serta kalimat yang memuat lebih dari satu opini dalam satu ulasan. Karakteristik tersebut menyebabkan pendekatan berbasis frekuensi kata seperti TF-IDF belum selalu mampu merepresentasikan makna kalimat secara utuh. Oleh karena itu, model berbasis *contextual embedding* seperti *IndoBERT* menjadi lebih sesuai karena mampu memahami hubungan antar kata berdasarkan konteks penggunaannya sehingga dapat meningkatkan kualitas klasifikasi sentimen pada teks berbahasa Indonesia.

Penelitian terdahulu menunjukkan bahwa model berbasis Transformer cenderung mencapai angka *accuracy* dan *F1-score* yang lebih tinggi daripada metode *machine learning* tradisional pada dataset berbahasa Indonesia [5]. Meskipun berbagai penelitian telah menunjukkan bahwa model berbasis Transformer mampu memberikan performa yang lebih baik dibandingkan metode *machine learning* tradisional pada data berbahasa Indonesia, sebagian besar penelitian masih berfokus pada data media sosial, *e-commerce*, maupun isu politik digital. Penelitian yang membandingkan *Support Vector Machine* (SVM), *Logistic Regression*, dan *IndoBERT* pada ulasan pengguna aplikasi *mobile banking*, khususnya WONDR by BNI, masih relatif terbatas. Selain itu, pemanfaatan ulasan pengguna WONDR by BNI sebagai sumber informasi untuk mengevaluasi kualitas layanan digital perbankan juga belum banyak dilakukan.

Berdasarkan uraian tersebut, terdapat *research gap* berupa keterbatasan penelitian yang membandingkan model *machine learning* tradisional dan model berbasis Transformer dalam proses analisis sentimen pada tanggapan pengguna aplikasi perbankan digital dalam bahasa Indonesia. Kondisi tersebut mendorong dilakukannya penelitian untuk menguji serta membandingkan kemampuan SVM, *Logistic Regression*, dan *IndoBERT* dalam mengidentifikasi serta mengklasifikasikan sentimen yang terkandung dalam tanggapan pengguna WONDR by BNI.

METODE PENELITIAN

Untuk mencapai tujuan penelitian, serangkaian tahapan penelitian dilakukan sebagaimana dijelaskan berikut ini:



Gambar 2.1 Tahapan Penelitian

Adapun penjelasan terkait alur metode penelitian adalah sebagai berikut:

Identifikasi Masalah

Penentuan permasalahan penelitian dilakukan dengan mengkaji umpan balik pengguna WONDR by BNI yang tersedia di *Google Play Store*. Data tersebut memuat berbagai bentuk penilaian terhadap kualitas aplikasi, sehingga dapat memberikan gambaran mengenai persepsi pengguna. Tanggapan yang terkumpul menunjukkan adanya variasi sentimen, baik yang positif, netral, maupun negatif, yang mencerminkan pengalaman pengguna dalam menggunakan aplikasi.

Penentuan Tujuan

Permasalahan yang telah ditemukan menjadi dasar dalam penentuan tujuan penelitian, yaitu membandingkan performa SVM, *Logistic Regression*, dan *IndoBERT* pada proses klasifikasi sentimen ulasan pengguna WONDR by BNI. Kinerja setiap model kemudian dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk memperoleh gambaran mengenai kemampuan masing-masing metode dalam menghasilkan klasifikasi yang akurat.

Studi Pustaka

Sebagai dasar penyusunan penelitian, dilakukan kajian literatur terhadap berbagai sumber yang relevan dengan topik yang dibahas. Referensi yang ditelaah mencakup teori analisis sentimen, *Natural Language Processing (NLP)*, SVM, *Logistic Regression*, *IndoBERT*, serta sejumlah penelitian terdahulu yang berkaitan dengan metode maupun objek penelitian.

Scraping Data

Sebagai sumber data penelitian, ulasan pengguna pada *Google Play Store* dikumpulkan melalui proses *scraping*. Kegiatan pengambilan data tersebut memanfaatkan bahasa pemrograman *Python* yang terintegrasi dengan pustaka *google-play-scraping* untuk mengekstrak informasi yang diperlukan dari halaman aplikasi. Data yang berhasil dikumpulkan berupa komentar dan penilaian dari pengguna aplikasi WONDR by BNI yang berisi nama pengguna, tanggal ulasan, penilaian, serta konten ulasan. Hasil *scraping* menghasilkan sebanyak 10.001 ulasan pengguna.

Preprocessing

Pada penelitian klasifikasi sentimen, *preprocessing* menjadi tahapan yang esensial guna meningkatkan kualitas data sebelum digunakan pada tahap klasifikasi. Tanggapan pengguna yang dikumpulkan dari *Google Play Store* masih berada dalam bentuk teks mentah yang belum terstruktur. Data tersebut memuat berbagai elemen, seperti angka, simbol, tanda baca, penggunaan bahasa nonbaku, serta kata-kata yang kurang memberikan informasi penting bagi proses analisis sentimen. Keberadaan elemen-elemen tersebut dapat mempengaruhi kualitas representasi data dan performa model yang digunakan. Oleh karena itu, tahap *preprocessing* diperlukan untuk membersihkan dan menormalkan data guna meningkatkan kualitas representasi teks serta kinerja model klasifikasi sentimen. Tahapan *preprocessing* yang dilakukan mencakup:

1. *Data Cleaning*

Pada tahap ini dilakukan pembersihan data guna mengurangi keberadaan informasi yang kurang relevan sehingga kualitas data menjadi lebih baik, seperti karakter khusus, URL, emoji, angka, simbol, data kosong (*missing value*), serta data duplikat yang dapat mempengaruhi hasil analisis.

2. *Case Folding*

Proses ini bertujuan untuk mengkonversi seluruh karakter ke huruf kecil dilakukan untuk menjaga keseragaman data teks, sehingga perbedaan penulisan huruf besar dan huruf kecil tidak memengaruhi proses pengolahan data.

3. *Stopword Removal*

Pada tahap ini dilakukan eliminasi terhadap kata-kata yang memiliki kontribusi informasi rendah meskipun sering muncul dalam dokumen teks. Kata-kata seperti *yang, dan, di, ke, dari, dan untuk* termasuk dalam kelompok *stopword* karena keberadaannya umumnya tidak memengaruhi klasifikasi sentimen. Dengan menghapus kata-kata tersebut, proses analisis dapat lebih menitikberatkan pada kata-kata yang merepresentasikan opini atau penilaian pengguna.

4. *Tokenization*

Tahap tokenisasi dilakukan dengan memecah teks menjadi token-token individual sehingga dapat digunakan pada proses pengolahan dan analisis data selanjutnya.

5. *Stemming*

Tahap ini berfungsi untuk menormalkan variasi bentuk kata melalui proses pengubahan kata berimbuhan menjadi kata dasar. Hasilnya, kata-kata yang memiliki akar makna yang sama akan memiliki representasi yang lebih konsisten dalam dataset. Sebagai contoh, kata *membantu*, *dibantu*, dan *membantukan* dapat direduksi menjadi kata dasar *bantu*.

Tabel 2.1 Contoh Hasil Transformasi Data

Tahapan	Hasil Transformasi
Teks Mentah	Aplikasi WONDR bagus banget!!! tapi gak bisa login
<i>Case Folding</i>	aplikasi wondr bagus banget!!! tapi gak bisa login
<i>Data Cleaning</i>	aplikasi wondr bagus banget tapi gak bisa login
<i>Stopword Removal</i>	aplikasi wondr bagus banget gak bisa login
<i>Stemming</i>	aplikasi wondr bagus banget tidak bisa login
<i>Tokenization</i>	["aplikasi", "wondr", "bagus", "banget", "tidak", "bisa", "login"]

Pelabelan Sentimen

Pelabelan sentimen adalah tahapan pengklasifikasian ulasan pengguna ke dalam kategori sentimen tertentu sesuai dengan opini yang disampaikan. Dalam penelitian ini, pelabelan dilakukan dengan bantuan Groq API berbasis *Large Language Model* (LLM). Pada proses pelabelan sentimen, Groq API digunakan dengan teknik zero-shot prompting, yaitu model diminta mengklasifikasikan setiap ulasan tanpa diberikan contoh sebelumnya. Prompt berisi instruksi agar model hanya menghasilkan salah satu label, yaitu *positif*, *negatif*, atau *netral*, berdasarkan isi ulasan yang diberikan. Kelas sentimen positif menunjukkan bahwa pengguna memberikan tanggapan yang baik serta merasa puas terhadap layanan atau fitur yang tersedia pada aplikasi, sentimen negatif menunjukkan adanya keluhan atau ketidakpuasan terhadap layanan dan fitur aplikasi sedangkan sentimen netral menunjukkan opini yang bersifat informatif. Untuk memastikan kualitas hasil pelabelan, dilakukan validasi terhadap 100 data ulasan yang dipilih secara acak. Hasil pelabelan dari Groq API kemudian dibandingkan dengan pelabelan manual yang dilakukan oleh peneliti.

Pembagian Data

Dataset yang telah melalui proses pelabelan kemudian dipartisi menjadi dua kelompok data, yaitu *training set* dan *testing set*. Pemisahan dilakukan

menggunakan metode *hold-out validation* dengan perbandingan 80:20, di mana sebagian besar data dimanfaatkan untuk membangun model, sedangkan sisanya digunakan untuk mengukur performa model terhadap data baru.

Dari keseluruhan 9.998 ulasan yang telah melalui proses pelabelan sentimen, sebanyak 7.998 data dialokasikan untuk tahap pelatihan model, sedangkan 2.000 data sisanya digunakan sebagai data pengujian. Data pelatihan dimanfaatkan untuk membangun model klasifikasi menggunakan SVM, *Logistic Regression*, dan *IndoBERT*. Adapun data pengujian berperan sebagai dasar untuk menilai kemampuan masing-masing model dalam mengklasifikasikan data baru yang belum pernah digunakan selama proses pembelajaran.

Pembagian dataset ke dalam data pelatihan dan data pengujian dilakukan untuk memastikan model memperoleh pembelajaran dari data yang tersedia sekaligus dapat diuji menggunakan data yang belum pernah dikenali sebelumnya. Pendekatan ini memungkinkan penilaian performa model dilakukan secara lebih objektif dalam menggambarkan kemampuannya menghadapi data baru.

Pelatihan Model

Tahapan ini berfokus pada pengembangan model klasifikasi berdasarkan data yang telah melalui proses pembersihan, pelabelan, dan pemisahan dataset. Tiga model yang digunakan, yaitu SVM, *Logistic Regression*, dan *IndoBERT*, dibangun menggunakan data pelatihan dan kemudian dievaluasi menggunakan data pengujian. Evaluasi tersebut bertujuan untuk menilai kemampuan setiap model dalam mengenali kecenderungan sentimen yang terkandung dalam ulasan pengguna serta mengklasifikasikannya ke dalam kategori negatif, positif, dan netral.

1. Transformasi teks menjadi fitur TF-IDF

Sebelum model SVM dan *Logistic Regression* dilatih, data ulasan yang masih dalam bentuk teks terlebih dahulu diubah menjadi representasi numerik. Proses tersebut dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) yang bertujuan menghasilkan vektor fitur sebagai input bagi algoritma *machine learning*.

Dalam TF-IDF, setiap kata diberi bobot berdasarkan kombinasi antara frekuensi kemunculannya dalam dokumen dan distribusinya pada seluruh korpus. Pendekatan ini menyebabkan kata yang bersifat spesifik terhadap suatu dokumen mendapatkan nilai yang lebih tinggi, sementara kata yang bersifat umum dan sering muncul di berbagai dokumen memiliki nilai yang lebih rendah karena kontribusinya dalam membedakan dokumen dianggap kecil.

Komputasi nilai TF-IDF dilakukan menggunakan Persamaan (1).

$$TFIDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

dengan:

$$IDF(t) = \log \frac{N}{df(t)} \quad (2)$$

Keterangan:

TF(t,d) = jumlah kemunculan kata t dalam dokumen d .

IDF(t) = nilai *inverse document frequency* dari kata *t*.
N = jumlah seluruh dokumen.
df(t) = banyaknya dokumen yang mengandung kata *t*.

Vektor fitur yang dihasilkan dari proses pembobotan menggunakan metode TF-IDF selanjutnya digunakan sebagai representasi numerik teks dan dijadikan masukan (*input*) dalam pada tahap pelatihan model sekaligus proses klasifikasi sentimen yang dibangun menggunakan metode *Support Vector Machine (SVM)* dan *Logistic Regression*.

2. *Support Vector Machine (SVM)*

Dalam SVM, proses klasifikasi dilakukan dengan mencari batas pemisah terbaik antara kelompok data. Batas tersebut berupa *hyperplane* yang dipilih sedemikian rupa agar pemisahan antar kelas dapat berlangsung secara optimal. Algoritma ini dikenal efektif dalam mengolah data dengan jumlah fitur yang besar atau berdimensi tinggi, sehingga sering dimanfaatkan pada berbagai penelitian klasifikasi teks, termasuk analisis sentimen.

Fungsi keputusan pada SVM dinyatakan pada Persamaan (3).

$$f(x) = wTx + b \quad (3)$$

dengan:

w = vektor bobot (*weight*).

x = vektor data masukan.

b = nilai bias.

Pada proses pelatihan, model SVM menerima input berupa fitur hasil pembobotan TF-IDF yang berasal dari kumpulan data ulasan pengguna aplikasi WONDR by BNI. Representasi tersebut digunakan untuk mempelajari pola kemunculan kata yang berkaitan dengan kecenderungan sentimen sehingga model dapat melakukan pengelompokan kelas secara lebih akurat.

3. *Logistic Regression*

Pada penelitian ini, *Logistic Regression* dimanfaatkan sebagai algoritma klasifikasi untuk mengelompokkan sentimen yang terdapat pada ulasan pengguna. Model ini menggunakan fungsi sigmoid untuk mengonversi hasil perhitungan menjadi nilai probabilitas yang mencerminkan kecenderungan data terhadap kategori spesifik. Algoritma ini banyak diterapkan dalam berbagai tugas klasifikasi teks karena memiliki mekanisme yang sederhana, efisien secara komputasi, dan mampu menghasilkan klasifikasi yang baik.

Fungsi sigmoid pada *Logistic Regression* ditunjukkan pada Persamaan (4).

$$P(y) = \frac{1}{1+e^{-z}} \quad (4)$$

dengan:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (5)$$

Keterangan:

$P(y)$ = probabilitas suatu kelas.
 β = parameter model.
 x = fitur masukan.

4. *IndoBERT*

IndoBERT adalah model bahasa yang didasarkan pada arsitektur *Transformer* yang telah dilatih sebelumnya dengan kumpulan data besar bahasa Indonesia sehingga model dapat menangkap konteks dan struktur kebahasaan Indonesia dengan lebih baik. Berbeda dengan model SVM dan *Logistic Regression*, *IndoBERT* tidak memerlukan tahap ekstraksi fitur menggunakan metode TF-IDF karena model telah memiliki mekanisme representasi bahasa secara kontekstual.

Pada penelitian ini, proses pelatihan dilakukan melalui tahap *fine-tuning* menggunakan data latih ulasan aplikasi WONDR by BNI yang telah diberi label sentimen. Sebelum memasuki tahap pelatihan, setiap teks ulasan terlebih dahulu dikonversi ke dalam bentuk token melalui mekanisme tokenisasi menggunakan *WordPiece Tokenizer*. Selanjutnya, model mempelajari hubungan kontekstual antar kata dan menyesuaikan parameter model berdasarkan dataset penelitian sehingga model mampu mengelompokkan sentimen menjadi tiga kelas, yakni positif, negatif, dan netral.

Evaluasi Model

Setelah proses pelatihan selesai, masing-masing model dievaluasi menggunakan data testing yang mencakup 20% dari keseluruhan dataset, sedangkan 80% sisanya digunakan sebagai data training. Proses evaluasi dilakukan terhadap tiga model, yaitu *Support Vector Machine (SVM)*, *Logistic Regression*, dan *IndoBERT Fine-tuned*.

Kinerja setiap model diukur menggunakan empat metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, *classification report* digunakan untuk mengevaluasi performa masing-masing model pada setiap kelas sentimen, yaitu positif, negatif, dan netral. Selanjutnya, hasil evaluasi dari ketiga model dibandingkan untuk menentukan metode yang memberikan performa terbaik dalam melakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi WONDR by BNI.

HASIL DAN PEMBAHASAN

Setelah proses pelatihan selesai, dilakukan pengujian terhadap model *Support Vector Machine (SVM)*, *Logistic Regression*, dan *IndoBERT Fine-tuned* menggunakan data testing. Pengujian bertujuan untuk mengetahui kemampuan masing-masing model dalam mengklasifikasikan sentimen ulasan pengguna aplikasi WONDR by BNI. Hasil evaluasi disajikan berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*, kemudian dianalisis untuk mengetahui kelebihan serta keterbatasan setiap model. Selain itu, ringkasan performa klasifikasi pada setiap kelas sentimen (positif, negatif, dan netral) untuk ketiga model disajikan pada Tabel 3.1 sebagai dasar dalam analisis hasil pengujian.

Tabel 3.1 Ringkasan *Classification Report* Ketiga Model

Model	Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
<i>Support Vector Machine</i>	Negatif	0.89	0.95	0.92
	Netral	0.50	0.22	0.31
	Positif	0.82	0.81	0.81
<i>Logistic Regression</i>	Negatif	0.87	0.98	0.92
	Netral	0.58	0.13	0.21
	Positif	0.84	0.78	0.81
<i>IndoBERT</i>		0.89	0.88	0.89

1. Evaluasi Kinerja Model *Support Vector Machine*

Berdasarkan hasil evaluasi yang telah dilakukan, model Support Vector Machine (SVM) menghasilkan nilai accuracy sebesar 0,855, precision sebesar 0,836, recall sebesar 0,855, dan F1-score sebesar 0,840. Perolehan nilai tersebut menunjukkan bahwa model SVM memiliki kemampuan yang baik dalam mengidentifikasi serta mengelompokkan sentimen pada data ulasan yang diberikan oleh pengguna aplikasi WONDR by BNI.

Berdasarkan *classification report*, memperlihatkan bahwa algoritma SVM memiliki tingkat keberhasilan yang tinggi dalam mendeteksi kelas sentimen negatif, sebagaimana ditunjukkan oleh nilai *recall* yang besar. Kondisi tersebut menunjukkan bahwa mayoritas data berlabel negatif dapat dikenali secara akurat oleh model. Di sisi lain, performa klasifikasi pada kelas sentimen netral masih perlu ditingkatkan karena tingkat pengenalannya belum sebaik kelas negatif. Rendahnya performa pada kelas ini dapat dipengaruhi oleh karakteristik ulasan yang bersifat lebih objektif dan kurang menunjukkan kecenderungan sentimen yang jelas, sehingga batas antara kelas netral dengan kelas sentimen lainnya menjadi lebih sulit dibedakan oleh model.

Selain itu, penggunaan representasi TF-IDF menyebabkan model hanya mempertimbangkan frekuensi kemunculan kata tanpa memperhatikan konteks serta keterkaitan antar kata dalam kalimat. Dampaknya,, SVM mengalami kesulitan ketika menghadapi ulasan yang memiliki makna kontekstual atau menggunakan variasi bahasa yang tidak baku. Meskipun demikian, SVM tetap menjadi metode yang efisien karena memiliki efisiensi komputasi yang baik dengan waktu pelatihan yang relatif singkat.

2. Evaluasi Kinerja Model *Logistic Regression*

Pengujian yang dilakukan menunjukkan bahwa *Logistic Regression* memperoleh nilai *accuracy* sebanyak 0,855, *precision* sebanyak 0,835, *recall* sebanyak 0,855, serta *F1-score* sebanyak 0,830. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan sentimen umpan balik pengguna aplikasi WONDR by BNI dengan performa yang cukup baik.

Berdasarkan hasil evaluasi yang diperoleh, performa *Logistic Regression* menghasilkan tingkat kinerja yang tidak jauh berbeda dibandingkan dengan model *Support Vector Machine (SVM)*, sehingga kedua model memiliki kemampuan klasifikasi yang relatif setara. Namun, kemampuan model dalam mengklasifikasikan sentimen netral masih belum optimal. Kondisi tersebut menunjukkan bahwa pendekatan berbasis TF-IDF dan klasifikasi linear cenderung lebih efektif dalam mengenali pola kata yang eksplisit.

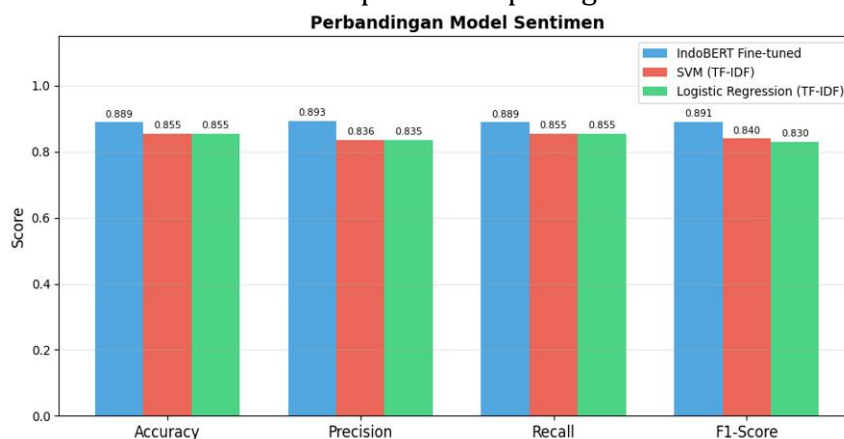
Sebagai model klasifikasi linear, *Logistic Regression* mengasumsikan hubungan yang sederhana antar fitur sehingga kurang mampu menangkap hubungan semantik yang kompleks pada data teks. Meskipun demikian, model ini tetap memberikan performa yang konsisten dengan proses pelatihan yang cepat serta implementasi yang sederhana, sehingga model ini masih relevan dipakai sebagai *baseline* dalam analisis sentimen.

3. Evaluasi Kinerja Model *IndoBERT*

Model *IndoBERT Fine-tuned* menunjukkan performa terbaik dibandingkan model lainnya dengan *accuracy* sebesar 0,889, *precision* 0,893, *recall* 0,889, dan *F1-score* 0,891.

Temuan penelitian menunjukkan bahwa penggunaan berbasis *transformer* mampu meningkatkan kualitas klasifikasi sentimen dibandingkan metode *machine learning* klasik. Peningkatan performa terlihat di setiap metrik evaluasi, serta kemampuan model yang lebih baik dalam mengklasifikasikan kelas netral yang menjadi kelemahan SVM dan *Logistic Regression*.

Penentuan model yang paling optimal dilakukan melalui perbandingan nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang diperoleh oleh setiap model. Ringkasan hasil evaluasi tersebut dapat dilihat pada gambar berikut.



Gambar 3.1 Perbandingan Performa Ketiga Model

Gambar 3.1 memperlihatkan bahwa *IndoBERT Fine-tuned* memperoleh performa terbaik pada seluruh metrik evaluasi dengan *accuracy* 0,889, *precision* 0,893, *recall* 0,889, dan *F1-score* 0,891. Dibandingkan dengan SVM dan *Logistic*

Regression, *IndoBERT* menunjukkan peningkatan *accuracy* sekitar 3,4%. Hal ini disebabkan karena *IndoBERT* menggunakan kemampuan *self-attention* pada model berbasis *transformer*, memungkinkan analisis keterkaitan kata dalam kalimat secara menyeluruh. Berbeda dengan TF-IDF yang hanya merepresentasikan teks ke dalam bentuk numerik berdasarkan tingkat kepentingan kata, *IndoBERT* memiliki kemampuan untuk mempelajari konteks dan hubungan antar kata sehingga makna yang terkandung dalam kalimat dapat dipahami dengan lebih baik. Kemampuan tersebut membuat *IndoBERT* lebih unggul dalam mengidentifikasi sentimen pada ulasan yang menggunakan beragam variasi bahasa, kalimat ambigu, maupun penggunaan bahasa tidak baku yang umum ditemukan pada ulasan pengguna aplikasi.

Meskipun demikian, hasil evaluasi menunjukkan bahwa masih terdapat sejumlah ulasan yang sulit diklasifikasikan secara tepat oleh seluruh model. Kesalahan klasifikasi terutama terjadi pada ulasan yang menggunakan bahasa tidak baku, singkatan, atau memuat lebih dari satu opini dalam satu kalimat sehingga batas antar kelas sentimen menjadi kurang jelas. Selain itu, kelas sentimen netral masih menjadi tantangan karena karakteristik ulasannya cenderung bersifat informatif dan tidak memiliki kecenderungan sentimen yang kuat. Kondisi tersebut menyebabkan model berbasis TF-IDF, seperti SVM dan Logistic Regression, lebih sering mengalami kesalahan klasifikasi pada kelas netral. Sementara itu, *IndoBERT* mampu mengurangi kesalahan tersebut karena mekanisme *self-attention* memungkinkan model memahami hubungan antar kata berdasarkan konteks kalimat secara lebih menyeluruh.

Perolehan hasil dalam penelitian ini selaras dengan berbagai penelitian terdahulu yang menyatakan bahwa model *transformer* cenderung memberikan performa yang lebih tinggi daripada algoritma *machine learning* klasik ketika digunakan untuk mengklasifikasikan teks berbahasa Indonesia. Dominasi sentimen negatif pada dataset juga dipengaruhi oleh periode pengambilan data yang dilakukan pada awal peluncuran aplikasi, sehingga banyak ulasan berisi keluhan terkait bug dan performa sistem.

Berdasarkan seluruh hasil evaluasi yang telah dilakukan, *IndoBERT Fine-tuned* menunjukkan performa terbaik dibandingkan model lainnya dalam mengklasifikasikan sentimen ulasan aplikasi *mobile banking* berbahasa Indonesia dan berpotensi menjadi pendekatan yang lebih akurat dalam mendukung evaluasi layanan digital.

KESIMPULAN

Penelitian ini berhasil melakukan evaluasi dan perbandingan terhadap kinerja algoritma klasifikasi yang terdiri atas Support Vector Machine (SVM), Logistic Regression, dan *IndoBERT Fine-tuned* pada umpan balik (feedback) pengguna aplikasi WONDR by BNI. Berdasarkan hasil pengujian, *IndoBERT Fine-tuned* menunjukkan performa terbaik dengan nilai *accuracy* sebesar 0,889, *precision* sebesar 0,893, *recall* sebesar 0,889, dan *F1-score* sebesar 0,891. Sementara itu, SVM dan Logistic Regression menghasilkan performa yang relatif setara dengan nilai *accuracy* sebesar 0,855. Pada konfigurasi dan representasi fitur

yang digunakan dalam penelitian ini, kedua model tersebut masih memiliki keterbatasan dalam memahami konteks dan makna kalimat secara menyeluruh karena mengandalkan representasi fitur berbasis TF-IDF.

Temuan penelitian menunjukkan bahwa pendekatan berbasis Transformer lebih mampu menangkap karakteristik sentimen pada teks berbahasa Indonesia dibandingkan metode *machine learning* konvensional. Kemampuan model dalam memahami konteks dan keterkaitan antar kata berkontribusi terhadap peningkatan performa klasifikasi yang diperoleh. Bagi pihak industri perbankan, hasil analisis sentimen ini dapat dimanfaatkan sebagai bahan evaluasi untuk mengidentifikasi permasalahan yang paling banyak dikeluhkan pengguna, menentukan prioritas perbaikan fitur maupun stabilitas aplikasi, serta meningkatkan kualitas layanan digital berdasarkan masukan pengguna. Dengan demikian, hasil penelitian ini diharapkan dapat menjadi referensi dalam pemilihan metode analisis sentimen, khususnya pada ulasan aplikasi *mobile banking*. Untuk penelitian selanjutnya, disarankan penggunaan sumber data yang lebih beragam serta pengujian terhadap berbagai model berbasis *Transformer* lainnya guna memperoleh pemahaman yang lebih komprehensif dan meningkatkan kualitas hasil klasifikasi sentimen.

DAFTAR PUSTAKA

- [1] B. Setiawan, "A Review of Sentiment Analysis Applications in Indonesia Between 2023–2024," *Journal of Information Engineering and Educational Technology*, vol. 8, no. 2, pp. 71–83, 2024.
- [2] B. Satya, M. S. Hasan, M. Rahardi, and F. F. Abdulloh, "Sentiment Analysis of Review Sestyc Using Support Vector Machine, Naive Bayes, and Logistic Regression Algorithm," in *Proceedings of the 2022 International Conference on Information and Communications Technology (ICOIACT)*, 2022.
- [3] M. I. Abidin and E. W. Pamungkas, "Analisis Sentimen Terhadap Timnas Indonesia di Piala Asia 2023 dengan Model Transformer Berbahasa Indonesia," *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, 2025.
- [4] V. E. Sidauruk and W. Herowati, "IndoBERT-Based Sentiment Analysis of Political Discourse on Platform X: The Case of Prabowo-Gibran Administration," *Journal of Applied Informatics and Computing*, vol. 10, no. 1, 2026.
- [5] H. Imaduddin, F. Y. A'la, and Y. S. Nugroho, "Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, 2023.
- [6] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," *arXiv:2011.00677*, 2020.
- [7] B. Wilie et al., "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," *arXiv:2009.05387*, 2020.
- [8] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 9, pp. 4305–4313, 2022.

- [9] I. H. Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 8, no. 3, pp. 302–307, 2023.
- [10] W. A. Hidayat dan V. R. S. Nastiti, "Perbandingan Kinerja Pre-Trained Indobert-Base dan Indobert-Lite pada Klasifikasi Sentimen Ulasan TikTok Tokopedia Seller Center," *JSII (Jurnal Sistem Informasi)*, vol. 11, no. 2, pp. 13–20, 2024.
- [11] U. Khairani, V. Mutiawani, dan H. Ahmadian, "Pengaruh Tahapan Preprocessing Terhadap Model IndoBERT dan IndoBERTtweet untuk Mendeteksi Emosi pada Komentar Akun Berita Instagram," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 11, no. 4, pp. 887-894, 2024.
- [12] M. V. S. Handayani dan Muljono, "Arsitektur Hibrida IndoBERTtweet-CNN untuk Klasifikasi Ujaran Kebencian Berbahasa Gaul di Media Sosial," *Infotekmesin*, vol. 17, no. 1, pp. 39-47, 2026.
- [13] L. B. Wijaya, Y. N. Wicaksana, S. S. Widhiasari, dan A. Saptawijaya, "Pengembangan Model Deteksi Hoaks Berbahasa Indonesia Menggunakan Kombinasi IndoBERT dan BiLSTM," *Buletin Pagelaran Mahasiswa Nasional Bidang TIK*, vol. 2, no. 1, pp. 12-16, 2024.
- [14] N. Fadilah, B. A. Putra, dan M. I. Pratama, "Optimasi Model BiLSTM Berbasis FastText pada Data Augmentasi Semantik IndoBERT untuk Klasifikasi Teks Bahasa Indonesia," *PISCES*, vol. 4, no. 1, 2026.
- [15] R. Santosa, A. B. Nusantara, dan S. Imron, "Comparative Analysis of SVM and IndoBERT for Intent Classification in Indonesian Overtime Chatbots," *JSCE (Journal of System and Computer Engineering)*, vol. 6, no. 3, pp. 258-270, 2025.
- [16] R. Jullfikar, R. Megasari, dan R. A. Sukamto, "Pengembangan Chatbot Menggunakan BERT untuk Pengelolaan Komunikasi Beasiswa Ikatan Alumni Universitas Pendidikan Indonesia," *Digital Transformation Technology (Digitech)*, vol. 5, no. 1, pp. 391-400, 2025.
- [17] M. S. Putra, M. M. H. Pandiangan, dan M. Farid, "Tinjauan Sistematis Chatbot Multibahasa: Arsitektur, Penerapan, dan Tantangan Linguistik pada Bahasa Sumber Daya Rendah," *Edusola: Journal Education, Sociology and Law*, vol. 1, no. 4, pp. 1560-1583, 2025.
- [18] D. F. Sjoraida, B. W. K. Guna, dan D. Yudhokusuma, "Analisis Sentimen Film Dirty Vote Menggunakan BERT (Bidirectional Encoder Representations from Transformers)," *Jurnal JTIK*, vol. 8, no. 2, pp. 393-404, 2024.
- [19] A. S. Rizky dan E. Y. Hidayat, "Klasifikasi Emosi pada Teks Bahasa Indonesia Menggunakan IndoBERT," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 4, pp. 2963-2970, 2025.
- [20] K. R. R. Wardani, I. Martina, dan J. F. X. Wern, "Analisis Pengaruh Karakteristik Masukan Teks terhadap Kinerja MiniLM v2-L6-H384 dan BERT-Base-Uncased pada Quora Question Pairs," *Jurnal Telematika*, vol. 20, no. 2, 2025.