Evaluasi Persepsi Pengguna SATUSEHAT Mobile Melalui Analisis Sentimen dan Topik Modeling Berbasis Machine Learning dan Transformer

User Perception Evaluation of SATUSEHAT Mobile Using Sentiment Analysis and Topic Modeling with Machine Learning and Transformer Approaches

Rizal Wahyu Pratama[1], Khulika Malkan*[2], Khulika Malkan[3], Ardelia Rachma Laksita[4], Habibah Ratna Dadhila Islami Hana[5]

[1,2,3,4,5] Program Studi Sains Data, Fakultas Informatika, Universitas Telkom

e-mail: ***[2]khulika@student.telkomuniversity.ac.id**

Abstrak -*Platform SATUSEHAT resmi diluncurkan Kementerian Kesehatan RI pada 26 Juli 2022 sebagai sistem integrasi data rekam medis nasional, dan pada 1 Maret 2023 PeduliLindungi bertransformasi menjadi SATUSEHAT Mobile untuk digunakan masyarakat umum. Hingga saat ini angka unduhan aplikasi melampaui 50 juta dengan rating 3.4 dari 5 bintang di Google Play Store pada 6 Juni 2026. Penelitian ini mengevaluasi persepsi 16.950 pengguna melalui klasifikasi sentimen dan identifikasi topik keluhan. Enam model diuji secara komparatif, di mana tiga pendekatan berbasis TF-IDF dengan algoritma Complement Naive Bayes, Support Vector Machine, dan Logistic Regression, serta tiga variasi IndoBERT meliputi ekstraksi fitur dengan SVM, ekstraksi fitur dengan Logistic Regression berparameter C=0,01, dan fine-tuning penuh dengan classification head. Topic modeling menggunakan Latent Dirichlet Allocation dengan delapan topik yang ditentukan berdasarkan Coherence Score tertinggi 0,5487. IndoBERT Fine-tuned mencapai F1-Macro 0,8736 dan akurasi 0,9339, melampaui baseline. Keluhan dominan adalah kegagalan aplikasi setelah pembaruan aplikasi, yaitu sebesar 23,3 persen dari total ulasan negatif, disusul penilaian aplikasi yang menyusahkan sebesar 17,6 persen dan kegagalan pendaftaran akun sebesar 17,3 persen.*

Kata kunci – Analisis Sentimen, IndoBERT, IndoBERT Fine-tuned, SATUSEHAT Mobile, Transformasi Kesehatan.

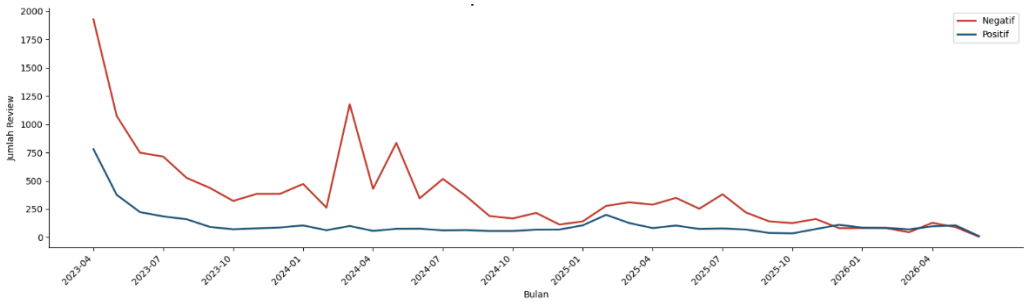
Abstract *-The SATUSEHAT platform was officially launched by Ministry of Health of the Republic of Indonesia on July 26, 2022, as a national health record data integration system, and on March 1, 2023, PeduliLindungi was formally transformed into SATUSEHAT Mobile for public use. The application has surpassed 50 million downloads with a rating of 3.4 from 5 starts on Google Play Store as of June 6, 2026. This study evaluates the perception of 16,950 users through sentiment classification and complaint topic identification. Six models were compared: three TF-IDF-based approaches using Complement Naive Bayes, Support Vector Machine, and Logistic Regression algorithms, alongside three IndoBERT variants comprising feature extraction with SVM, feature extraction with Logistic Regression at C=0.01, and full fine-tuning with a classification head. Topic modeling employed Latent Dirichlet Allocation with eight topics determined by the highest Coherence Score of 0.5487. Fine-tuned IndoBERT achieved the best performance with an F1-Macro of 0.8736 and accuracy of 0.9339, surpassing all baselines. The dominant complaint identified was application failure following updates, accounting for 23.3 percent of total negative reviews, followed by poor usability at 17.6 percent and account registration failures at 17.3 percent.*

Keywords – Health Transformation, IndoBERT, IndoBERT Fine-tuned, Sentiment Analysis, SATUSEHAT Mobile.

I. PENDAHULUAN

Sektor kesehatan Indonesia mengalami digitalisasi yang luar biasa sejak pandemi COVID-19 yang memaksa pemerintah mengakselerasi adopsi teknologi informasi dalam layanan publik secara masif dan dalam waktu yang singkat [1]. Kementerian Kesehatan merespons dengan meluncurkan platform SATUSEHAT pada 26 Juli 2022 sebagai sistem integrasi data rekam medis nasional berbasis *Indonesia Health Services*, menghubungkan rumah sakit, puskesmas, laboratorium, dan apotek ke dalam satu ekosistem data kesehatan terpadu [2]. Visi jangka panjangnya adalah mengurangi ketergantungan pasien dalam membawa dokumen fisik saat berpidah fasilitas kesehatan, karena seluruh riwayat medis sudah tersimpan dan dapat diakses secara digital. Setahun kemudian, pada 1 Maret 2023, PeduliLindungi resmi bertransformasi menjadi SATUSEHAT Mobile. Ambisi yang tergolong besar ini belum sepenuhnya diiimbangi oleh pengalaman pengguna yang memuaskan di lapangan.

Google Play Store mencatat lebih dari 50 juta unduhan SATUSEHAT Mobile per Juni 2026, namun rating rata-ratanya bertahan di angka 3,4 bintang dari 5. Dari 20.000 ulasan yang dikumpulkan dalam penelitian ini, lebih dari 55 persen memberikan rating satu bintang, distribusi yang sangat condong ke arah negatif sebagaimana terlihat pada Gambar 1. Ketimpangan antara skala yang masif dan tingkat kepuasan yang rendah ini bukan anomali kecil yang bisa diabaikan, melainkan sinyal bahwa ada permasalahan sistemik yang belum teridentifikasi secara terstruktur dan menyeluruh [3], [4]. Pemerintah telah menginvestasikan sumber daya yang signifikan untuk membangun platform ini, sehingga memahami akar permasalahan dari sudut pandang pengguna nyata menjadi prioritas yang tidak bisa ditunda.



Gambar 1. Distribusi Rating dan Tren Ulasan Pengguna SATUSEHAT Mobile

Menganalisis ulasan pengguna secara manual tidak praktis pada skala jutaan entri yang terus bertambah setiap harinya. Pendekatan berbasis pemrosesan bahasa alami menjadi solusi yang relevan, dengan analisis sentimen untuk mengklasifikasikan polaritas opini dan topic modeling untuk mengungkap pola tematik tersembunyi di balik keluhan tersebut [5], [6]. Kombinasi keduanya menghasilkan gambaran yang jauh lebih lengkap dibandingkan sekadar menghitung berapa banyak ulasan negatif yang masuk, karena ia juga menjawab pertanyaan yang jauh lebih penting bagi pengembang, di aspek mana tepatnya pengguna merasa tidak puas dan seberapa sering keluhan itu berulang.

Model berbasis Transformer, khususnya turunan BERT, telah menggeser standar performa analisis sentimen Bahasa Indonesia dalam beberapa tahun terakhir dan menjadi pendekatan

yang semakin banyak diadopsi dalam penelitian NLP lokal [7]. IndoBERT yang dilatih pada kumpulan dokumen besar teks Indonesia menawarkan pemahaman kontekstual yang lebih kaya dibandingkan representasi statistik konvensional seperti TF-IDF, karena mampu membedakan makna kata yang sama dalam konteks kalimat yang berbeda [8]. Meski demikian, pertanyaan praktis yang relevan bagi peneliti dan praktisi adalah seberapa signifikan keunggulan itu untuk domain ulasan aplikasi yang kaya bahasa informal dan singkatan, dan apakah biaya komputasi yang jauh lebih besar sepadan dengan peningkatan performa yang diperoleh dibandingkan baseline sederhana [9].

Studi terdahulu yang menganalisis ulasan aplikasi pemerintah Indonesia umumnya mengandalkan satu metode klasifikasi tunggal tanpa mengintegrasikan analisis topik keluhan yang lebih mendalam [10]. Perbandingan sistematis antara pendekatan Machine Learning konvensional dan Transformer untuk domain aplikasi kesehatan digital berbahasa Indonesia masih sangat langka dalam literatur yang tersedia, terutama untuk platform yang baru diluncurkan dan masih dalam fase pengembangan aktif seperti SATUSEHAT Mobile [11].

Penelitian ini mengusulkan pendekatan yang menggabungkan komparasi enam model klasifikasi sentimen dengan topic modeling berbasis LDA terhadap data hasil *scraping* Google Play Store periode April 2023 hingga Juni 2026. Rentang waktu ini sengaja dipilih agar mencakup tiga fase yang berbeda karakternya, yaitu fase transisi dari PeduliLindungi di mana pengguna lama beradaptasi dengan platform baru, fase stabilisasi di mana keluhan mulai mencerminkan masalah fungsional yang lebih mendasar, dan kondisi terkini yang memberikan gambaran paling aktual tentang persepsi pengguna.

II. TINJAUAN TEORI

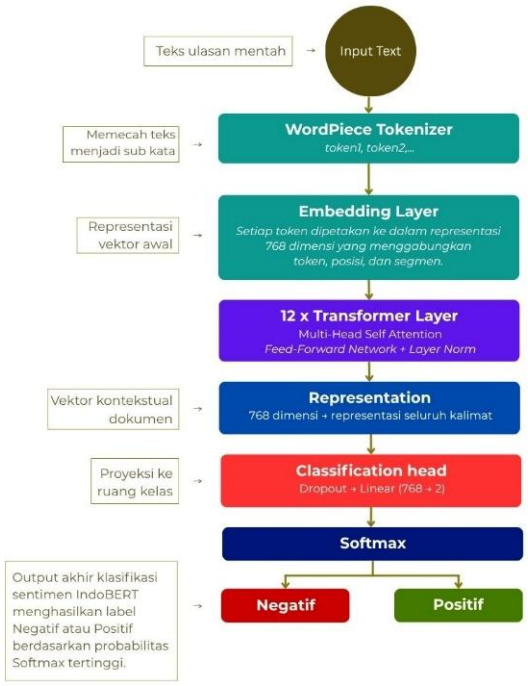
2.1 Analisis Sentimen

Analisis sentimen adalah cabang pemrosesan bahasa alami yang dapat mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini subjektif dari teks ke dalam kategori seperti positif, negatif, atau netral [12]. Bidang ini telah berkembang pesat seiring bertambahnya volume data ulasan daring yang dihasilkan pengguna di berbagai platform digital. Pendekatan berbasis Machine Learning mendominasi penelitian modern karena kemampuannya belajar dari pola data tanpa harus mendefinisikan aturan secara manual, menjadikannya lebih adaptif terhadap variasi linguistik yang kaya dalam bahasa informal [13]. Dua representasi fitur yang paling umum digunakan adalah TF-IDF untuk pendekatan statistik berbasis frekuensi kata, dan embedding kontekstual dari model Transformer yang mampu menangkap nuansa makna berdasarkan konteks kalimat secara keseluruhan.

TF-IDF mengukur relevansi kata terhadap dokumen dengan mempertimbangkan frekuensi kemunculannya sekaligus kelangkaannya di seluruh kumpulan dokumen, sehingga kata yang spesifik dan informatif mendapat bobot lebih tinggi dibandingkan kata umum [14]. Keterbatasan utamanya adalah ketidakmampuan menangkap makna kontekstual, sehingga kalimat tidak bagus dan sangat bagus bisa menghasilkan representasi yang serupa padahal polaritasnya berlawanan secara semantik. Meski demikian, kombinasi TF-IDF dengan algoritma klasifikasi yang tepat tetap menghasilkan performa kompetitif untuk berbagai domain teks termasuk ulasan berbahasa Indonesia, terutama ketika fitur dioptimalkan dengan pemilihan ngram yang tepat [9], [15].

2.2 Arsitektur Transformer dan IndoBERT

Transformer diperkenalkan oleh Vaswani et al. pada 2017 sebagai alternatif arsitektur rekuren yang memproses seluruh sekuens input secara paralel melalui mekanisme self-attention, yang menghitung bobot relevansi setiap token terhadap seluruh token lainnya dalam satu langkah komputasi [16]. Kemampuan menangkap dependensi jangka panjang tanpa batasan jarak antar token menjadikannya unggul dibandingkan LSTM dan GRU untuk tugas pemahaman teks yang kompleks [7]. BERT mengembangkan Transformer dengan pendekatan bidireksional, membaca konteks dari kiri ke kanan dan kanan ke kiri secara bersamaan untuk menghasilkan representasi token yang lebih kaya secara semantik, kemudian dilatih dengan dua tugas *pre-training* yaitu *Masked Language Modeling* dan *Next Sentence Prediction* [17]. Arsitektur IndoBERT untuk klasifikasi sentimen dalam penelitian ini diilustrasikan pada Gambar 2, yang menunjukkan alur dari input teks hingga prediksi kelas sentimen.



Gambar 2. Arsitektur IndoBERT untuk Klasifikasi Sentimen

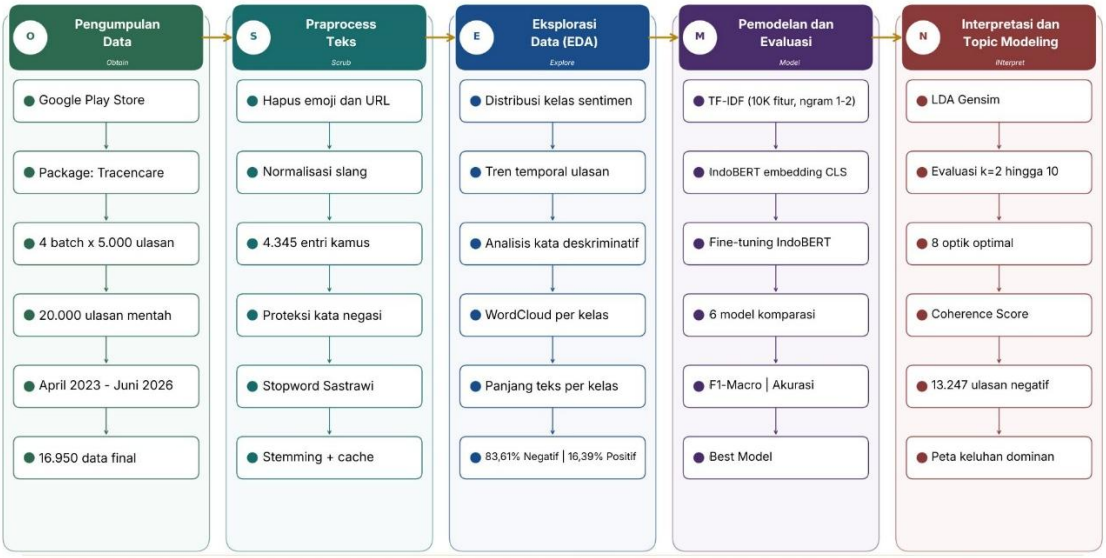
2.3 Topic Modeling dengan LDA

Latent Dirichlet Allocation (LDA) adalah model probabilistik generatif yang mengasumsikan setiap dokumen sebagai pengelompokan berdasarkan inti bahasannya dan setiap topik sebagai distribusi probabilitas atas kosakata dalam kumpulan dokumen [4]. Model ini dilatih secara *unsupervised* sehingga tidak memerlukan data berlabel, menjadikannya cocok untuk eksplorasi tematik pada kumpulan teks berskala besar tanpa melakukan pelabelan manual. Proses inferensi LDA menggunakan pendekatan Bayesian untuk menemukan distribusi topik yang paling mungkin menghasilkan kumpulan dokumen yang diamati, dengan parameter alpha yang mengontrol *sparsity* distribusi topik per dokumen dan parameter eta yang mengontrol *sparsity* distribusi kata per topik [6]. Kualitas model dievaluasi menggunakan Coherence Score untuk mengukur koherensi semantik antar kata teratas per topik, dan Log Perplexity untuk mengukur kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya, di mana nilai Coherence Score yang lebih tinggi dan

Log Perplexity yang lebih rendah mengindikasikan model yang lebih baik [18].

III. METODE

Setiap tahapan dalam penelitian ini, dari cara data dikumpulkan hingga bagaimana model dievaluasi, dibuat berdasarkan temuan dari tahapan sebelumnya sehingga seluruh proses terhubung secara logis seperti yang terlihat pada Gambar 3.



Gambar 3. Alur Metodologi Penelitian

3.1 Pengumpulan Data

Data dikumpulkan menggunakan library google-play-scraper pada Python dengan mengakses ulasan aplikasi SATUSEHAT Mobile yang memiliki package name com.telkom.tracencare di Google Play Store. *Scraping* dilakukan dalam empat *batch* masing-masing 5.000 ulasan menggunakan mekanisme pagination berbasis *continuation token* untuk menjaga konsistensi urutan data, menghasilkan 20.000 ulasan mentah dengan rentang waktu April 2023 hingga Juni 2026. Ulasan dengan rating bintang tiga dieliminasi karena secara semantik ambigu dan tidak merepresentasikan polaritas sentimen yang jelas, mengikuti pendekatan yang konsisten dengan beberapa penelitian sentimen berbasis rating sebelumnya [14], [19]. Rating bintang satu dan dua diberi label Negatif, sementara bintang empat dan lima diberi label Positif, menghasilkan dataset berlabel dua kelas yang siap diproses lebih lanjut.

Tabel 1. Ringkasan Lima Data Teratas Ulasan SATUSEHAT Mobile

No	reviewId	content	score	at	appVersion
1	25e741a8-d0ba...	Aplikasinya Bagus	5	2026-06-05	8.8.1
2	d702f73c-7b8c...	Proses *login* sangat mengecewakan. *Login* bermasalah...	1	2026-06-05	8.8.1
3	a8187492-9c25...	cukup baik hanya bbrp hasil med. chek dr kemkes...	5	2026-06-04	8.8.1
4	c75d6f30-57d5...	tgl lahir saya tidak dapat di ubah	1	2026-06-04	None

No	reviewId	content	score	at	appVersion
		mandiri, ja...			
5	94057e2f-9efc...	baik dan menyenangkan	5	2026-06-04	None

3.2 Praproses Teks

Desain *pipeline* praproses didasarkan pada temuan eksplorasi data, bukan asumsi. Eksplorasi mengungkap bahwa data mengandung emoji masif, bahasa informal dan singkatan yang sangat beragam, serta kata negasi yang kerap menjadi pembeda polaritas utama antara ulasan negatif dan positif. Pipeline dirancang dengan enam tahap berurutan. Pembersihan teks menghapus emoji dan karakter non-ASCII, URL, mention, angka, dan tanda baca. Normalisasi kata tidak baku menggunakan kamus colloquial-indonesian-lexicon dengan 4.345 entri. Kata negasi termasuk tidak, bukan, belum, jangan, dan bisa dilindungi dari penghapusan stopword menggunakan mekanisme whitelist untuk menjaga polaritas kalimat tetap utuh. Penghapusan stopword menggunakan 126 kata Sastrawi dengan *whitelist* negasi yang aktif, dilanjutkan *stemming* menggunakan Sastrawi Stemmer dengan mekanisme *cache* yang menghasilkan 13.182 kata unik dan mempercepat komputasi secara signifikan. Pipeline TF-IDF diterapkan *custom stopword* tambahan pada kata-kata yang terbukti muncul merata di kedua kelas berdasarkan analisis rasio kemunculan. Ringkasan parameter tersaji pada Tabel 2.

Tabel 2. Parameter Praproses Teks

Tahap	Konfigurasi	Keterangan
Normalisasi *slang*	4.345 entri kamus	Kamus colloquial + *domain* khusus
Proteksi negasi	tidak, bukan, belum, jangan, bisa	*Whitelist* dari stopword Sastrawi
Stopword removal	126 kata Sastrawi	Dengan *whitelist* negasi aktif
Stemming	Sastrawi Stemmer + *cache*	13.182 kata unik ter-*cache*
Filter panjang	≥ 10 karakter setelah *cleaning*	Eliminasi ulasan tanpa konteks
Custom stopword TF-IDF	aplikasi, vaksin, sertifikat, mobile	Kata tidak diskriminatif antar kelas

3.3 Pembagian Data dan Penanganan Ketidakseimbangan Kelas

Dataset final sebanyak 16.950 ulasan memiliki distribusi yang sangat timpang yaitu 83,61 persen kelas Negatif dan hanya 16,39 persen kelas Positif. Ketimpangan ini bukan artefak dari proses pengumpulan data, melainkan cerminan nyata dari perilaku pengguna di mana mereka cenderung lebih terdorong menulis ulasan panjang dan ekspresif ketika mengalami masalah dibandingkan ketika merasa puas. Pembagian data menggunakan stratified split 80:20 untuk memastikan proporsi kelas terjaga dengan baik di data latih maupun data uji, menghasilkan 13.560 data latih dan 3.390 data uji. Ketidakseimbangan ditangani menggunakan pembobotan kelas pada seluruh model tanpa augmentasi data sintetis, karena distribusi ini mencerminkan realitas yang perlu dipertahankan keasliannya agar model dapat belajar dari kondisi yang representatif [15]. Bobot dihitung dengan metode *invers* frekuensi kelas, menghasilkan bobot 0,598 untuk kelas Negatif dan 3,051 untuk kelas Positif.

3.4 Konfigurasi Model

Dua grup model diuji secara komparatif untuk membandingkan pendekatan representasi teks yang berbeda. Grup pertama menggunakan TF-IDF dengan 10.000 fitur, ngram range (1,2) yang mencakup unigram dan bigram, minimum document frequency 2 untuk menyaring kata sangat jarang, dan maximum document frequency 0,95 untuk menyaring kata sangat umum. Tiga algoritma klasifikasi diuji pada representasi ini yaitu CNB dengan sample weight balanced, SVM LinearSVC dengan class weight balanced, dan Logistic Regression dengan class *weight balanced* serta parameter regularisasi C=1,0. Grup kedua menggunakan IndoBERT model indobenchmark/indobert-base-p1 dalam tiga variasi yang berbeda. Konfigurasi lengkap seluruh model tersaji pada Tabel 3.

Tabel 3. Konfigurasi Enam Model Klasifikasi

Grup	Model	Representasi	Parameter Utama
TF-IDF	CNB	TF-IDF 10K fitur	sample weight balanced
TF-IDF	SVM	TF-IDF 10K fitur	class weight balanced
TF-IDF	LR	TF-IDF 10K fitur	C=1,0, class weight balanced
IndoBERT	Emb + SVM	CLS 768 dim	class weight balanced
IndoBERT	Emb + LR	CLS 768 dim	C=0,01, class weight balanced
IndoBERT	Fine-tuned	Full fine-tuning	lr=2e-5, batch=16, epoch maks 5, early stop patience 2

3.5 Topic Modeling

Topic modeling diterapkan pada 13.247 ulasan kelas Negatif menggunakan LDA dari library Gensim dengan parameter alpha dan eta yang disetel secara adaptif. Kosakata difilter untuk menghapus kata yang muncul di bawah 10 dokumen agar model tidak terdistraksi oleh kata sangat jarang yang tidak informatif, dan di atas 70 persen dokumen agar kata terlalu umum yang hadir di hampir setiap ulasan tidak mendominasi topik. Jumlah topik optimal dipilih dengan mengevaluasi Coherence Score dan Log Perplexity untuk nilai k dari 2 hingga 10, masing-masing dilatih dengan 15 iterasi untuk memastikan konvergensi yang memadai. Topik yang terbentuk kemudian diberi nama berdasarkan interpretasi manual terhadap sepuluh kata dengan probabilitas tertinggi di setiap topik, dan distribusi ulasan per topik dihitung berdasarkan penetapan topik dominan untuk setiap dokumen.

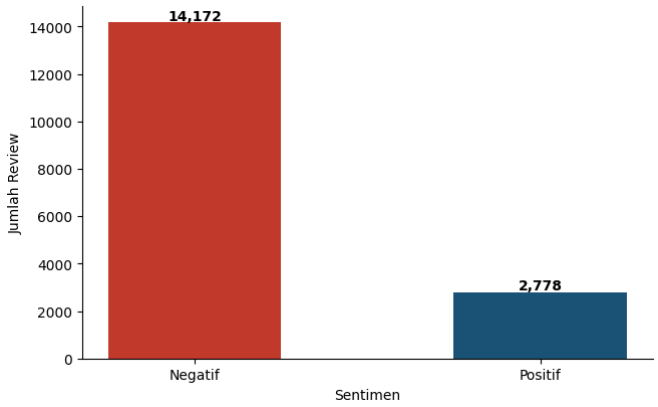
IV. HASIL DAN PEMBAHASAN

4.1 Karakteristik Dataset

Diperoleh sebanyak 20.000 ulasan mentah hasil *scraping*, sebanyak 775 dengan rating tiga dieliminasi dan 19.225 ulasan berlabel tersisa untuk diproses lebih lanjut. Tahap praproses kemudian menyingkirkan 269 ulasan yang menjadi kosong setelah pembersihan, sebagian besar berupa deretan emoji tanpa teks bermakna seperti baris panjang simbol bintang yang menghabiskan batas 500 karakter, dan 1.006 ulasan duplikat berdasarkan konten aslinya. Dataset final berjumlah 16.950 ulasan dengan rentang waktu April 2023 hingga Juni 2026.

Analisis temporal mengungkap dua puncak ulasan negatif yang menonjol dan dapat

dijelaskan secara kontekstual. Puncak pertama terjadi April 2023, bertepatan tepat dengan masa transisi dari PeduliLindungi ketika jutaan pengguna lama tiba-tiba menghadapi aplikasi baru dengan antarmuka, alur *login*, dan fitur yang berbeda dari yang selama ini mereka kenal. Puncak kedua muncul Maret hingga April 2024, kemungkinan besar dipicu pembaruan besar yang tidak kompatibel dengan sebagian perangkat pengguna, memunculkan gelombang keluhan serupa. Distribusi kelas akhir setelah praproses ditunjukkan pada Gambar 4, memperlihatkan 83,61 persen Negatif dan 16,39 persen Positif, sebuah ketimpangan yang mencerminkan kondisi nyata dan bukan artefak dari proses pengumpulan data.



Gambar 4. Distribusi Kelas Sentimen

4.2 Perbandingan Performa Model Klasifikasi

Hasil evaluasi keenam model pada data uji tersaji pada Tabel 4, dengan F1-Macro sebagai metrik utama karena lebih representatif untuk data yang tidak seimbang dibandingkan akurasi semata. IndoBERT Fine-tuned meraih performa terbaik di semua metrik dengan F1-Macro 0,8736, F1-Negatif 0,9609, F1-Positif 0,7863, dan akurasi 0,9339. Di antara model berbasis TF-IDF, Logistic Regression unggul dengan F1-Macro 0,8224, sementara SVM menghasilkan 0,8135 dan CNB 0,8181.

Tabel 4. Perbandingan Performa Model Klasifikasi Sentimen

Model	F1-Macro	F1-Negatif	F1-Positif	Akurasi
TF-IDF + CNB	0,8181	0,9415	0,6948	0,9018
TF-IDF + SVM	0,8135	0,9322	0,6948	0,8891
TF-IDF + LR	0,8224	0,9348	0,7100	0,8935
IndoBERT Emb + SVM	0,8113	0,9297	0,6930	0,8855
IndoBERT Emb + LR	0,8342	0,9396	0,7287	0,9012
IndoBERT Fine-tuned	0,8736	0,9609	0,7863	0,9339

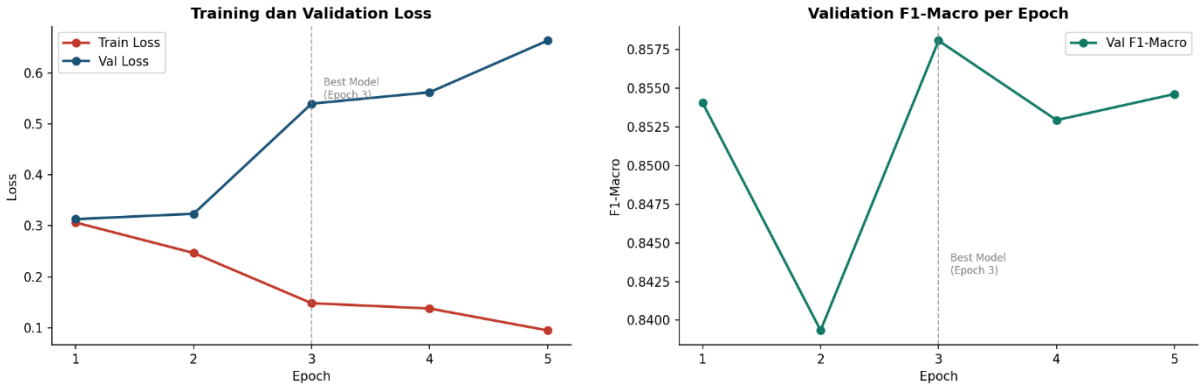
Satu temuan yang paling layak dicermati adalah performa IndoBERT dalam mode ekstraksi fitur statis yang justru lebih rendah dari TF-IDF Logistic Regression, perbedaan yang secara statistik bermakna. Fenomena ini dapat dijelaskan dari karakteristik linguistik datanya. Ulasan pengguna SATUSEHAT Mobile dipenuhi bahasa informal, singkatan seperti apk dan

gak, serta ekspresi frustrasi yang jauh dari distribusi teks formal yang mendominasi korpus pre-*training* IndoBERT. Ketika bobot model tidak diperbarui melalui fine-tuning, bias domain bawaan ini justru kontraproduktif untuk tugas klasifikasi di domain yang berbeda signifikan dari data pelatihan awal.

IndoBERT Fine-tuned tidak hanya melampaui semua *baseline* dengan *margin* yang cukup besar, tetapi juga mencatat peningkatan F1-Positif sebesar 7,63 poin dibandingkan TF-IDF + LR, dari 0,7100 menjadi 0,7863. Peningkatan terbesar justru terjadi pada kelas minoritas Positif yang secara historis paling sulit dipelajari model akibat ketidakseimbangan data yang ekstrem. Ini mengkonfirmasi bahwa *fine-tuning* kontekstual yang dikombinasikan dengan pembobotan kelas adalah strategi yang sangat efektif untuk klasifikasi sentimen pada data ulasan berbahasa Indonesia dengan distribusi yang tidak seimbang, dan bahwa biaya komputasi tambahan *fine-tuning* sebanding dengan peningkatan performa yang diperoleh.

4.3 Analisis *Training* Curve Fine-tuning

Kurva *training* loss dan F1-Macro validasi selama proses fine-tuning ditunjukkan pada Gambar 6. *Training* loss turun konsisten dari 0,307 di epoch pertama hingga 0,095 di epoch kelima, memperlihatkan bahwa model belajar dari data latih dengan baik sepanjang proses. Namun validation loss bergerak ke arah sebaliknya: dari 0,313 di epoch pertama, sempat turun ke 0,324 di epoch kedua, lalu melonjak ke 0,540 di epoch ketiga dan terus merangkak ke 0,664 di epoch kelima. Pola divergensi antara *training* loss yang terus turun dan validation loss yang naik setelah epoch kedua adalah tanda klasik overfitting yang terjadi ketika model mulai menghafal pola spesifik data latih yang tidak dapat digeneralisasi ke data baru.



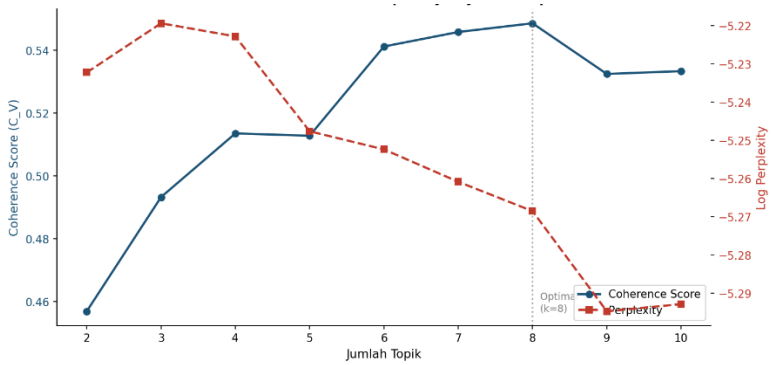
Gambar 6. Kurva *Training* Loss dan Validation F1-Macro per Epoch

F1-Macro validasi mencapai puncaknya di epoch ketiga dengan nilai 0,8581, turun ke 0,8529 di epoch keempat, lalu sedikit membaik ke 0,8546 di epoch kelima. Selisih antar epoch yang tidak cukup besar untuk memicu *early stopping* dengan *patience* 2 membuat *training* menyelesaikan seluruh 5 epoch. Model epoch ketiga dipilih sebagai model final berdasarkan mekanisme load best model, dan ketika dievaluasi pada data uji yang sepenuhnya terpisah, menghasilkan F1-Macro 0,8736, jauh di atas performa validasi terbaiknya sendiri, mengindikasikan bahwa model berhasil menangkap pola yang cukup general meski tanda-tanda overfitting mulai muncul.

4.4 Penentuan Jumlah Topik Optimal

Coherence Score dan Log Perplexity untuk k dari 2 hingga 10 ditampilkan pada Gambar 7. Coherence Score naik dari 0,4569 di k=2 hingga mencapai puncak 0,5487 di k=8, kemudian turun ke 0,5325 di k=9 dan sedikit naik ke 0,5334 di k=10, menunjukkan bahwa menambah

topik melampaui 8 tidak memberikan peningkatan koherensi yang berarti. Log Perplexity menurun secara monoton sepanjang rentang yang diuji, namun laju penurunannya melambat setelah k=8. Karena kedua metrik secara bersamaan paling mendukung k=8, yakni Coherence Score tertinggi dan Log Perplexity terendah pada titik yang sama, jumlah ini dipilih sebagai konfigurasi optimal dengan justifikasi yang kuat dari dua dimensi evaluasi sekaligus.



Gambar 7. Coherence Score dan Perplexity untuk Penentuan Jumlah Topik Optimal

4.5 Hasil dan Interpretasi Topik Modeling

Delapan topik keluhan yang teridentifikasi beserta distribusinya tersaji pada Tabel 5 dan visualisasi *WordCloud* masing-masing topik ditampilkan pada Gambar 8.

Tabel 5. Delapan Topik Keluhan Pengguna SATUSEHAT Mobile

No	Topik	Jumlah	Persentase
1	Sertifikat Vaksin Tidak Dapat Diunduh	1.403	10,6%
2	Sertifikat Vaksin Hilang atau Tidak Muncul	1.729	13,1%
3	Kegagalan Verifikasi OTP dan *Login*	1.506	11,4%
4	Ketidaksesuaian Data NIK dan KTP	263	2,0%
5	Kegagalan Pendaftaran dan Akses Akun	2.296	17,3%
6	Aplikasi Tidak Bisa Dibuka Setelah Pembaruan	3.087	23,3%
7	Kekecewaan Transisi dari PeduliLindungi	627	4,7%
8	Aplikasi Dinilai Menyusahkan Pengguna	2.336	17,6%

Gambar 8. Distribusi dan WordCloud Delapan Topik Keluhan Pengguna

Topik 6 tentang kegagalan aplikasi setelah pembaruan mendominasi dengan 23,3 persen dari seluruh ulasan negatif. Kata-kata penentu topik ini yaitu bisa, aplikasi, tidak, buka, dan update menggambarkan skenario yang sangat spesifik, pengguna yang sebelumnya dapat menggunakan aplikasi dengan normal mendapatinya tidak bisa dibuka sama sekali setelah proses pembaruan dilakukan. Ini bukan sekadar *bug* acak melainkan indikasi masalah regresi yang berulang dalam siklus pengembangan, di mana setiap rilis baru memicu gelombang ketidakkompatibilan pada sebagian perangkat pengguna yang menggunakan versi Android atau spesifikasi *hardware* tertentu. Pola seperti ini umumnya mengindikasikan kurangnya pengujian regresi yang komprehensif sebelum rilis ke publik.

Topik 5 tentang kegagalan pendaftaran dan akses akun mencatat 17,3 persen, dengan keluhan berputar pada proses *login* yang berulang kali gagal, pendaftaran akun baru yang tidak bisa diselesaikan, dan akses melalui *email* yang bermasalah. Secara kumulatif, Topik 1 dan 2 tentang sertifikat vaksin mencakup 23,7 persen dari seluruh keluhan negatif, menjadikannya kategori terbesar bila diakumulasikan dan menggarisbawahi bahwa fitur yang paling banyak dicari pengguna justru paling sering bermasalah. Topik 7 tentang kekecewaan transisi dari PeduliLindungi yang hanya 4,7 persen dan Topik 4 tentang ketidaksesuaian NIK yang 2,0 persen menunjukkan permasalahan yang bersifat lebih spesifik dan mulai mereda seiring waktu, berbeda dari keluhan teknis lain seperti Topik 5, 6, dan 8 yang memperlihatkan persistensi lebih tinggi sepanjang periode pengamatan.

V. KESIMPULAN

Dari komparasi enam model terhadap 16.950 ulasan SATUSEHAT Mobile, fine-tuned IndoBERT terbukti sebagai pendekatan terbaik dengan F1-Macro 0,8736 dan akurasi 0,9339, menjawab pertanyaan tentang model mana yang paling efektif untuk domain ini, sementara di antara model TF-IDF, Logistic Regression unggul dengan F1-Macro 0,8224 dan bahkan melampaui IndoBERT dalam mode ekstraksi fitur statis, sebuah temuan yang menegaskan bahwa keunggulan Transformer baru sepenuhnya terealisasi ketika fine-tuning dilakukan pada domain target. Topic modeling LDA dengan delapan topik mengungkap bahwa kegagalan aplikasi setelah pembaruan mendominasi keluhan sebesar 23,3 persen, disusul penilaian aplikasi yang menyusahkan sebesar 17,6 persen dan kegagalan pendaftaran akun sebesar 17,3 persen, sementara permasalahan sertifikat vaksin secara kumulatif mencakup 23,7 persen dari seluruh ulasan negatif, memberikan peta keluhan konkret yang dapat

langsung ditindaklanjuti pengembang dan Kementerian Kesehatan untuk memprioritaskan perbaikan yang berdampak paling besar bagi pengguna.

VI. DAFTAR PUSTAKA

[1] Muhammad Alfarizi dan Ngatindriatun, "Digital Health 5.0 for Health Equity in Rural Developing Regions: A Bibliometric and Systematic Review," Mar 2026. [Daring]. Tersedia pada: https://journal.unesa.ac.id/index.php/jorris

[2] Biro Komunikasi dan Pelayanan Masyarakat Kementerian Kesehatan RI, "Kemenkes Luncurkan Platform SATUSEHAT Untuk Integrasikan Data Kesehatan Nasional," Jakarta, Jul 2022. Diakses: 7 Juni 2026. [Daring]. Tersedia pada: https://kemkes.go.id/id/kemenkes-ri-resmi-luncurkan-platform-integrasi-data-layanan-kesehatan-bernama-satusehat

[3] M. Hadwan, M. Al-Sarem, F. Saeed, dan M. A. Al-Hagery, "An Improved Sentiment Classification Approach for Measuring User Satisfaction toward Governmental Services' Mobile Apps Using Machine Learning Methods with Feature Engineering and SMOTE Technique," *Applied Sciences*, vol. 12, no. 11, hlm. 5547, Mei 2022, doi: 10.3390/app12115547.

[4] Y. Zhai, X. Song, Y. Chen, dan W. Lu, "A Study of Mobile Medical App User Satisfaction Incorporating Theme Analysis and Review Sentiment Tendencies," *Int. J. Environ. Res. Public Health*, vol. 19, no. 12, hlm. 7466, Jun 2022, doi: 10.3390/ijerph19127466.

[5] E. L. Funnell, B. Spadaro, N. Martin-Key, T. Metcalfe, dan S. Bahn, "mHealth Solutions for Mental Health Screening and Diagnosis: A Review of App User Perspectives Using Sentiment and Thematic Analysis," *Front. Psychiatry*, vol. 13, Apr 2022, doi: 10.3389/fpsyt.2022.857304.

[6] Y. Shan, M. Ji, W. Xie, K.-Y. Lam, dan C.-Y. Chow, "Public Trust in Artificial Intelligence Applications in Mental Health Care: Topic Modeling Analysis," *JMIR Hum. Factors*, vol. 9, no. 4, hlm. e38799, Des 2022, doi: 10.2196/38799.

[7] H. Imaduddin, F. Y. A'la, dan 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, doi: 10.14569/IJACSA.2023.0140813.

[8] F. Koto, A. Rahimi, J. H. Lau, dan T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," dalam *Proceedings of the 28th International Conference on Computational Linguistics*, Stroudsburg, PA, USA: International Committee on Computational Linguistics, 2020, hlm. 757–770. doi: 10.18653/v1/2020.coling-main.66.

[9] H. Murfi, Syamsyuriani, T. Gowandi, G. Ardaneswari, dan S. Nurrohmah, "BERT-based combination of convolutional and recurrent neural network for indonesian sentiment analysis," *Appl. Soft Comput.*, vol. 151, hlm. 111112, Jan 2024, doi: 10.1016/j.asoc.2023.111112.

[10] S. Amrie, S. Kurniawan, J. H. Windiatmaja, dan Y. Ruldeviyani, "Analysis of Google Play Store's Sentiment Review on Indonesia's P2P Fintech Platform," dalam *2022 IEEE Delhi Section Conference (DELCON)*, IEEE, Feb 2022, hlm. 1–5. doi: 10.1109/DELCON54057.2022.9753108.

[11] R. I. Perwira, V. A. Permadi, D. I. Purnamasari, dan R. P. Agusdin, "Domain-Specific Fine-Tuning of IndoBERT for Aspect-Based Sentiment Analysis in Indonesian Travel User-Generated Content," *Journal of Information Systems Engineering and Business Intelligence*, vol. 11, no. 1, hlm. 30–40, Mar 2025, doi: 10.20473/jisebi.11.1.30-40.

[12] K. L. Tan, C. P. Lee, dan K. M. Lim, "A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research," *Applied Sciences*, vol. 13, no. 7, hlm. 4550, Apr 2023, doi: 10.3390/app13074550.

[13] A. Palanivinayagam, C. Z. El-Bayeh, dan R. Damaševičius, "Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review," *Algorithms*, vol. 16, no. 5, hlm. 236, Apr 2023, doi: 10.3390/a16050236.

[14] A. D. Gustiavani dan M. Muljono, "Sentiment Classification of Indonesian E-Government Application Reviews Using Advanced Learning Models," *Journal of Applied Informatics and Computing*, vol. 10, no. 2, hlm. 1662–1673, Apr 2026, doi: 10.30871/jaic.v10i2.12217.

[15] I. G. B. A. Budaya dan I. K. P. Suniantara, "Comparison of Sentiment Analysis Algorithms with SMOTE Oversampling and TF-IDF Implementation on Google Reviews for Public Health Centers," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, hlm. 1077–1086, Jul 2024, doi: 10.57152/malcom.v4i3.1459.

[16] A. Vaswani *dkk.*, "Attention Is All You Need," 2017. Diakses: 7 Juni 2026. [Daring]. Tersedia pada: https://arxiv.org/abs/1706.03762

[17] J. Devlin, M.-W. Chang, K. Lee, dan K. Toutanova, "BERT: Pre-*training* of Deep Bidirectional Transformers for Language Understanding," dalam *Proceedings of the 2019 Conference of the North*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, hlm. 4171–4186. doi: 10.18653/v1/N19-1423.

[18] J. D. Romero, M. A. Feijoo-Garcia, G. Nanda, B. Newell, dan A. J. Magana, "Evaluating the Performance of Topic Modeling Techniques with Human Validation to Support Qualitative Analysis," *Big Data and Cognitive Computing*, vol. 8, no. 10, hlm. 132, Okt 2024, doi: 10.3390/bdcc8100132.

[19] A. Bau dan G. D. Kapitan, "Sentiment Analysis of National Health Insurance Participants' Reviews on Google Reviews," *Jurnal Jaminan Kesehatan Nasional*, vol. 5, no. 2, hlm. 365–378, Des 2025, doi: 10.53756/jjkn.v5i2.344.

Halaman ini sengaja dikosongkan.