

Aspect-Based Sentiment Analysis pada Ulasan Pengguna Aplikasi Mobile JKN Menggunakan Model Berbasis Transformer

Akhmad Syukur Pangestu¹⁾, Efrans Christian²⁾, Ariesta Lestari³⁾

¹⁾²⁾³⁾Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya
Tunjung Nyaho Jl. Yos Sudarso Palangka Raya 73112

¹⁾Pangestu.02@mhs.eng.upr.ac.id

²⁾efrans@it.upr.ac.id

³⁾ariesta@it.upr.ac.id

Abstrak

Peningkatan penggunaan aplikasi mobile menuntut adanya evaluasi kualitas layanan berdasarkan umpan balik pengguna. Penelitian ini bertujuan untuk melakukan Analisis Sentimen Berbasis Aspek (Aspect-Based Sentiment Analysis/ABSA) terhadap ulasan pengguna aplikasi Mobile JKN yang diambil dari Google Play Store. Data dikumpulkan melalui teknik scraping menggunakan pustaka google-play-scraper dan difilter agar hanya mencakup ulasan dalam bahasa Indonesia dengan jumlah minimal lima kata, menghasilkan sebanyak 55.688 data ulasan.

Tahapan preprocessing yang diterapkan mencakup pembersihan data, case folding, normalisasi, tokenisasi, penghapusan stopword, dan lemmatization. Pemodelan topik dilakukan menggunakan algoritma Latent Dirichlet Allocation (LDA), dan pelabelan aspek dibantu oleh teknologi Generative AI Google Gemini versi 2.0 Flash. Proses klasifikasi aspek dan sentimen dilakukan secara multitugas menggunakan tiga model berbasis transformer, yaitu IndoBERT, RoBERTa, dan DistilBERT.

Evaluasi menunjukkan bahwa model IndoBERT memperoleh akurasi tertinggi pada klasifikasi sentimen sebesar 96% dan akurasi klasifikasi aspek sebesar 91%. Dengan demikian, IndoBERT dipilih untuk diimplementasikan dalam antarmuka web berbasis Streamlit. Penelitian ini menunjukkan bahwa pendekatan ABSA berbasis transformer dan AI generatif dapat memberikan pemahaman lebih dalam terhadap opini pengguna dan menjadi dasar dalam perbaikan layanan aplikasi digital.

Kata kunci: Analisis Sentimen, ABSA, Mobile JKN, Transformer, BERT, Google Gemini, LDA

Abstract

The increasing use of mobile applications demands an evaluation of service quality based on user feedback. This study aims to conduct Aspect-Based Sentiment Analysis (ABSA) on user reviews of the JKN Mobile application taken from the Google Play Store. Data was collected through scraping techniques using the google-play-scraper library and filtered to only include reviews in Indonesian with a minimum of five words, resulting in 55,688 review data.

The preprocessing stages applied include data cleaning, case folding, normalization, tokenization, stopword removal, and lemmatization. Topic modeling was performed using the Latent Dirichlet Allocation (LDA) algorithm, and aspect labeling was assisted by Google Gemini's Generative AI technology version 2.0 Flash. The aspect and sentiment classification process was performed in a multi-tasking manner using three transformer-based models, namely IndoBERT, RoBERTa, and DistilBERT.

The evaluation showed that the IndoBERT model obtained the highest accuracy on sentiment classification at 96% and aspect classification accuracy at 91%. Thus, IndoBERT was chosen to be implemented in the Streamlit-based web interface. This research shows that the transformer-based ABSA approach and generative AI can provide a deeper understanding of user opinions and become the basis for improving digital application services.

Keywords: Sentiment Analysis, ABSA, Mobile JKN, Transformer, BERT, Google Gemini, LDA

1. PENDAHULUAN

Transformasi digital di Indonesia telah mendorong masyarakat untuk memanfaatkan aplikasi mobile dalam berbagai aktivitas harian. Berdasarkan data BPS tahun 2022, sebanyak 66,48% penduduk Indonesia mengakses internet, sebagian besar melalui perangkat mobile. Hal ini menunjukkan bahwa aplikasi digital memiliki peran penting dalam mendukung kebutuhan masyarakat, termasuk dalam sektor kesehatan. Salah satu aplikasi strategis adalah Mobile JKN, yang dikembangkan oleh BPJS Kesehatan untuk mempermudah akses layanan kesehatan secara daring. Hingga tahun 2025, aplikasi ini telah diunduh lebih dari 50 juta kali di Google Play Store, mencerminkan tingginya adopsi masyarakat terhadap layanan digital kesehatan.

Ulasan pengguna di Google Play Store menjadi sumber informasi yang kaya untuk mengevaluasi kepuasan pengguna dan mengidentifikasi area layanan yang perlu ditingkatkan. Ulasan tersebut mencerminkan beragam persepsi, mulai dari pengalaman positif hingga keluhan terhadap fitur tertentu. Namun, ulasan yang tidak terstruktur menimbulkan tantangan dalam pengolahan dan interpretasi data secara sistematis. Oleh karena itu, pendekatan analisis berbasis aspek (Aspect-Based Sentiment Analysis/ABSA) menjadi solusi yang tepat untuk mengevaluasi ulasan berdasarkan aspek layanan dan polaritas sentimen yang menyertainya.

Penelitian ini menggabungkan metode Latent Dirichlet Allocation (LDA) untuk pemodelan topik, Generative AI (Gemini) untuk pelabelan aspek, serta model transformer-based seperti IndoBERT, RoBERTa, dan DistilBERT untuk klasifikasi sentimen. Fokus utama penelitian adalah mengidentifikasi aspek-aspek dominan yang paling sering muncul dalam ulasan pengguna dan menganalisis sentimen terhadap aspek tersebut. Dengan pendekatan ini, diharapkan hasil analisis dapat memberikan rekomendasi berbasis data bagi pengembang aplikasi Mobile JKN dalam meningkatkan layanan dan pengalaman pengguna secara lebih terarah.

2. TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Analisis Sentimen adalah teknologi yang dirancang untuk memahami emosi serta pandangan pengguna terhadap suatu produk, layanan, atau topik tertentu. Dengan mendeteksi opini pengguna, baik positif, negatif, maupun netral, analisis ini dapat menggali wawasan mendalam mengenai suatu entitas atau aspek tertentu [1]

2.2 Analisis Sentimen Berbasis Aspek

Aspect-Based Sentiment Analysis (ABSA) merupakan cabang dari opinion mining yang mengekstraksi informasi spesifik mengenai aspek-aspek tertentu dalam ulasan pengguna. ABSA memungkinkan sistem untuk mengenali aspek yang menjadi perhatian dan menilai sentimen terhadap aspek tersebut, apakah positif, negatif, atau netral. Teknologi ini memberikan wawasan rinci tentang elemen penting dari produk atau layanan [2].

2.3 Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) adalah model probabilistik generatif yang digunakan untuk menemukan topik tersembunyi (latent topics) dalam kumpulan teks. Dalam model ini, setiap dokumen dianggap sebagai campuran dari beberapa topik, dan setiap topik merupakan distribusi atas kata-kata [3]. LDA memungkinkan pengungkapan struktur laten dalam data teks dengan memodelkan proses generatif dokumen menggunakan distribusi Dirichlet dan Multinomial [4].

2.4 Generative Artificial Intelligence

Generative AI merupakan cabang dari kecerdasan buatan yang fokus pada penciptaan konten baru seperti teks, gambar, audio, atau video, berdasarkan pola yang telah dipelajari [5]. Salah satu model terkini adalah Google Gemini, sebuah model bahasa multimodal dari Google DeepMind,

yang mampu menangani berbagai jenis data sekaligus. Dalam konteks penelitian ini, kemampuan Gemini dalam pemrosesan bahasa alami (Natural Language Processing) relevan untuk mendukung proses analisis sentimen berbasis aspek (ABSA) secara kontekstual [6].

2.5 Transformer

Transformer adalah arsitektur jaringan saraf yang diperkenalkan oleh untuk memproses data berurutan, seperti teks, dengan mekanisme self-attention guna memahami konteks dan hubungan antar elemen [7]. Berbeda dengan model RNN yang memproses secara berurutan, transformer memproses seluruh input secara paralel, memungkinkan pemahaman konteks lebih baik dan efisien dalam tugas NLP [8].

2.6 IndoBERT

IndoBERT merupakan adaptasi dari BERT yang dilatih khusus dengan korpus bahasa Indonesia. Dengan arsitektur bidirectional, model ini memahami kata dalam konteks kalimat secara menyeluruh, baik sebelum maupun sesudahnya, sehingga lebih akurat dalam memproses teks bahasa Indonesia [9].

2.7 RoBERTa

RoBERTa (Robustly Optimized BERT Pretraining Approach) adalah penyempurnaan dari BERT yang menghilangkan tugas Next Sentence Prediction, menggunakan teknik masking dinamis, serta melibatkan data pelatihan lebih besar. Dalam konteks ABSA, RoBERTa mampu mengenali kata bermuatan sentimen secara lebih akurat [10].

2.8 DistilBERT

DistilBERT merupakan versi ringan dari BERT yang dikembangkan melalui teknik knowledge distillation. Model ini mempertahankan sekitar 97% kinerja BERT dengan ukuran 60% lebih kecil dan kecepatan inferensi yang lebih cepat, sehingga cocok untuk implementasi yang membutuhkan efisiensi tinggi [11].

2.9 Evaluasi Model

Evaluasi kinerja model klasifikasi dilakukan menggunakan metrik seperti Accuracy, Precision, Recall, dan F1-score [12]. Nilai-nilai ini dihitung berdasarkan Confusion Matrix, yang menggambarkan jumlah prediksi benar dan salah dari tiap kelas.

Confusion Matrix terdiri dari empat komponen:

- True Positive (TP): data positif yang diprediksi benar,
- True Negative (TN): data negatif yang diprediksi benar,
- False Positive (FP): data negatif yang salah diprediksi positif,
- False Negative (FN): data positif yang salah diprediksi negatif.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Gambar 1. Confusion Matrix

Berikut rumus evaluasi model:

Accuracy sebuah algoritma dilihat dari perbandingan data yang telah terklasifikasi (TP + TN) dengan total data (TP + TN + FP + FN). Rumus dari metrik *accuracy* ditunjukkan pada Rumus 1.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \tag{1}$$

Precision adalah sebuah algoritma dilihat dari perbandingan data yang terklasifikasi dengan benar (TP) dengan total data yang telah terprediksi benar (TP + FP). Rumus dari metrik precision ditunjukkan pada Rumus 2.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Recall didefinisikan sebagai perbandingan antara data yang terklasifikasi dengan benar (TP) dengan total data yang sebenarnya benar (TP + FN). Rumus dari metrik recall ditunjukkan pada Rumus 3.

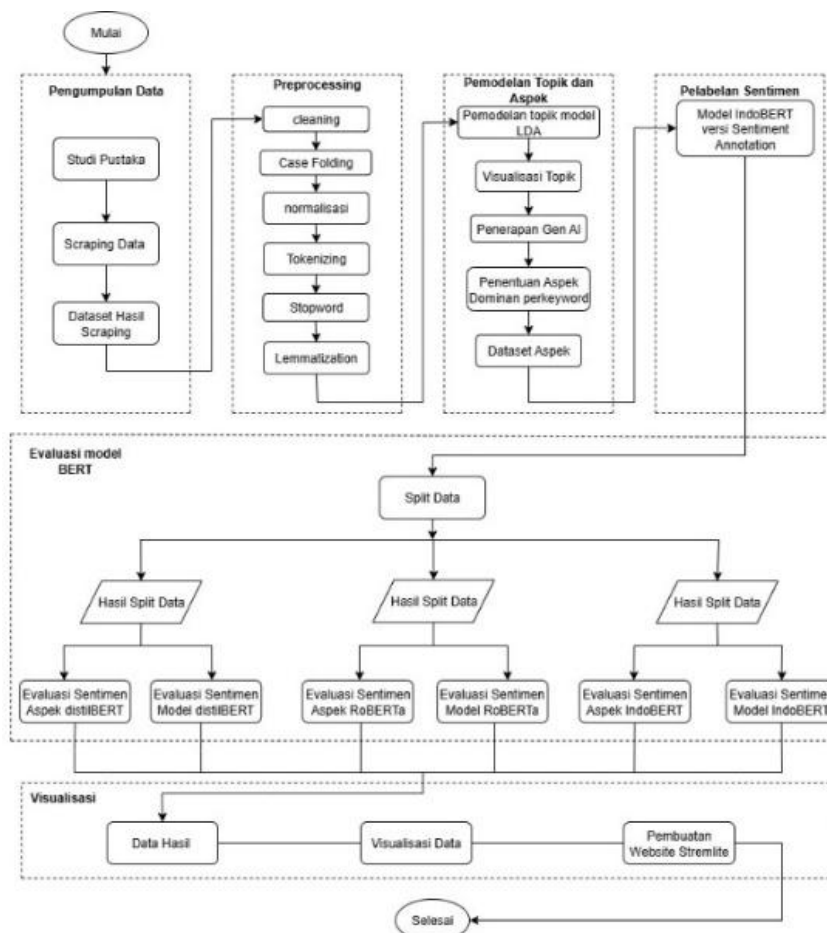
$$Recall = \frac{TP}{TP+FN} \tag{3}$$

F1 Score juga merupakan pengukur F. F1 score menyatakan keseimbangan antara precision dan recall. Rumus dari metrik F1 Score ditunjukkan pada Rumus 4.

$$F1Score = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{4}$$

3. METODE PENELITIAN

3.1 Alur Penelitian



Gambar 2. Alur Penelitian

Penelitian ini terdiri atas enam tahapan utama, yaitu: (1) pengumpulan data, (2) pra-pemrosesan data, (3) pemodelan topik dan aspek, (4) pelabelan sentimen, (5) evaluasi model, dan

(6) visualisasi data. Seluruh tahapan tersebut saling terintegrasi secara sistematis untuk menghasilkan analisis yang valid. Rangkaian alur penelitian ditunjukkan pada Gambar 2.

3.2 Pengumpulan Data

Data dikumpulkan melalui dua pendekatan: studi pustaka dan web scraping. Studi pustaka digunakan untuk memperoleh landasan teori terkait analisis sentimen berbasis aspek dan model BERT seperti IndoBERT, RoBERTa, dan DistilBERT. Sementara itu, data ulasan pengguna aplikasi Mobile JKN dikumpulkan dari Google Play Store menggunakan pustaka google-play-scraper. Ulasan yang diambil berbahasa Indonesia, mencakup periode 2021 hingga 2025, dan disimpan dalam format CSV.

3.3 Preprocessing Data

Tahap ini bertujuan untuk membersihkan dan menyiapkan data sebelum dianalisis. Proses yang dilakukan meliputi:

- a) Cleaning: menghapus karakter tidak relevan seperti angka, simbol, URL, dan tanda baca.
- b) Case folding: mengubah seluruh teks menjadi huruf kecil.
- c) Normalisasi: menyamakan kata tidak baku menjadi bentuk standar.
- d) Tokenisasi: memecah teks menjadi token kata.
- e) Stopword removal: menghapus kata-kata umum yang tidak memberikan makna penting.
- f) Lematisasi: mengubah kata menjadi bentuk dasarnya sesuai konteks kalimat.

Langkah-langkah ini memastikan data dalam kondisi optimal untuk dianalisis oleh model berbasis Transformer.

3.4 Pemodelan Topik dan Aspek

Pemodelan topik dilakukan menggunakan algoritma Latent Dirichlet Allocation (LDA) untuk mengekstraksi topik-topik laten dari ulasan pengguna. Evaluasi kualitas topik dilakukan menggunakan metrik coherence score, dan visualisasinya disajikan melalui pustaka pyLDAvis.

Identifikasi aspek dilakukan menggunakan pendekatan Generative AI berdasarkan kata-kata kunci dominan dari setiap topik. Hasil akhir berupa enam aspek utama yang mewakili tema ulasan terhadap layanan aplikasi Mobile JKN.

3.5 Pelabelan Sentimen

Pelabelan sentimen dilakukan secara otomatis menggunakan model IndoBERT. Proses dimulai dengan tokenisasi melalui AutoTokenizer dari HuggingFace, diikuti klasifikasi sentimen menjadi positif atau negatif berdasarkan skor probabilitas tertinggi. Pendekatan ini memungkinkan pelabelan dalam skala besar secara efisien tanpa memerlukan anotasi manual.

3.6 Evaluasi Model

Evaluasi performa model dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score. Tiga model utama yang dibandingkan adalah IndoBERT, RoBERTa, dan DistilBERT, dengan dua baseline: SVM dan Naïve Bayes. Hasil menunjukkan bahwa IndoBERT memiliki akurasi tertinggi ($\pm 92\%$), diikuti oleh RoBERTa ($\pm 90\%$) dan DistilBERT ($\pm 88\%$). Hal ini menunjukkan efektivitas model berbasis Transformer dalam analisis sentimen berbahasa Indonesia.

3.7 Visualisasi Data

Visualisasi digunakan untuk memperkuat interpretasi hasil. Teknik yang digunakan antara lain word cloud, bar chart, dan pie chart. Selain itu, confusion matrix digunakan untuk evaluasi klasifikasi model, dan heatmap dimanfaatkan untuk menunjukkan relasi antar aspek. Visualisasi ini membantu memberikan pemahaman lebih dalam terhadap distribusi sentimen dan topik yang dominan dalam ulasan.

4. PEMBAHASAN

4.1 Pengumpulan Data

Data penelitian diperoleh melalui teknik web scraping terhadap ulasan pengguna aplikasi Mobile JKN di Google Play Store, menggunakan pustaka Python google-play-scraper. Data dikumpulkan dalam rentang Oktober 2021 hingga Juni 2025, dengan parameter seperti app_id, bahasa Indonesia, dan wilayah Indonesia. Hasil scraping awal menghasilkan 55.688 ulasan dalam format CSV, berisi informasi seperti nama pengguna, isi ulasan, skor (1–5), tanggal, dan versi aplikasi. Untuk menjaga kualitas data, ulasan yang terdiri dari kurang dari lima kata dihapus karena dianggap tidak informatif. Penyaringan ini mengacu pada prinsip bahwa ulasan singkat cenderung kurang kontekstual dan bisa menimbulkan noise dalam analisis sentimen.

4.2 Preprocessing Data

Tahap pra-pemrosesan dilakukan untuk meningkatkan kualitas data sebelum analisis lebih lanjut. Beberapa tahapan yang dilakukan meliputi:

- 1) *Cleaning* – Menghapus karakter khusus, angka, emoji, URL, serta tanda baca yang tidak relevan.
- 2) *Case Folding* - Mengubah seluruh teks menjadi huruf kecil agar seragam dalam analisis.
- 3) *Normalisasi* - Mengubah kata-kata tidak baku atau informal menjadi bentuk standar menggunakan kamus normalisasi yang diperoleh dari platform Kaggle.
- 4) *Tokenizing* – Memecah kalimat menjadi kata-kata (token) individual untuk mempermudah proses analisis berikutnya.
- 5) *Stopword Removal* – Menghilangkan kata-kata umum dalam Bahasa Indonesia yang tidak memiliki nilai semantik tinggi, seperti “yang”, “dan”, “di”.
- 6) *Lemmatization* – Mengembalikan kata ke bentuk dasarnya menggunakan lemmatizer Bahasa Indonesia guna mengurangi variasi kata.

Hasil dari tahap ini berupa data ulasan yang bersih dan siap digunakan untuk proses pemodelan topik dan analisis sentimen.

4.3 Pemodelan Topik dan Aspek

Pemodelan topik dilakukan menggunakan algoritma Latent Dirichlet Allocation (LDA) untuk mengelompokkan ulasan pengguna aplikasi Mobile JKN. Proses ini terdiri atas beberapa tahapan utama sebagaimana digambarkan pada Gambar 3.



Gambar 3. Alur Topic Modelling

4.3.1 Bag of Words (BoW)

Tahapan awal melibatkan pembentukan representasi vektor kata menggunakan pendekatan Bag of Words melalui pustaka Gensim. Vektor ini mewakili frekuensi kata berdasarkan kamus (dictionary) yang dihasilkan dari data ulasan, dan menjadi input utama bagi model LDA. Di bagian Bag of Words ini, menggunakan input data ulasan yang sudah di lemmatization. Kemudian, data tersebut diubah menjadi kamus. Setelah membuat kamus, langkah selanjutnya adalah membuat Bag of Words dari korpus. Hasil dari proses ini adalah data yang ditampung pada variabel id2word dan bow_corpus.

Dokumen	Tokenisasi	Bag of Words
aplikasi mobile jkn sangat bantu akses layanan sehat proses daftar cari fasilitas jadwal kontrol mudah cepat bantu fitur notifikasi jadwal	aplikasi mobile jkn bantu akses layanan sehat proses daftar cari fasilitas jadwal kontrol mudah cepat bantu fitur notifikasi jadwal	[('akses', 1), ('aplikasi', 1), ('bantu', 2), ('cari', 1), ('cepat', 1), ('daftar', 1), ('fasilitas', 1), ('fitur', 1), ('jadwal', 2), ('jkn', 1), ('kontrol', 1), ('layanan', 1), ('mobile', 1), ('mudah', 1), ('notifikasi', 1), ('proses', 1), ('sehat', 1)]
sangat kecewa aplikasi mobile jkn error login verifikasi	kecewa aplikasi mobile jkn error login verifikasi	[('aplikasi', 1), ('jkn', 1), ('mobile', 1), ('error', 1), ('kecewa', 1), ('login', 1), ('verifikasi', 1)]
alam aplikasi mobile jkn sangat positif layanan sedia modern cepat percaya	alam aplikasi mobile jkn positif layanan sedia modern cepat percaya	[('aplikasi', 1), ('cepat', 1), ('jkn', 1), ('layanan', 1), ('mobile', 1), ('alam', 1), ('modern', 1), ('percaya', 1), ('positif', 1), ('sedia', 1)]
antarmuka aplikasi bingung tidak responsif	antarmuka aplikasi bingung tidak responsif	[('aplikasi', 1), ('antarmuka', 1), ('bingung', 1), ('responsif', 1)]
aplikasi standar sistem verifikasi login sangat tidak stabil	aplikasi standar sistem verifikasi login stabil	[('aplikasi', 1), ('login', 1), ('verifikasi', 1), ('sistem', 1), ('stabil', 1), ('standar', 1)]

Gambar 4. Kamus kata dari Bag of Words

4.3.2 Latent Dirichlet Allocation Model

Pada tahap ini, setelah dilakukan proses pembobotan menggunakan metode Bag of Words (BoW), selanjutnya dibangun model Latent Dirichlet Allocation (LDA) untuk mengekstraksi topik-topik yang muncul dalam ulasan pengguna aplikasi Mobile JKN. Model LDA ini dievaluasi menggunakan coherence score guna menilai sejauh mana topik yang dihasilkan bersifat koheren dan dapat diinterpretasikan.

Dalam proses ini, coherence score digunakan untuk memilih jumlah topik optimal berdasarkan keterpahaman topik oleh manusia. Model dievaluasi dengan batasan jumlah topik maksimum (limit) sebanyak lima, minimum (start) dua, dan penambahan jumlah topik (step) satu per iterasi. Berikut adalah hasil nilai coherence score untuk masing-masing jumlah topik.

Tabel 1. Hasil *Coherence Value* Pada Setiap Topik

Jumlah Topik	Nilai Coherence
2	0.533
3	0.593
4	0.582
5	0.595
6	0.600

4.3.3 Penerapan Gen AI Penentuan Aspek

Setelah topik diperoleh melalui algoritma LDA, setiap kelompok kata kunci belum secara langsung menunjukkan makna atau tema spesifik. Oleh karena itu, digunakan Gen AI berbasis Google Gemini versi *gemini-2.0-flash* untuk membantu menginterpretasikan setiap kumpulan kata menjadi satu aspek utama yang paling relevan. Dalam proses penentuan aspek dari hasil pemodelan topik menggunakan metode *Latent Dirichlet Allocation* (LDA), setiap topik diwakili oleh sejumlah kata kunci yang memiliki probabilitas tertinggi dalam topik tersebut. Umumnya, jumlah kata yang digunakan untuk merepresentasikan satu topik berkisar antara 10 hingga 15 kata.

Tabel 2. Uraian kata per topik pada setiap aspek

Topik	Aspek	Kata Per Topik
0	Proses Login dan Perangkat	'login', 'update', 'buka', 'captcha', 'emulator', 'nik', 'device', 'salah'
1	Akses Aplikasi Mobile JKN	'mudah', 'bantu', 'layanan', 'jkn', 'mobile', 'bpjs', 'cepat', 'akses', 'lambat'
2	Perubahan Data BPJS	'faskes', 'ubah', 'data', 'bayar', 'pindah', 'rubah', 'alamat', 'bpjs', 'kantor'
3	Kendala Penggunaan Aplikasi	'aplikasi', 'susah', 'sulit', 'ribet', 'bikin', 'error', 'pakai', 'daftar'

4	Registrasi dan Verifikasi Awal	'daftar', 'verifikasi', 'otp', 'kode', 'nomor', 'kirim', 'sms', 'masuk'
5	Pendaftaran Online dan Antrian	'online', 'antrian', 'sakit', 'rujuk', 'dokter', 'rumah', 'faskes', 'kartu', 'kontrol'

4.3.4 Penentuan Dominan Aspek Per Topik

Dataset dibagi ke dalam masing-masing topik yang telah ditentukan sebelumnya. Hal ini dilakukan dengan menambahkan interpretasi topik berdasarkan nilai dominan topik yang diperoleh. Proses interpretasi topik berdasarkan nilai dominan topik yang diperoleh dilakukan dengan cara mengamati kata-kata kunci (*keywords*) yang dihasilkan oleh model LDA pada setiap topik. Setelah model LDA mengelompokkan kalimat atau ulasan ke dalam beberapa topik dan memberikan proporsi kontribusi dari masing-masing topik terhadap ulasan tersebut, langkah selanjutnya adalah menentukan topik dominan, yaitu topik dengan kontribusi terbesar pada setiap kalimat.

Dominant_Topic	Perc_Contrib	Topic_Keywords	Text	Aspect
0	1	0.583 sangat, mudah, bantu, jkn, lambat, layan, kasi...	aplikasi mobile jkn sangat bantu akses layan s...	Akses Aplikasi Mobile JKN
1	3	0.360 aplikasi, tidak, susah, baik, pakai, banget, d...	sangat kecewa aplikasi mobile jkn error login ...	Kendala Penggunaan Aplikasi
2	1	0.674 sangat, mudah, bantu, jkn, lambat, layan, kasi...	alam aplikasi mobile jkn sangat positif layan ...	Akses Aplikasi Mobile JKN
3	3	0.586 aplikasi, tidak, susah, baik, pakai, banget, d...	antarmuka aplikasi bingung tidak responsif	Kendala Penggunaan Aplikasi
4	3	0.424 aplikasi, tidak, susah, baik, pakai, banget, d...	aplikasi standar sistem verifikasi login sanga...	Kendala Penggunaan Aplikasi

Gambar 5. Hasil Penentuan Dominan Topik Per Kalimat

4.4 Pelabelan Sentimen

Pada tahap pelabelan sentimen ini, tujuannya adalah untuk memberi label pada dataset yang telah diproses sebelumnya dengan sentimen baik itu positif maupun negatif. Pelabelan dilakukan menggunakan metode IndoBERT. Berikut adalah tahapan pelabelan sentimen.

Model *IndoBERT* awalnya melalui tahap *pre-training*, yaitu pelatihan awal untuk memahami struktur dan pola umum dalam bahasa Indonesia menggunakan data besar seperti Wikipedia dan berita. Setelah itu, dilakukan *fine-tuning*, yaitu pelatihan lanjutan untuk menyesuaikan model agar mampu menyelesaikan tugas tertentu, dalam hal ini adalah klasifikasi sentimen. Model yang digunakan dalam penelitian ini adalah *mdhugol/indonesia-bert-sentiment-classification*, yang sudah di-fine-tune secara khusus untuk membedakan sentimen positif dan negatif dalam teks ulasan.

Tabel 3. Hasil labelling data menggunakan model *IndoBERT*

Teks	Sentimen	Score
sangat kecewa aplikasi mobile jkn error login verifikasi	Negative	0.99692
antarmuka aplikasi bingung tidak responsif	Negative	0.99714
sangat bantu terima kasih developer	Positive	0.95076

4.5 Evaluasi Model Klasifikasi Berbasis BERT

Penelitian ini menggunakan tiga model BERT berbahasa Indonesia untuk klasifikasi sentimen, yaitu IndoBERT, RoBERTa, dan DistilBERT. Ketiga model dilatih menggunakan parameter yang telah disesuaikan, seperti learning rate, batch size, dan jumlah epoch, guna menghasilkan kinerja terbaik dalam klasifikasi ulasan pengguna aplikasi Mobile JKN.

Model IndoBERT (*indobenchmark/IndoBERT-base-p1*) dilatih menggunakan pendekatan fine-tuning dengan parameter learning rate $2e-5$, batch size 16, dan sequence length 128. Optimizer AdamW dan fungsi loss CrossEntropy digunakan untuk mengoptimalkan proses pelatihan selama 3 epoch. Model ini menunjukkan performa tinggi dalam mendeteksi sentimen positif dan negatif secara seimbang.

Sementara itu, model RoBERTa (*w11wo/indonesian-RoBERTa-base-sentiment-classifier*) juga diuji dalam tugas klasifikasi yang sama. Model ini menggunakan tokenizer dan arsitektur

Transformer dari pustaka HuggingFace dan menunjukkan hasil evaluasi yang cukup kompetitif, meskipun sedikit lebih rendah pada sentimen positif dibandingkan IndoBERT.

Model ketiga, DistilBERT (cahya/DistilBERT-base-indonesian), digunakan sebagai model ringan yang lebih efisien secara komputasi. Model ini tetap mampu menghasilkan performa klasifikasi yang baik, mendekati hasil model BERT lainnya.

Hasil evaluasi masing-masing model ditampilkan dalam Tabel 4 berikut.

Tabel 4. Evaluasi Klasifikasi Sentimen untuk Masing-Masing Model

Model	Kelas Sentimen	Precision	Recall	F1-Score	Support
IndoBERT	Negatif	0.97	0.97	0.97	7,431
	Positif	0.93	0.94	0.94	3,628
	Accuracy			0.96	11,059
RoBERTa	Negatif	0.95	0.97	0.96	7,431
	Positif	0.93	0.91	0.92	3,628
	Accuracy			0.95	11,059
DistilBERT	Negatif	0.96	0.97	0.97	7,431
	Positif	0.94	0.92	0.93	3,628
	Accuracy			0.96	11,059

Model IndoBERT menunjukkan akurasi klasifikasi aspek tertinggi sebesar 91%, dengan performa stabil pada sebagian besar kategori. Sementara itu, model DistilBERT unggul pada aspek “Akses Aplikasi Mobile JKN” namun mengalami penurunan pada aspek “Perubahan Data BPJS”. RoBERTa memiliki performa yang seimbang tetapi cenderung lebih rendah dalam menangani ulasan pada aspek “Pendaftaran dan Perubahan Data”.

Tabel 5. Evaluasi Klasifikasi Aspek pada Masing-Masing Model

Kategori Aspek	Model	Precision	Recall	F1-Score	Support
Akses Aplikasi Mobile JKN	IndoBERT	0.92	0.97	0.94	3,295
	RoBERTa	0.90	0.95	0.93	3,295
	DistilBERT	0.94	0.95	0.95	3,295
Kendala Penggunaan Aplikasi	IndoBERT	0.92	0.92	0.92	4,331
	RoBERTa	0.91	0.90	0.90	4,331
	DistilBERT	0.93	0.89	0.91	4,331
Pendaftaran Online dan Antrian	IndoBERT	0.84	0.43	0.57	226
	RoBERTa	0.77	0.57	0.65	226
	DistilBERT	0.63	0.69	0.66	226
Perubahan Data BPJS	IndoBERT	0.84	0.50	0.62	248
	RoBERTa	0.66	0.67	0.67	248
	DistilBERT	0.77	0.58	0.66	248
Proses Login dan Perangkat	IndoBERT	0.80	0.66	0.72	383
	RoBERTa	0.73	0.68	0.70	383
	DistilBERT	0.69	0.75	0.71	383
Registrasi dan Verifikasi Awal	IndoBERT	0.90	0.93	0.92	2,576
	RoBERTa	0.91	0.90	0.91	2,576
	DistilBERT	0.89	0.93	0.91	2,576
Akurasi Total	IndoBERT	–	–	0.91	11,059
	RoBERTa	–	–	0.89	11,059
	DistilBERT	–	–	0.90	11,059

4.6 Visualisasi Data

Visualisasi Word Cloud ini menampilkan kata-kata yang paling sering muncul dalam ulasan atau komentar dengan sentimen positif. Kata-kata tersebut mencerminkan aspek-aspek yang diapresiasi oleh pengguna, seperti kemudahan, kecepatan, dan kualitas layanan atau produk. Berikut gambar visualisasi word cloud.



Gambar 6. Word cloud Kata Positif Yang Sering Muncul

Pada Word Cloud ini, ditampilkan kata-kata yang paling sering muncul dalam ulasan atau komentar dengan sentimen negatif. Kata-kata tersebut menunjukkan keluhan, masalah, atau hambatan yang dialami oleh pengguna selama menggunakan layanan atau produk. Berikut gambar visualisasi word cloud.



Gambar 7. Word cloud Kata Negatif Yang Sering Muncul

5. KESIMPULAN

Penelitian ini dilakukan untuk menganalisis opini pengguna terhadap aplikasi Mobile JKN menggunakan pendekatan Aspect-Based Sentiment Analysis (ABSA) berbasis Natural Language Processing (NLP). Data sebanyak 55.688 ulasan dikumpulkan melalui teknik web scraping dari Google Play Store. Tahapan preprocessing dilakukan mencakup pembersihan data, normalisasi, tokenisasi, penghapusan stopword, hingga lemmatization. Proses identifikasi aspek menggunakan algoritma LDA menghasilkan enam topik utama yang kemudian diberi label aspek melalui bantuan Gemini 2.0 Flash dengan teknik prompt engineering. Enam aspek yang dihasilkan meliputi Akses Aplikasi Mobile JKN, Kendala Penggunaan Aplikasi, Registrasi dan Verifikasi Awal, Proses Login dan Perangkat, Perubahan Data BPJS, serta Pendaftaran Online dan Antrian.

Sentimen pengguna dilabeli menggunakan model IndoBERT yang telah dilatih pada korpus Bahasa Indonesia. Hasil klasifikasi menunjukkan 18.136 ulasan bersentimen positif dan 37.153 ulasan negatif. Evaluasi kinerja model klasifikasi multitugas menggunakan IndoBERT, RoBERTa, dan DistilBERT menunjukkan bahwa IndoBERT memiliki performa terbaik dengan akurasi 96% untuk klasifikasi sentimen (F1-score 0.97 negatif dan 0.94 positif), serta akurasi 91% untuk klasifikasi aspek. F1-score tertinggi terdapat pada aspek Akses Aplikasi dan Kendala Penggunaan, sedangkan aspek Pendaftaran Online dan Perubahan Data BPJS memperoleh skor

lebih rendah akibat ketidakseimbangan jumlah data. Hasil ini membuktikan bahwa pendekatan ABSA efektif dalam menangkap sentimen dan isu utama yang dirasakan pengguna terhadap layanan aplikasi Mobile JKN.

DAFTAR PUSTAKA

- [1] W. K. Suhet and A. Nugroho, "Analisis Sentimen Penilaian Masyarakat Indonesia Terhadap Konferensi G20 Di Bali Dengan Menggunakan Metode Naive Bayes," *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 3, pp. 904–912, 2023, doi: 10.35870/jimik.v4i3.290.
- [2] D. Arianto and I. Budi, "Aspect-based Sentiment Analysis on Indonesia's Tourism Destinations Based on Google Maps User Code-Mixed Reviews (Study Case: Borobudur and Prambanan Temples)," *Proc. 34th Pacific Asia Conf. Lang. Inf. Comput.*, vol. 2019, no. PACLIC, pp. 359–367, 2020, [Online]. Available: <https://aclanthology.org/2020.paclic-1.41>
- [3] M. A. N. Febriansyach, F. Rashif, G. I. P. Nirvana, and N. A. Rakhmawati, "Implementasi LDA untuk Pengelompokan Topik Tweet Akun Bot Twitter bertagar #covid-19," *CogITO Smart J.*, vol. 7, no. 1, pp. 170–181, 2021, doi: 10.31154/cogito.v7i1.299.170-181.
- [4] Y. Sahria and D. Hatta Fudholi, "Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode Topic Modeling LDA (Latent Dirichlet Allocation)," *Masa Berlaku Mulai*, vol. 1, no. 3, pp. 336–344, 2017.
- [5] R. Z. Suchrady and A. Purwarianti, "Indo LEGO-ABSA: A Multitask Generative Aspect Based Sentiment Analysis for Indonesian Language," *Proc. Int. Conf. Electr. Eng. Informatics*, 2023, doi: 10.1109/ICEEI59426.2023.10346852.
- [6] V. H. Pranatawijaya, N. N. K. Sari, R. A. Rahman, E. Christian, and S. Geges, "Unveiling User Sentiment: Aspect-Based Analysis and Topic Modeling of Ride-Hailing and Google Play App Reviews," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 10, no. 3, pp. 328–339, 2024, doi: 10.20473/jisebi.10.3.328-339.
- [7] A. Bahari and K. E. Dewi, "Peringkasan Teks Otomatis Abstraktif Menggunakan Transformer Pada Teks Bahasa Indonesia," *Komputa J. Ilm. Komput. dan Inform.*, vol. 13, no. 1, pp. 83–91, 2024, doi: 10.34010/komputa.v13i1.11197.
- [8] U. A. A. Al-Faruq, "Implementasi Arsitektur Transformer pada Image Captioning dengan Bahasa Indonesia," *Automata*, 2021.
- [9] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," *Ilk. J. Ilm.*, vol. 14, no. 3, pp. 348–354, 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.
- [10] N. A. R. Putri and Ardiansyah, "Analisis Sentimen Terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa," *J. Sains dan Inform.*, vol. 9, no. 2, pp. 136–145, 2023, doi: 10.34128/jsi.v9i2.649.
- [11] Ade Chandra Saputra, Agus Sehatman Saragih, and Deddy Ronaldo, "Perbandingan Nilai Akurasi Distil Bert Dan Bert Pada Dataset Analisis Sentimen Lembaga Kursus," <https://e-journal.upr.ac.id/index.php/JTI>, vol. Vol 18 No 2, no. 2, pp. 160–171, 2024, doi: <https://doi.org/10.47111/JTI>.
- [12] Erlin, Josef Sianturi, Alyauma Hajjah, and Agustin, "Analisis Sentimen Prosesor AMD Ryzen menggunakan Metode Support Vector Machine," *SATIN - Sains dan Teknol. Inf.*, vol. 7, no. 2, pp. 129–141, 2021, doi: 10.33372/stn.v7i2.804.