OPEN ACCESS

Volume 9 (2), April-June 2025, 495-505

E-ISSN:2580-1643

Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)

DOI: https://doi.org/10.35870/jtik.v9i2.3340

Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google PlayStore Menggunakan *IndoBERT*

Tarwoto [1], Rizki Nugroho [2], Najmul Azka [3*], Wakhid Sayudha Rendra Graha [4]

[1,2,3,4] Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Kabupaten Banyumas, Provinsi Jawa Tengah, Indonesia.*

a r t i c l e i n f o

Article history:
Received 10 November 2024
Received in revised form
20 November 2024
Accepted 30 December 2024
Available online April 2025.

Keywords:
Sentiment Analysis;
IndoBERT; Machine Learning;
Sentiment Classification.

Kata Kunci:
Analisis Sentimen; IndoBERT;
Machine Learning; Klasifikasi
Sentimen.

a b s t r a c t

This research analyzes the sentiment of JKN mobile app reviews on Google PlayStore using the IndoBERT model, a deep learning-based language model designed for Indonesian text. The research process involved review data collection, text pre-processing, and sentiment classification into three categories: positive, negative, and neutral. The results show that the model performs very well, with an average accuracy of 97.28% and best metrics of 98.27% on accuracy, precision, recall, and F1 score. The specific contribution of this research is the development of a deep learning-based approach for sentiment analysis of Indonesian texts, particularly in the health sector through mobile applications. The findings not only provide insight into user perceptions of the JKN app, but also provide a basis for feature improvements and service enhancements. The implications of this research can support developers in designing strategies to improve the quality of digital-based health services in Indonesia.

a b s t r a k

Penelitian ini menganalisis sentimen ulasan aplikasi Mobile JKN di Google PlayStore menggunakan model IndoBERT, sebuah model bahasa berbasis deep learning yang dirancang untuk teks bahasa Indonesia. Proses penelitian melibatkan pengumpulan data ulasan, pra-pemrosesan teks, dan klasifikasi sentimen ke dalam tiga kategori: positif, negatif, dan netral. Hasilnya menunjukkan bahwa model memiliki kinerja sangat baik, dengan akurasi rata-rata 97,28% dan metrik terbaik sebesar 98,27% pada akurasi, precision, recall, dan F1 score. Kontribusi spesifik penelitian ini adalah pengembangan pendekatan berbasis deep learning untuk analisis sentimen teks Indonesia, khususnya di sektor kesehatan melalui aplikasi mobile. Temuan ini tidak hanya memberikan wawasan mengenai persepsi pengguna terhadap aplikasi JKN, tetapi juga menyediakan dasar untuk perbaikan fitur dan peningkatan layanan. Implikasi penelitian ini dapat mendukung pengembang dalam merancang strategi peningkatan kualitas layanan kesehatan berbasis digital di Indonesia.

ACM Computing Classification System (CCS)
EBSCOhost
Communication and Mass Media Complete (CMMC)

Corresponding Author. Email: najmulazka225@gmail.com [3].

1. Pendahuluan

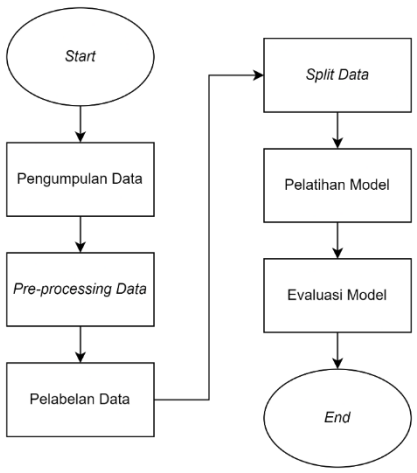
Dalam era digital yang terus berkembang, aplikasi mobile telah menjadi komponen krusial dalam berbagai aspek kehidupan, termasuk sektor kesehatan (Hadiwijaya *et al*., 2020). BPJS Kesehatan meluncurkan aplikasi Mobile JKN pada tahun 2017. Aplikasi ini dirancang untuk mempermudah peserta dalam melakukan berbagai kegiatan, mulai dari pemeriksaan mandiri terkait COVID-19, pendaftaran antrian secara daring, hingga konsultasi dengan tenaga medis melalui platform digital. Berbagai inovasi telah diperkenalkan oleh tim BPJS Kesehatan untuk meningkatkan aplikasi Mobile JKN (Tarwoto *et al*., 2019). Diharapkan dengan adanya aplikasi ini, masyarakat dapat lebih mudah mengakses informasi dan layanan kesehatan yang mereka perlukan. Meskipun aplikasi Mobile JKN dirancang untuk memberikan kemudahan, tantangan tetap ada (Sugianto *et al*., 2021). Berbagai ulasan dan tanggapan dari pengguna di platform seperti *Google Play Store* mencerminkan pengalaman mereka, yang sering kali berisi kritik dan masukan mengenai fitur, kinerja, serta kepuasan secara keseluruhan. Beberapa pengguna melaporkan kesulitan dalam navigasi aplikasi, adanya bug yang mengganggu, atau kurangnya informasi yang jelas mengenai layanan kesehatan yang tersedia. Hal ini menunjukkan adanya kebutuhan untuk memahami sentimen pengguna terhadap aplikasi ini guna mengidentifikasi area yang perlu diperbaiki (Baihaqi & Munandar, 2020).

Analisis sentimen terhadap ulasan pengguna dapat menjadi solusi untuk permasalahan ini (Hadiwijaya *et al*., 2019). Penelitian yang dilakukan oleh M. A. Hadiwijaya *et al*. (2019) menunjukkan bahwa model *IndoBERT* memiliki kinerja lebih baik dibandingkan dengan *Logistic Regression* dan *Support Vector Classification* (SVC), dengan akurasi mencapai 91%. Selain itu, penelitian lain (Gunawan & Kurniawan, 2020) mengungkapkan bahwa penggabungan model *IndoBERT* dengan *RCNN* menghasilkan akurasi tertinggi, yakni 95,16%. Sementara itu, penelitian oleh W. M. Baihaqi dan A. Munandar (2018) menyimpulkan bahwa model *IndoBERT* lebih unggul dibandingkan model *Naïve Bayes* dengan akurasi sebesar 85%. Berdasarkan tinjauan literatur, model *IndoBERT* menunjukkan keunggulan dalam analisis sentimen ulasan aplikasi Mobile JKN di *Google Play Store*. Dengan memanfaatkan teknologi pemrosesan bahasa alami, khususnya model *IndoBERT* yang dirancang untuk bahasa Indonesia, analisis ini dapat membantu dalam mengklasifikasikan ulasan menjadi kategori sentimen positif, negatif, dan netral (Widodo & Rahmat, 2020). Model *IndoBERT* memiliki kemampuan untuk memahami bahasa Indonesia dengan baik, yang memungkinkan untuk menghasilkan hasil analisis yang lebih akurat (Hadiwijaya *et al*., 2019). Dengan demikian, pengembang aplikasi dapat memperoleh pemahaman lebih mendalam mengenai persepsi pengguna serta mengidentifikasi masalah yang sering dihadapi dan aspek yang dihargai oleh pengguna. Hasil dari analisis sentimen ini diharapkan dapat memberikan gambaran yang jelas mengenai kepuasan pengguna terhadap aplikasi Mobile JKN. Dengan mengetahui sentimen yang dominan, pengembang dapat fokus pada perbaikan fitur yang dianggap kurang memadai oleh pengguna (Baihaqi & Munandar, 2020). Misalnya, jika banyak pengguna memberikan ulasan negatif terkait dengan kesulitan navigasi, pengembang dapat mempertimbangkan untuk merombak antarmuka pengguna agar lebih intuitif. Selain itu, analisis ini juga dapat membantu merancang strategi komunikasi yang lebih efektif untuk menjawab pertanyaan dan kekhawatiran pengguna.

Penelitian yang dilakukan oleh Widodo *et al.* (2020) menerapkan metode *BERT* pada ulasan aplikasi *Ruang Guru* untuk menganalisis sentimen pengguna. Hasil penelitian ini menunjukkan mayoritas pengguna *Ruang Guru* memberikan ulasan positif. Dari 5437 data yang diuji, 5254 ulasan dinyatakan positif, 16 netral, dan 167 negatif. Nilai *F1 Score* yang diperoleh dari presisi dan recall adalah 98,9%, sementara akurasi mencapai 99%. Hal ini menunjukkan bahwa metode pre-trained *BERT* sangat efektif untuk diterapkan dalam analisis. Nilai *F1 Score* yang dihasilkan dari presisi dan recall adalah 98,9%, sedangkan akurasi mencapai 99%. Ini menunjukkan bahwa metode pre-trained *BERT* sangat efektif untuk diterapkan dalam analisis. Selanjutnya, penelitian oleh Sugianto *et al.* (2021) menggunakan metode *BERT* untuk menganalisis sentimen terhadap pelayanan kesehatan berdasarkan ulasan *Google Maps*. Model *IndoBERT* menunjukkan performa yang baik dalam klasifikasi teks berbahasa Indonesia dan dapat diterapkan pada teks lainnya dalam bahasa Indonesia. Meskipun demikian, penting

untuk memperhatikan pengumpulan dataset untuk meningkatkan hasil yang diperoleh dalam setiap kelas. Penelitian ini juga menunjukkan bahwa penyedia layanan kesehatan telah mengalami kemajuan, namun masih ada masalah yang perlu dievaluasi kembali, yang menjadi perhatian utama masyarakat terkait kualitas pelayanan kesehatan. Penelitian ini tidak hanya bertujuan untuk meningkatkan kualitas aplikasi Mobile JKN, tetapi juga memberikan kontribusi pada pengembangan layanan kesehatan di Indonesia secara keseluruhan. Dengan memahami kebutuhan dan harapan pengguna, aplikasi ini dapat beradaptasi dan berkembang sesuai dengan perkembangan zaman. Oleh karena itu, penelitian ini berfokus pada solusi permasalahan yang dihadapi pengguna, mendukung tujuan utama aplikasi dalam meningkatkan akses serta kualitas layanan kesehatan di Indonesia. Dengan pendekatan berbasis data, diharapkan aplikasi Mobile JKN dapat menjadi alat yang lebih efektif dalam mendukung kesehatan masyarakat dan meningkatkan kepuasan pengguna.

2. Metodologi Penelitian



Gambar 1. Alur Penelitian

Pengumpulan Data

Metode pengumpulan data penelitian ini melibatkan proses pengunduhan ulasan pengguna aplikasi Mobile JKN yang terdapat di Google PlayStore. Tahapan awal dimulai dengan mencari aplikasi JKN pada platform, selanjutnya mengekstraksi data ulasan yang fokus pada konten teks. Koleksi data mencakup ragam ulasan yang terdiri dari tahun 2016 sampai 2024 dengan range rating 1-5 serta kategori positif, negatif, dan netral, yang bertujuan untuk memfasilitasi analisis sentimen secara menyeluruh dan mendalam.

Pre-processing Data

Tahap *pre-processing* data merupakan langkah kritis dalam mempersiapkan data ulasan untuk analisis sentimen menggunakan *IndoBERT*. Proses ini diawali dengan pembersihan data, yang meliputi penghapusan karakter khusus, tanda baca berlebihan, dan konten yang tidak relevan. Selanjutnya, dilakukan normalisasi teks dengan mengonversi seluruh teks menjadi huruf kecil (lowercase) untuk menstandarkan format. Proses tokenisasi akan dilakukan untuk mengurangi redundansi pada saat memecah teks ulasan menjadi unit-unit kata yang dapat diproses oleh model. Tahap pembersihan juga mencakup penghapusan kata-kata yang tidak bermakna (stopwords) dalam bahasa Indonesia, serta koreksi ejaan dan singkatan umum yang sering digunakan dalam ulasan daring. Selain itu, akan dilakukan proses stemming untuk mengubah kata ke dalam bentuk dasarnya, sehingga meningkatkan konsistensi dan akurasi analisis. Metode *pre-processing* ini bertujuan untuk menghasilkan dataset yang bersih, terstruktur, dan siap untuk dianalisis menggunakan model *IndoBERT*.

Pelabelan Data

Proses pelabelan data merupakan tahap krusial dalam mempersiapkan dataset untuk analisis sentimen menggunakan *IndoBERT*. Pelabelan dilakukan secara otomatis untuk meminimalkan kesalahan terjadi pada saat mengklasifikasikan setiap ulasan ke dalam tiga kategori sentimen: positif, negatif, dan netral. Setiap ulasan dinilai berdasarkan muatan emosional, intensitas ekspresi, dan konteks pernyataan pengguna. Tahap validasi akan dilakukan untuk memastikan konsistensi dan akurasi pelabelan, dengan melakukan penyesuaian jika terdapat ketidaksesuaian pelabelan otomatis.

Split Data

Splitting data dilaksanakan menggunakan pendekatan stratified random sampling untuk membagi dataset ulasan aplikasi JKN ke dalam tiga subset: data latih, validasi, dan pengujian. Proporsi pembagian dataset ditetapkan 70% untuk pelatihan, 15% untuk validasi,

dan 15% untuk pengujian. Metode ini dirancang untuk menjamin distribusi yang seimbang antarkelas sentimen, mencegah potensi bias, dan memastikan model *IndoBERT* dapat dilatih serta dievaluasi secara komprehensif. Teknik splitting akan mempertimbangkan representasi proporsional dari setiap kategori sentimen, sehingga model mampu belajar dan melakukan generalisasi dengan akurat.

Pelatihan Model

Pelatihan model *IndoBERT* untuk analisis sentimen ulasan aplikasi JKN dilakukan melalui serangkaian tahapan kompleks yang melibatkan fine-tuning arsitektur deep learning. Proses dimulai dengan inisialisasi model pra-latih *IndoBERT* yang telah dikembangkan dengan basis data teks bahasa Indonesia. Arsitektur model akan disesuaikan dengan karakteristik dataset ulasan, dengan menambahkan lapisan klasifikasi khusus untuk menghasilkan prediksi sentimen. *Hyperparameter* yang akan dioptimasi mencakup learning rate, batch size, jumlah *epoch*, dan *dropout rate*. Fungsi kerugian categorical cross-entropy akan digunakan untuk mengukur kesalahan prediksi, sementara algoritma optimizer Adam diterapkan untuk melakukan penyesuaian bobot jaringan. Proses pelatihan dilakukan secara iteratif dengan melakukan evaluasi performa model pada dataset validasi untuk mencegah *overfitting*. Metrik evaluasi yang digunakan meliputi akurasi, presisi, recall, dan F1-score untuk memberikan penilaian komprehensif terhadap kinerja model dalam mengklasifikasikan sentimen ulasan. Teknik regularisasi seperti early stopping dan learning *rate scheduling* akan diimplementasikan guna meningkatkan generalisasi model.

Evaluasi Model

Evaluasi model dilaksanakan melalui pendekatan komprehensif yang mencakup berbagai metrik performansi klasifikasi sentimen. Pengujian akan menggunakan dataset terpisah untuk menilai kemampuan generalisasi model mdhugol/indonesia-bert-sentiment-classification, dengan fokus pada akurasi, presisi, recall, dan F1-score. Metode *cross-validation* akan diterapkan untuk memastikan konsistensi performa model pada berbagai subset data. Analisis mendalam akan dilakukan melalui *confusion matrix* dan kurva ROC, memberikan wawasan detail tentang kemampuan model dalam mengklasifikasikan sentimen ulasan aplikasi JKN. Perbandingan dengan model klasifikasi tradisional akan memberikan konsep empiris terhadap keunggulan pendekatan *deep learning* berbasis *IndoBERT*.

3. Hasil dan Pembahasan

Hasil

Pengumpulan Data

Data ulasan pada aplikasi Mobile JKN dikumpulkan dari ulasan di google playstore dengan data yang diurutkan dari yang terbaru. Untuk mengambil data ulasan ini menggunakan library yang dibuat khusus untuk mengambil pada google playstore serta *library pandas* untuk membaca data sebagai data frame.

Tabel 1. Data Ulasan

No	Ulasan
1	buruk
2	Mudah dan mantab 👍
3	Cepat klo daftar berobat
...	...
5404	GK ribet mantap

Data mentah yang telah diambil sebanyak 5404 ulasan. Data ini yang nantinya akan dilakukan proses *pre-processing*.

Pre-processing Data

Sebelum memproses dataset, langkah *preprocessing* data dilakukan agar klasifikasi dapat dilakukan dengan lebih mudah. Preprocessing adalah tahap krusial untuk membersihkan data mentah, dengan tujuan meningkatkan akurasi dan efisiensi model. Berikut adalah langkah-langkah yang terlibat dalam proses *preprocessing*. *Data cleaning* adalah langkah untuk mengurangi gangguan dengan menghilangkan elemen-elemen teks yang tidak diperlukan.

Tabel 2. Data *Cleaning*

No	Ulasan	*Pre-processing*
1	buruk	buruk
2	Mudah dan mantab 👍	Mudah dan mantab
3	Cepat klo daftar berobat	cepat klo daftar berobat

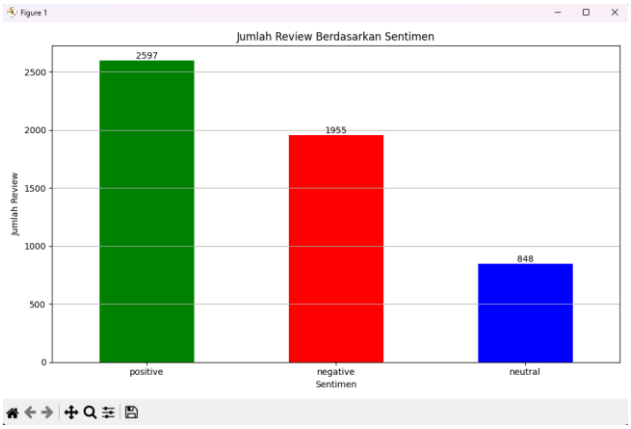
Penghapusan *stopword* adalah proses yang dilakukan untuk menghilangkan kata-kata yang sering muncul namun tidak memberikan kontribusi signifikan terhadap makna atau biasanya kata penghubung.

Tabel 3. *Stopword*

No	Ulasan	*Stopword*
1	buruk	buruk
2	mudah dan mantab	mudah mantab
3	cepat klo daftar berobat	cepat daftar berobat

Pelabelan Data

Proses pelabelan adalah tahap di mana label atau klasifikasi diberikan pada data yang akan dianalisis, agar model dapat lebih mudah dalam mengklasifikasikan sentimen. Dalam penelitian ini, pelabelan data dilakukan secara otomatis untuk menetapkan label pada setiap ulasan dalam dataset, yang dikategorikan sebagai sentimen positif, netral, atau negatif.



Gambar 2. Hasil Label

Pada gambar 2 tertera sebuah 3 sentimen positif, negatif dan netral yang dihasilkan dari proses labeling. Untuk sentimen positif sebanyak 2597 ulasan, sentimen negatif sebanyak 1955 ulasan dan sentimen netral 848 ulasan.

Split Data

Berikut tabel data yang akan di latih dan diuji dengan proporsi 70:15:15 dimana 70% data yang akan dilatih, 15% data yang akan di uji dan 15 data yang akan digunakan untuk validasi.

Tabel 4. Hasil Split Data

No.	Sentimen	Tipe Data	Jumlah
1.	Positif	latih	1853
		uji	372
		validasi	372
2.	Negatif	latih	1346
		uji	304
		validasi	305
3.	Netral	latih	581
		uji	134
		validasi	133

Tabel 4 menggambarkan hasil pembagian data berdasarkan jenis sentimen, yaitu positif, negatif, dan netral, yang telah disiapkan untuk proses pelatihan, pengujian, dan validasi model analisis sentimen. Untuk kategori sentimen positif, terdapat 1.853 contoh data yang digunakan untuk pelatihan, 372 untuk pengujian, dan 372 untuk validasi. Sedangkan untuk sentimen negatif, model dilatih dengan 1.346 contoh data, diuji dengan 304 contoh, dan divalidasi dengan 305 contoh. Di sisi lain, untuk sentimen netral, terdapat 581 contoh data untuk pelatihan, 134 untuk pengujian, dan 133 untuk validasi. Pembagian ini bertujuan agar model dapat dengan efektif mengenali dan mengklasifikasikan setiap jenis sentimen, sehingga mampu menghasilkan analisis yang tepat dan dapat dipercaya. Dengan distribusi data yang seimbang, diharapkan model dapat belajar dari berbagai pola yang terdapat dalam setiap kategori sentimen dan dapat menggeneralisasi dengan baik ketika dihadapkan pada data baru.

Pelatihan Model *(Modeling)*

Pelatihan model, atau *modeling*, merupakan tahap krusial dalam pengembangan sistem analisis sentimen. Pada fase ini, data yang telah dibagi menjadi set pelatihan, pengujian, dan validasi digunakan untuk mengajarkan model bagaimana mengenali dan mengklasifikasikan sentimen berdasarkan pola yang ada. Dalam pelatihan ini, dilakukan selama 12 epoch dengan ukuran batch sebesar 64, yang memungkinkan model untuk memproses data dalam kelompok yang lebih kecil dan meningkatkan efisiensi pembelajaran. Algoritma yang digunakan dioptimalkan untuk mencapai model terbaik berdasarkan metrik f1, yang memberikan keseimbangan antara presisi dan *recall*. Selain itu, learning rate yang diterapkan adalah 6e-6, yang dirancang untuk memastikan bahwa model

belajar dengan stabil tanpa melewatkan titik minimum yang optimal. Namun, pelatihan model dihentikan pada epoch ke-10 karena penerapan early stopping dengan nilai patience sebesar 10, yang berarti model tidak menunjukkan peningkatan performa selama tiga epoch berturut-turut. Hal ini bertujuan untuk mencegah overfitting dan memastikan bahwa model yang dihasilkan tetap generalisasi dengan baik saat dihadapkan pada data baru. Setelah proses pelatihan selesai, model kemudian diuji menggunakan data pengujian untuk mengevaluasi kinerjanya dan memastikan bahwa ia dapat mengenali sentimen dengan akurasi yang tinggi. Dengan demikian, pelatihan model yang efektif sangat penting untuk menghasilkan sistem analisis sentimen yang handal dan akurat.

Tabel 5. Hasil Evaluasi Model

Epoch	*Loss*	*Accuracy*	*Precision*	*Recall*	*F1*
1	7,36%	98,02%	98,06%	98,02%	98,02%
2	6,59%	97,53%	97,55%	97,53%	97,50%
3	5,70%	97,41%	97,43%	97,41%	97,37%
4	5,11%	98,40%	98,40%	98,40%	98,39%
5	5,75%	97,53%	97,54%	97,53%	97,50%
6	9,59%	96,54%	96,73%	96,54%	96,49%
7	8,32%	97,16%	97,19%	97,16%	97,12%
8	9,66%	96,42%	96,44%	96,42%	96,36%
9	12,28%	96,54%	96,61%	96,54%	96,52%
10	8,45%	97,28%	97,32%	97,28%	97,28%
AVG	7,88%	97,28%	97,33%	97,28%	97,26%

Tabel 5 yang disajikan menunjukkan hasil evaluasi model *machine learning* atau *deep learning* selama sepuluh *epoch* pelatihan. Setiap *epoch* mencerminkan satu siklus penuh di mana seluruh dataset digunakan untuk memperbarui bobot model. Nilai loss, yang berkisar antara 5,11% hingga 12,28%, menunjukkan fluktuasi dalam performa model, dengan tren penurunan di awal epoch, namun mengalami peningkatan pada beberapa epoch, yang mungkin mengindikasikan overfitting. Akurasi model secara keseluruhan sangat tinggi, dengan nilai antara 96,54% hingga 98,40%, menunjukkan kemampuan model dalam mengklasifikasikan data dengan baik. Selain itu, *precision* dan *recall* juga menunjukkan performa yang solid, masing-masing berkisar antara 96,44% hingga 98,40%, yang menandakan bahwa model efektif dalam mengidentifikasi kelas positif tanpa banyak menghasilkan *false positives*. F1 score, yang merupakan rata-rata harmonis dari precision dan recall, berkisar antara 96,36% hingga 98,39%, menunjukkan keseimbangan yang baik antara kedua metrik tersebut. Epoch 4 mencatat performa terbaik dengan akurasi, precision, recall, dan F1 score tertinggi, sehingga menjadi titik fokus untuk evaluasi lebih lanjut. Hasil ini menunjukkan bahwa model memiliki performa yang sangat baik dan dapat diandalkan untuk aplikasi praktis, meskipun pengujian lebih lanjut pada dataset yang berbeda diperlukan untuk memastikan generalisasi model.

Tabel 6. Hasil Metrik Terbaik

Accuracy	*Precision*	*Recall*	*F1 Score*
98.27%	98.27%	98.27%	98.27%

Tabel 6 menunjukkan hasil metrik terbaik yang dicapai setelah evaluasi model selama sepuluh *epoch* pelatihan. Semua metrik akurasi, *precision*, *recall*, dan F1 score mencatat nilai yang sama, yaitu 98,27%. Angka ini mencerminkan performa optimal model dalam mengklasifikasikan data, di mana model tidak hanya mampu mencapai tingkat akurasi yang tinggi, tetapi juga menunjukkan kemampuan yang seimbang dalam mengidentifikasi kelas positif dengan baik. *Precision* yang tinggi menunjukkan bahwa sebagian besar prediksi positif yang dihasilkan oleh model adalah benar, sementara recall yang setara menandakan bahwa model berhasil mendeteksi sebagian besar contoh positif yang ada. F1 score yang juga mencapai

98,27% menegaskan keseimbangan yang baik antara precision dan recall, membuat model ini sangat efektif untuk aplikasi yang membutuhkan keakuratan tinggi dalam klasifikasi. Hasil-hasil ini menunjukkan bahwa model telah dilatih dengan baik dan dapat diandalkan untuk digunakan.

Tabel 7. Perbandingan Model Penelitian Sebelumnya

Peneliti	Metode	*Accuracy*	*Precision*	*Recall*	F1-score
Kami	*IndoBERT*	98.27%	98.27%	98.27%	98.27%
Maulida (2024)	SVM	89.53%	88.17%	45.96%	86%
Roiqoh (2023)	*Naive Bayes*	94.75%	64.32%	70.3%	64.55%

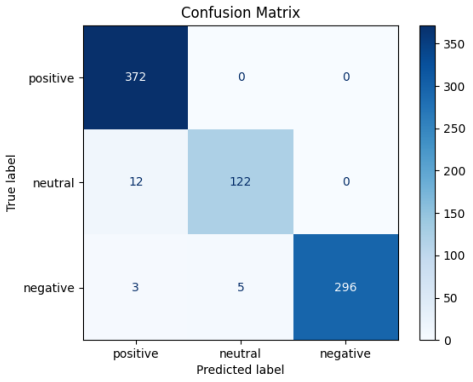
Pada Tabel 7 membandingkan performa model *IndoBERT* dengan metode *Naive Bayes* (Roiqoh, 2023) dan SVM (Maulida, 2024) berdasarkan metrik evaluasi utama: *Accuracy, Precision, Recall, dan F1-score*. *IndoBERT* menunjukkan hasil terbaik dengan nilai 98,27% di semua metrik, yang menegaskan kemampuannya dalam memahami pola kompleks pada teks bahasa Indonesia untuk klasifikasi sentimen. Keunggulan ini menunjukkan bahwa arsitektur berbasis transformer seperti *IndoBERT* lebih efektif dalam memahami konteks dan makna teks dibandingkan dengan metode klasik.

Sementara itu, model SVM hanya mencapai *Accuracy* 89,53% dan *Precision* 88,17%, namun memiliki kelemahan pada *Recall* yang rendah, yaitu 45,96%. Rendahnya nilai *Recall* menunjukkan SVM kurang optimal dalam mendeteksi semua sampel relevan. Di sisi lain, *Naive Bayes* memiliki *Accuracy* 94,75% dan *Recall* 70,3%, lebih baik daripada SVM, namun *Precision*-nya hanya 64,32% dan F1-score-nya 64,55%, yang menunjukkan bahwa *Naive Bayes* masih kurang dalam menangkap kompleksitas data teks. Dari perbandingan ini, dapat disimpulkan bahwa *IndoBERT* unggul dalam semua metrik evaluasi dan lebih efektif dalam mengklasifikasikan sentimen dibandingkan dengan metode berbasis *machine learning* klasik seperti SVM dan *Naive Bayes*.

Tabel 8. Hasil Metrik *Training Loss*

Train Loss	*Train Step/second*	*Train Samples/second*
5,7%	0.6/s	42.659/s

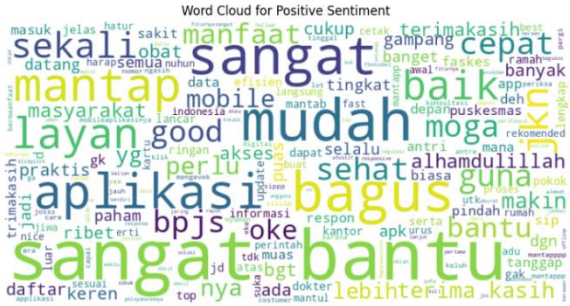
Pada Tabel 8, Train Loss sebesar 5,7% menunjukkan tingkat kesalahan model dalam memahami pola bahasa Indonesia untuk mengklasifikasikan sentimen ulasan aplikasi JKN. Nilai ini menunjukkan performa yang baik, meskipun masih memungkinkan optimasi lebih lanjut. Kecepatan pelatihan dengan Train Step 0,6 langkah per detik mencerminkan waktu yang dibutuhkan untuk menyelesaikan satu langkah pelatihan. Kecepatan ini dipengaruhi oleh kompleksitas model dan ukuran data. *Train Samples* sebesar 42.659 sampel per detik menunjukkan efisiensi tinggi dalam memproses data ulasan. Secara keseluruhan, *IndoBERT* menunjukkan performa baik dalam klasifikasi sentimen ulasan aplikasi JKN. Matriks konfusi menunjukkan kinerja model klasifikasi. Baris mewakili label sebenarnya, sementara kolom mewakili label yang diprediksi. Elemen diagonal menunjukkan contoh yang diklasifikasikan dengan benar, sedangkan elemen di luar diagonal menunjukkan contoh yang salah klasifikasi. Dalam hal ini, model telah mengklasifikasikan dengan benar 372 contoh positif, 122 contoh netral, dan 296 contoh negatif. Model salah mengklasifikasikan 12 contoh positif sebagai netral, 3 contoh positif sebagai negatif, dan 5 contoh netral sebagai negatif.



Gambar 3. *Confusion Matrix*

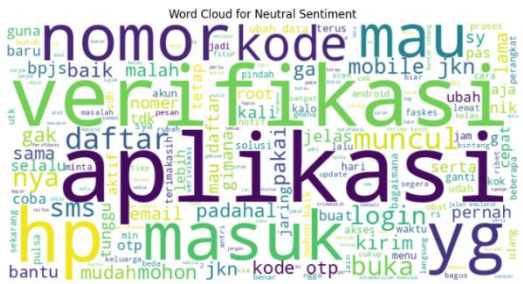
Gambar *word cloud* positif di bawah ini menampilkan kata-kata yang paling sering muncul dalam ulasan dan komentar dengan sentimen positif. Kata-kata besar

dalam visualisasi ini mencerminkan elemen-elemen yang paling dihargai oleh responden, seperti "bagus," "mudah," dan "baik." Dengan menganalisis kata-kata ini, kita dapat memahami aspek-aspek yang membuat pengalaman pengguna menjadi positif dan menarik, serta faktor-faktor yang berkontribusi pada kepuasan mereka.



Gambar 4. *Word Cloud* Sentimen Positif

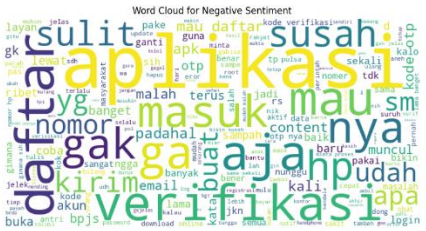
Word cloud netral yang ditampilkan di bawah ini menggambarkan kata-kata yang muncul dalam ulasan dan komentar dengan sentimen netral. Dalam visualisasi ini, kata-kata seperti "masuk," "verifikasi," dan "aplikasi" terlihat menonjol, menunjukkan bahwa responden memiliki pandangan yang lebih seimbang atau tidak terlalu kuat terhadap pengalaman mereka. Analisis kata-kata ini memberikan wawasan tentang elemen-elemen yang mungkin tidak mencolok tetapi tetap penting dalam membentuk persepsi keseluruhan, serta membantu kita memahami area yang mungkin memerlukan perhatian lebih lanjut.



Gambar 5. *Word Cloud* Sentimen Netral

Gambar *word cloud* negatif di bawah ini menunjukkan kata-kata yang paling sering muncul dalam ulasan dan komentar dengan sentimen negatif. Kata-kata besar seperti "susah," "sulit," dan "gak" mencerminkan pengalaman yang kurang memuaskan dan menunjukkan area di mana pengguna merasa tidak puas.

Dengan menganalisis kata-kata dalam word cloud ini, kita dapat mengidentifikasi isu-isu yang perlu diperbaiki dan memahami lebih dalam tentang keluhan yang sering diungkapkan oleh responden, sehingga dapat mengambil langkah yang tepat untuk meningkatkan kualitas dan kepuasan pengguna di masa mendatang.



Gambar 6. *Word Cloud* Sentimen Negatif

Tabel 9. Kata Yang Sering Muncul

Negative	aplikasi	857
	verifikasi	576
	daftar	535
	mau	480
	gak	404
	nya	399
	kode	395
	no	386
	ga	377
Neutral	masuk	371
	aplikasi	295
	daftar	232
	verifikasi	214
	kode	210
	masuk	181
	otp	176
	hp	165
	nomor	156
	mau	143
Positive	jkn	119
	sangat	864
	bantu	653
	mudah	444
	bagus	372
	mantap	249
	aplikasi	238
	baik	202
	layan	192
	good	146
	jkn	126

Berdasarkan analisis sentimen, pengembang dapat mengambil beberapa langkah untuk meningkatkan pengalaman pengguna. Untuk sentimen negatif, kata seperti "gak", "ga", "no", dan "masuk" menunjukkan masalah dalam pendaftaran atau login, sehingga pengembang perlu menyederhanakan proses ini dan meningkatkan stabilitas aplikasi. Kata "kode" dan "aplikasi" juga menunjukkan masalah teknis, yang memerlukan pengujian kode dan perbaikan bug. Pada sentimen positif, kata seperti "sangat", "bantu", "mudah", dan "bagus" menunjukkan kepuasan pengguna. Pengembang dapat memperkuat fitur yang ada dan menambah fitur baru, serta fokus pada layanan JKN yang mendapat respons baik. Untuk sentimen netral, kata seperti "otp", "hp", dan "nomor" menandakan kebingungan dalam proses verifikasi. Pengembang perlu memperjelas instruksi dan mempercepat proses verifikasi untuk mengurangi kebingungan dan meningkatkan kepuasan pengguna.

Salah satu keterbatasan penelitian ini adalah representativitas dataset yang digunakan, yang mungkin tidak mencakup seluruh kelompok pengguna aplikasi JKN. Jika sebagian besar ulasan berasal dari pengguna dengan pengalaman negatif, sentimen negatif bisa mendominasi hasil, sehingga kesimpulan yang dihasilkan kurang mencerminkan pengalaman semua pengguna. Oleh karena itu, penting menggunakan dataset yang lebih beragam. Selain itu, model *IndoBERT* mungkin kesulitan mengklasifikasikan sentimen pada ulasan yang ambigu atau campuran antara positif dan negatif, seperti ulasan yang mengandung keluhan dan pujian. Pengembangan lebih lanjut diperlukan agar model dapat mengenali sentimen yang lebih kompleks. Keterbatasan lain adalah kesulitan dalam membedakan sentimen netral dan positif. Ulasan yang tampak netral namun menyimpan ketidakpuasan bisa salah diidentifikasi sebagai positif. Model perlu diperbaiki agar dapat lebih akurat dalam memahami konteks dan nuansa sentimen.

Pembahasan

Hasil analisis sentimen terhadap ulasan pengguna aplikasi Mobile JKN menunjukkan bahwa model *IndoBERT* mampu memberikan hasil yang akurat dalam mengklasifikasikan sentimen ulasan pengguna ke dalam kategori positif, negatif, dan netral. Model ini menunjukkan kinerja yang sangat baik, terutama dalam menangani ulasan berbahasa Indonesia. Penelitian oleh Hendriyanto *et al.* (2022) menunjukkan bahwa teknik analisis sentimen menggunakan algoritma *Support Vector Machine* (SVM) juga berhasil diterapkan untuk ulasan aplikasi serupa, meskipun model *IndoBERT* memberikan akurasi yang lebih tinggi dalam konteks ulasan berbahasa Indonesia. Selain itu, penggabungan model *IndoBERT* dengan *RCNN* dalam penelitian oleh Jayadianti *et al.* (2022) meningkatkan akurasi lebih lanjut, menunjukkan bahwa fine-tuning model dengan teknik lain dapat menghasilkan performa yang lebih optimal dalam analisis sentimen aplikasi. Berdasarkan temuan ini, aplikasi Mobile JKN dapat lebih responsif dalam menghadapi masukan dari pengguna. Sebagai contoh, jika banyak pengguna memberikan ulasan negatif mengenai kesulitan dalam navigasi aplikasi atau bug yang mengganggu, pengembang dapat memprioritaskan perbaikan pada antarmuka pengguna atau stabilitas aplikasi. Khotimah (2022) dalam penelitiannya juga mengungkapkan pentingnya kualitas sistem dan kualitas layanan dalam aplikasi Mobile JKN, yang secara langsung mempengaruhi kepuasan peserta BPJS Kesehatan di wilayah JABODETABEK. Dengan memanfaatkan hasil analisis sentimen, pengembang dapat mengetahui aspek mana yang perlu diperbaiki untuk meningkatkan kualitas pengalaman pengguna.

Pentingnya analisis sentimen ini juga didukung oleh penelitian oleh Fahlevvi (2022), yang menemukan bahwa aplikasi dengan analisis sentimen yang tepat dapat memberikan gambaran yang lebih jelas mengenai kepuasan pengguna. Dalam hal ini, model *IndoBERT* memberikan pemahaman yang lebih baik mengenai perasaan pengguna terhadap fitur-fitur tertentu dalam aplikasi Mobile JKN. Sebagai contoh, jika pengguna banyak mengeluhkan kurangnya informasi yang jelas mengenai layanan kesehatan yang tersedia, hal ini menunjukkan bahwa pengembang perlu menyempurnakan fitur penyampaian informasi di dalam aplikasi agar lebih terstruktur dan mudah dipahami oleh pengguna. Selain itu, penelitian oleh Wijaya *et al.* (2023) menunjukkan bahwa penerapan model *IndoBERT* dalam analisis sentimen pada berbagai platform media sosial menunjukkan kemampuannya untuk menangani data teks berbahasa Indonesia dengan efektif. Hal ini memperkuat

keyakinan bahwa model yang sama dapat diimplementasikan untuk aplikasi seperti Mobile JKN untuk lebih memahami kebutuhan dan harapan pengguna terkait layanan kesehatan. Dengan demikian, pengembang aplikasi dapat menggunakan hasil analisis sentimen ini untuk merancang strategi komunikasi yang lebih efektif, serta merancang pembaruan aplikasi yang lebih sesuai dengan harapan pengguna. Dari perspektif perbaikan aplikasi, analisis sentimen ini dapat dijadikan dasar untuk merancang fitur-fitur baru yang lebih sesuai dengan kebutuhan masyarakat. Seperti yang dikemukakan oleh Baihaqi dan Munandar (2023), aplikasi yang dapat menyesuaikan diri dengan masukan pengguna akan meningkatkan kepuasan dan retensi pengguna. Misalnya, jika analisis sentimen menunjukkan bahwa pengguna lebih menghargai fitur konseling medis yang lebih mudah diakses, maka pengembang dapat lebih fokus pada pengembangan fitur tersebut.

Aplikasi Mobile JKN juga dapat berperan penting dalam memperbaiki kualitas layanan kesehatan di Indonesia, terutama dengan mengidentifikasi masalah yang sering dihadapi oleh pengguna dan mencari solusi berdasarkan data ulasan yang telah dikumpulkan. Penerapan analisis sentimen dengan model *IndoBERT* dapat memberikan pemahaman yang lebih mendalam mengenai persepsi pengguna terhadap aplikasi Mobile JKN. Hal ini membuka peluang untuk pengembangan aplikasi yang lebih responsif dan efektif dalam meningkatkan kualitas layanan kesehatan di Indonesia. Dengan berfokus pada perbaikan berdasarkan masukan pengguna yang terstruktur, aplikasi ini dapat lebih optimal dalam mendukung kesehatan masyarakat dan meningkatkan kepuasan pengguna, sebagaimana dijelaskan oleh Maulida *et al.* (2024) dalam penelitian mereka yang menggunakan *Support Vector Machine* berbasis *Particle Swarm Optimization*.

4. Kesimpulan

Penelitian ini berhasil menunjukkan efektivitas model *IndoBERT* dalam menganalisis sentimen ulasan aplikasi Mobile JKN di *Google PlayStore* dengan tingkat akurasi yang sangat tinggi. Model ini berhasil mengklasifikasikan sentimen ulasan pengguna dengan akurasi rata-rata 97,28% dan mencapai metrik terbaik sebesar 98,27% dalam hal akurasi, precision, recall, dan *F1 score*. Hasil ini menunjukkan potensi besar penggunaan model *IndoBERT* dalam analisis sentimen teks berbahasa Indonesia, khususnya dalam aplikasi kesehatan. Meskipun model ini menunjukkan hasil yang sangat baik, penelitian ini juga mengidentifikasi tantangan dalam memahami sentimen yang lebih kompleks atau bercampur. Oleh karena itu, penelitian lanjutan disarankan untuk mengembangkan model yang lebih canggih serta memperluas dataset yang digunakan untuk meningkatkan akurasi dan keterwakilan analisis sentimen. Temuan ini memberikan wawasan yang berharga bagi pengembang aplikasi kesehatan, terutama dalam meningkatkan fitur aplikasi dan pengalaman pengguna berdasarkan analisis sentimen yang lebih akurat. Penelitian ini juga membuka peluang untuk penelitian lebih lanjut dalam pengembangan model analisis sentimen yang lebih adaptif dan dapat diintegrasikan dengan sistem rekomendasi untuk aplikasi kesehatan digital.

5. Daftar Pustaka

Ardiansyah, A. S. W., Qodri, K. N., & Saputro, F. E. N. (2023). Nisrina Akbar Rizky P,"Analisis sentimen terhadap pelayanan Kesehatan berdasarkan ulasan Google Maps menggunakan BERT," J.

Baihaqi, W. M., & Munandar, A. (2023). Sentiment Analysis of Student Comment on the College Performance Evaluation Questionnaire Using Naïve Bayes and *IndoBERT*. *JUITA: Jurnal Informatika*, *11*(2), 213-220. https://doi.org/10.30595/juita.v11i2.17336.

Erfina, A., & Wardani, N. R. (2022). Analisis Sentimen Perguruan Tinggi Termewah Di Indonesia Menurut Ulasan Google Maps Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Manajemen Informatika dan Sistem Informasi*, *5*(1), 77-85. https://doi.org/10.36595/misi.v5i1.591.

Fahlevvi, M. R. (2022). Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi dan Dokumentasi Kementerian Dalam Negeri

Republik Indonesia di Google Playstore Menggunakan Metode Support Vector Machine. *Jurnal Teknologi dan Komunikasi Pemerintahan*, *4*(1), 1-13. https://doi.org/10.33701/jtkp.v4i1.2701.

Hadiwijaya, M. A., Pirdaus, F. P., Andrews, D., Achmad, S., & Sutoyo, R. (2023, November). Sentiment Analysis on Tokopedia Product Reviews using Natural Language Processing. In *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)* (pp. 380-386). IEEE. https://doi.org/10.1109/ICIMCIS60089.2023.10348996.

Hendriyanto, M. D., Ridha, A. A., & Enri, U. (2022). Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine. *INTECOMS: Journal of Information Technology and Computer Science*, *5*(1), 1-7. https://doi.org/10.31539/intecoms.v5i1.3708.

Jayadianti, H., Kaswidjanti, W., Utomo, A. T., Saifullah, S., Dwiyanto, F. A., & Drezewski, R. (2022). Sentiment analysis of Indonesian reviews using fine-tuning *IndoBERT* and R-CNN. *ILKOM Jurnal Ilmiah*, *14*(3), 348-354.

Khotimah, N. (2022). Pengaruh Kualitas Sistem, Kualitas Layanan, dan Kualitas Informasi pada Aplikasi Mobile JKN Terhadap Kepuasan Peserta BPJS Kesehatan Di Wilayah Jabodetabek. *Jurnal Akuntansi Dan Manajemen Bisnis*, *2*(2), 69-76.

Kusuma, R. M. R. W. P., & Yustanti, W. (2021). Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers). *Journal of Emerging Information System and Business Intelligence (JEISBI)*, *2*(3).

Limia Budiarti, R., & Adriana, W. (2019). Pemanfaatan Google Maps API dalam Pemetaan dan Pemberdayaan Pariwisata Desa Di Indonesia Berbasis Web-Mobile. *Indonesian Journal of Computer Science*, *8*(1), 55-65.

Maarif, M. M., & Setiyawati, N. (2024). Analisis Sentimen Review Aplikasi LinkedIn di Google Play Store Menggunakan Support Vector Machine. *Progresif: Jurnal Ilmiah Komputer*, *20*(1), 454-464. https://doi.org/10.35889/progresif.v20i1.1614.

Pribadi, T. I., Fahry, F., Muharis, M., & Marswandi, E. D. P. (2024). Analysis of Tourist Sentiment towards Tourist Attractions in the Mandalika Special Economic Zone Using the Naïve Bayes Method. *Jurnal Bumigora Information Technology (BITe)*, *6*(1), 105-114.

Raffi, M., Suharso, A., & Maulana, I. (2023). Analisis Sentimen Ulasan Aplikasi Binar Pada Google Play Store Menggunakan Algoritma NaÃ¯ve Bayes. *INTECOMS: Journal of Information Technology and Computer Science*, *6*(1), 450-462.

Roiqoh, S., Zaman, B., & Kartono, K. (2023). Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Mobile JKN dengan Lexicon Based dan Naïve Bayes. *Jurnal Media Informatika Budidarma*, *7*(3), 1582-1592.

Wardiah, R., Izhar, M. D., & Lanita, U. (2022). Studi Efektivitas Aplikasi Mobile Jkn Pada Peserta Jaminan Kesehatan Nasional (Jkn) Kota Jambi. *Jurnal Endurance*, *7*(3), 607-614. https://doi.org/10.22216/jen.v7i3.1664.

Wijaya, K. A., Romadhony, A., & Richasdy, D. (2023). Implementasi Model *IndoBERT* pada Dashboard Sentimen Media Sosial (Studi Kasus Universitas XYZ). *eProceedings of Engineering*, *10*(4).

Yuspita, E., & Suryono, R. R. (2024). Perbandingan Berbagai Metode Klasifikasi Teks Untuk Sentimen Kebijakan Makan Gratis Di Indonesia. *The Indonesian Journal of Computer Science*, *13*(5). https://doi.org/10.30656/jsii.v11i2.9168.