

Aspect-Based Sentiment Analysis Ulasan Pengguna Aplikasi Polri Presisi

Noviyanti Santoso^{1*}, Ahmad Abror¹

Program Studi D4 Statistika Bisnis, Fakultas Vokasi, Institut Teknologi Sepuluh Nopember, Jawa Timur, Indonesia

*Email: noviyanti.santoso@its.ac.id

Info Artikel

Kata Kunci :

aspect-based sentiment analysis, indoBERT, latent dirichlet allocation (LDA), long short-term memory (LSTM), net reputation score, polri presisi

Keywords :

aspect-based sentiment analysis, indoBERT, latent dirichlet allocation (LDA), long short-term memory (LSTM), net reputation score, polri presisi

Tanggal Artikel

Dikirim : 29 Januari 2026

Direvisi : 12 Juni 2026

Diterima : 21 Oktober 2026

Abstrak

Pemanfaatan aplikasi *mobile* dalam layanan publik memerlukan evaluasi kualitas layanan berbasis persepsi pengguna. Penelitian terdahulu umumnya melakukan klasifikasi sentiment secara umum, atau memisahkan identifikasi aspek dan klasifikasi sentiment tanpa menghubungkannya dengan metrik prioritas perbaikan yang terukur. Penelitian ini bertujuan menjawab celah tersebut dengan mengusulkan *pipeline Aspect-Based Sentiment Analysis* (ABSA) dengan melakukan integrasi antara *Latent Dirichlet Allocation* (LDA), *IndoBERT*, *Net Reputation Score* (NRS) dan *Long Short-Term Memory* (LSTM) pada Aplikasi Polri Presisi. Data berupa 2640 ulasan pengguna versi 2.0.27 dikumpulkan dari *Google Play Store* dan menjadi 2458 ulasan setelah *text preprocessing* dengan pustaka *Sastrawi* dan kamus normalisasi *slang* Bahasa Indonesia. Identifikasi aspek menggunakan LDA dengan koherensi *C_v* menghasilkan empat aspek utama, yaitu *Pengurusan SKCK*, *Efisiensi Pendaftaran*, *Isu Teknis*, dan *Kepuasan Layanan*. *IndoBERT* berfungsi sebagai *penganotasi sentiment otomatis* yang divalidasi dengan *rating pengguna*. *Prioritas perbaikan* ditentukan menggunakan *NRS*. Aspek *Isu Teknik* memperoleh *NRS -19,84%* dan menjadi *prioritas utama*, diikuti *Efisiensi Pendaftaran (51,51%)*, *Pengurusan SKCK (75,78%)* dan *Kepuasan Layanan (77,57%)*. *Klasifikasi topik dan sentimen* dievaluasi melalui *LSTM* dengan pembagian data terstratifikasi 80:20, menghasilkan akurasi 85,07 untuk topik dan 86,91% untuk sentiment. Kontribusi penelitian ini meliputi *pipeline ABSA terpadu* yang mengintegrasikan pemodelan topik, klasifikasi sentimen berbasis *transformer*, evaluasi *deep learning*, dan metrik prioritas *NRS*, penerapan pada domain layanan publik kepolisian digital yang masih terbatas kajiannya, serta pemberian dasar kuantitatif bagi pengelola aplikasi pemerintah dalam mengambil Keputusan perbaikan.

Abstract

The use of mobile applications in public services requires evaluation of service quality based on user perception. Previous studies generally perform sentiment classification in a general manner, or separate aspects and sentiment classification without linking them to the measured improvement priority metrics. This study aims to address this gap by proposing an Aspect-Based Sentiment Analysis (ABSA) pipeline by integrating Latent Dirichlet Allocation (LDA), IndoBERT, Net Reputation Score (NRS), and Long Short-Term Memory (LSTM) in the Presisi Police Application. Data in the form of 2,640 user reviews of version 2.0.27 were collected from the Google Play Store and reduced to 2,458 reviews after text preprocessing with the Sastrawi library and the Indonesian slang dictionary. Aspect identification using LDA with C_v coherence resulted in four main aspects: SKCK Management, Registration Efficiency, Technical Issues, and Service Satisfaction. IndoBERT functions as an automatic sentiment annotator validated with user ratings. Improvement priorities are determined using NRS. The Technical Issues aspect obtained an NRS of -19.84% and became the top priority, followed by Registration Efficiency (51.51%), SKCK Management (75.78%) and Service Satisfaction

(77.57%). Topic and sentiment classification were evaluated through LSTM with stratified data splitting 80:20, resulting in an accuracy of 85.07 for topics and 86.91% for sentiment. The contribution of this research includes an integrated ABSA pipeline that integrates topic modeling, transformer-based sentiment classification, deep learning evaluation, and NRS priority metrics, application to the digital police public service domain which is still limited in its studies and providing a quantitative basis for government application managers in making improvement decisions.

1. PENDAHULUAN

Era digital telah mengubah cara masyarakat berinteraksi dengan sektor publik, mendorong pemanfaatan teknologi informasi dan komunikasi sebagai sarana peningkatan kualitas dan efisiensi pelayanan. Aplikasi mobile menjadi salah satu strategi utama pemerintah dalam memperluas jangkauan layanan serta meningkatkan kemudahan akses bagi masyarakat [1]. Dalam konteks Indonesia, Kepolisian Negara Republik Indonesia mengembangkan aplikasi Polri Presisi sebagai platform layanan digital terpadu yang menyediakan berbagai layanan kepolisian, termasuk pengurusan Surat Keterangan Catatan Kepolisian (SKCK), perpanjangan SIM, e-Tilang, layanan pengaduan PROPAM dan DUMAS, serta layanan panggilan darurat 110. Aplikasi ini dirancang untuk mendukung pelayanan publik yang lebih responsif, transparan, dan mudah diakses sesuai dengan visi Presisi (Prediktif, Responsibilitas, Transparansi Berkeadilan).

Keberhasilan implementasi aplikasi layanan publik sangat bergantung pada pengalaman dan penerimaan pengguna. Ulasan pengguna yang dipublikasikan pada *Google Play Store* merepresentasikan tingkat kepuasan, keluhan, serta harapan pengguna terhadap layanan yang diberikan [2]. Beragamnya opini yang muncul menunjukkan bahwa meskipun Aplikasi Polri Presisi memberikan manfaat, masih terdapat aspek-aspek layanan yang memerlukan evaluasi dan perbaikan. Akan tetapi, jumlah ulasan yang besar dan bersifat tidak terstruktur menyebabkan evaluasi secara manual menjadi tidak efektif dan berpotensi bias, sehingga diperlukan pendekatan otomatis berbasis *Natural Language Processing* (NLP) [3].

Sejumlah penelitian terdahulu telah menerapkan analisis sentiment pada ulasan aplikasi layanan public berbasis *Goggle Play Store*, namun masing-masing keterbatasan yang relevan dengan penelitian ini. Fatmasari dkk. [1] dan Fristtikasari dkk. [2] menerapkan *Naive Bayes* pada ulasan aplikasi Universitas Terbuka dan Kitalulus, sementara Rizaldi dan Alam [3] menerapkan pendekatan serupa pada aplikasi JMO. Ketiganya hanya menghasilkan klasifikasi sentimen tingkat dokumen tanpa mengaitkan sentiment dengan aspek spesifik layanan, sehingga *actionable insight* untuk perbaikan aplikasi terbatas. Poria dkk. [4] secara teoritis menegaskan bahwa ABSA membutuhkan integrasi antara identifikasi aspek dan klasifikasi sentimen yang *context-aware*, namun penerapan integrative di ranah layanan publik digital indonesai masih terbatas. Hidayat dan Gata [5] mengombinasikan IndoBERT dan LSTM pada ulasan aplikasi Depok *Single Window* dengan akurasi 97%, tetapi tidak melakukan identifikasi aspek dan tidak menyediakan metrik prioritas perbaikan. Hammami [6] menerapkan ABSA pada aplikasi Identitas Kependudukan Digital (IKD) dengan kombinasi LDA, IndoBERT, dan CNN serta memperoleh akurasi 84% (aspek) dan 83% (sentimen). Meskipun telah menggunakan NRS untuk prioritas perbaikan, penelitian tersebut menggunakan CNN sebagai pengklasifikasi sehingga belum mengeksplorasi potensi arsitektur LSTM, khususnya dalam menangkap dependensi sekuensial yang lazim terdapat dalam teks ulasan. Tyas dkk. [7] menunjukkan keunggulan model *transformer-based* seperti BERT (akurasi 81%) disbanding *Logistic Regression*, *LSTM* dan *Graph Convolutional Network* pada klasifikasi berita saham, namun konteksnya bukan ulasan layanan publik.

Berdasarkan hasil penelusuran literatur, terdapat tiga kesenjangan penelitian (*research gap*) yang masih terbuka. Pertama, sejauh ini belum ada penelitian yang mengintegrasikan topic modeling (LDA), klasifikasi sentimen *context-aware* berbahasa Indonesia (IndoBERT), metrik prioritas perbaikan (NRS), dan evaluasi performa *deep learning* (LSTM) dalam satu *pipeline* analisis layanan publik digital. Kedua, kajian tentang ABSA yang secara khusus menasar pada aplikasi kepolisian digital Indonesia belum banyak ditemukan. Ketiga kuantifikasi prioritas perbaikan layanan publik berbasis sentimen masih jarang dilakukan, sehingga rekomendasi perbaikan cenderung deskriptif dan kurang terukur.

Sebagai Upaya untuk mengisi kekosongan tersebut, kebaruan (*novelty*) penelitian ini dapat disoroti pada tiga dimensi utama. Dimensi pertama bersifat metodologis, yakni pengusulan *pipeline* yang menggabungkan LDA, IndoBERT, NRS dan LSTM secara berkesinambungan. LDA digunakan untuk identifikasi aspek tanpa pelabelan awal, IndoBERT untuk pelabelan sentimen kontekstual berbahasa Indonesia, LSTM untuk evaluasi performa klasifikasi

terhadap label aspek dan sentimen yang dihasilkan, dan NRS untuk menerjemahkan distribusi sentimen menjadi prioritas perbaikan yang terukur. Dimensi edua, pada objek penelitian, yakni Aplikasi Polri Presisi sebagai layanan digital kepolisian Indonesia yang belum banyak dikaji secara ilmiah. Dimensi ketiga menyentuh aspek terapan, berupa peringkat prioritas perbaikan aplikasi yang dapat langsung dimanfaatkan oleh pengelola aplikasi pemerintah.

Berdasarkan latar belakang dan *research gap* tersebut, penelitian ini diarahkan untuk menjawab tiga pertanyaan penelitian berikut.

1. Aspek-aspek layanan utama apa yang dapat diidentifikasi dari ulasan pengguna Aplikasi Polri Presisi menggunakan *topic modeling* LDA?
2. Bagaimana distribusi sentimen pengguna pada masing-masing aspek setelah diklasifikasi menggunakan IndoBERT dan bagaimana prioritas perbaikan ditentukan berdasarkan NRS?
3. Bagaimana evaluasi performa LSTM dalam mengklasifikasikan aspek dan sentimen dari ulasan pengguna Aplikasi Polri Presisi?

Dari sisi akademik, penelitian ini memberikan kontribusi pada bidang *text mining* dan NLP bahasa Indonesia, khususnya dalam pengembangan *pipeline* ABSA yang menghubungkan *topic modeling* dengan *transformer-based sentiment classification* dan metrik reputasi layanan. Secara praktis, hasil penelitian memberikan dasar pengambilan keputusan berbasis data bagi Polri untuk memprioritaskan perbaikan aplikasi yang memiliki dampak terbesar terhadap persepsi masyarakat. Sisa makalah ini disusun sebagai berikut: Bagian 2 memaparkan metode penelitian, Bagian 3 menyajikan hasil dan pembahasan, dan Bagian 4 memuat kesimpulan, keterbatasan, serta arah penelitian lanjutan.

2. METODE PENELITIAN

Berikut adalah tahapan yang akan dilakukan dalam penelitian ini.

2.1 Sumber dan Etika Data

Data yang digunakan merupakan data primer berupa 2.460 ulasan pengguna Aplikasi Polri Presisi versi 2.0.27 yang diperoleh dari Google Play Store melalui proses *crawling* menggunakan *Google Colab*. Ulasan yang digunakan berada pada rentang waktu 14 Februari 2025 hingga 8 Desember 2025 dan seluruhnya berbahasa Indonesia. Setelah filter ulasan dengan jumlah kata lebih dari dua dan penghapusan ulasan kosong setelah *stopword removal*, diperoleh 2458 ulasan valid untuk analisis lanjutan.

Aspek etika penelitian dijaga dengan ketentuan sebagai berikut. Pertama, data ulasan bersifat publik dan dapat diakses secara terbuka melalui *Google Play Store*. Kedua, tidak ada informasi pengenalan pribadi (*personally identifiable information*) yang digunakan. Ketiga, pengolahan data semata untuk tujuan akademik dan yang terakhir hasil disajikan secara agregat tanpa mengaitkan ulasan tertentu dengan identitas pengguna.

2.2 Pipeline Analisis Terpadu

Penelitian ini mengusulkan *pipeline* ABSA terpadu yang menghubungkan empat komponen dengan relasi sebagai berikut. Secara lebih rinci, metode yang digunakan mencakup LDA [8] untuk mengekstraksi topik laten, IndoBERT [9] untuk mengklasifikasikan sentimen secara kontekstual, NRS [10] untuk mengukur reputasi per aspek, dan LSTM [11] untuk mengevaluasi performa klasifikasi. Arsitektur transformer [12] menjadi dasar dari model IndoBERT, sementara koherensi topik hasil LDA dievaluasi menggunakan *coherence score* [13]. Peningkatan performa klasifikasi dapat didukung oleh mekanisme atensi [14], namun dalam penelitian ini LSTM dilatih dari awal tanpa menggunakan embedding dari IndoBERT. Sebelum masuk ke tahap pemodelan, teks ulasan terlebih dahulu melalui tahapan *preprocessing* [15] untuk membersihkan dan menstandarisasi data. LSTM yang digunakan adalah *bidirectional* LSTM [16] agar mampu menangkap konteks dari dua arah. Adapun NRS mengacu pada konsep reputasi bersih [17] dengan representasi vektor kata yang dilatih dari awal menggunakan *embedding layer* [18][19]. Pemetaan peran tiap komponen disajikan pada Tabel 1.

Tabel 1. Peran, Input, dan Output Setiap Komponen Pipeline ABSA

Komponen	Peran	Input	Output
LDA	Identifikasi aspek (<i>unsupervised topic modeling</i>)	Token preprocessed (<i>bag-of-words</i>)	Empat label aspek (<i>dominant topic per dokumen</i>)

IndoBERT	Pelabelan sentimen otomatis	Teks ulasan <i>preprocessed</i>	Tiga label sentimen (positif, netral, negatif)
LSTM	Evaluasi performa klasifikasi aspek dan sentimen	<i>Integer indices</i> dari Keras <i>Tokenizer</i>	Klasifikasi 4 kelas topik atau 3 kelas sentimen (dua model terpisah)
NRS	Penilaian prioritas perbaikan layanan	Distribusi sentimen IndoBERT per aspek	Skor reputasi per aspek (%) dan urutan prioritas

2.3 Text Processing

Tahap preprocessing sangat penting bagi kualitas analisis sentimen pada teks ulasan yang bersifat informal [15]. Implementasi pada penelitian ini menggunakan pustaka re, *emoji*, NLTK, dan *Sastrawi* dengan tahapan:

- (1) normalisasi duplikasi karakter (mis. “bagusss” → “bagus”) menggunakan *regex* `(.)\{2,}`
- (2) penanganan emoji melalui *emoji.demojize()* diikuti penghapusan token emoji
- (3) penghapusan URL, mention, hashtag, angka, dan tanda baca dengan *regular expression*
- (4) *case folding* ke huruf kecil
- (5) normalisasi kata tidak baku menggunakan kamus slang bahasa Indonesia berisi lebih dari 70 entri khusus domain ulasan aplikasi (contoh: “gak/ga/gk” → “tidak”) dengan *word-boundary regex* agar penggantian aman dari *partial match*
- (6) *stopword removal* menggunakan *Sastrawi StopWordRemoverFactory* digabung dengan daftar tambahan domain-spesifik dari berkas *stopwordbahasa.txt*
- (7) stemming menggunakan *Sastrawi Stemmer (StemmerFactory)* untuk mengembalikan kata ke bentuk dasar bahasa Indonesia (mis. “mengurus” → “urus”).

Contoh hasil tahap *preprocessing* disajikan pada Tabel 2.

Tabel 2. Contoh Hasil Preprocessing Ulasan Pengguna

<i>Sebelum Preprocessing</i>	<i>Setelah Case Folding, Normalization, Removal</i>	<i>Setelah Stemming & Stopwords Removal</i>
tim it nya ayolah, masa hanya mengirim kode otp untuk verifikasi email saja lama nya minta ampun, ngurus apa apa susah kalo gini cara nya makan waktu	tim it nya ayolah masa hanya mengirim kode otp untuk verifikasi email saja lama nya minta ampun mengurus apa apa susah kalo gini cara nya makan waktu	tim ayo kirim kode otp verifikasi email ampun urus susah kalo gin makan
aplikasi gak berguna cuma nyusahin aja, setelah update gak bisa masuk lagi data yang tersimpan jadi hilang semua	aplikasi tidak berguna cuma nyusahin saja setelah update tidak bisa masuk lagi data yang tersimpan jadi hilang semua	aplikasi nyusahin update masuk data simpan hilang
kenapa susah sekali untuk foto ktp, selalu hasilnya blur padahal sudah benar fotonya	kenapa susah sekali untuk foto ktp selalu hasilnya blur padahal sudah benar fotonya	susah foto ktp hasil blur foto

Pada tahap *stopword removal*, terjadi pengurangan ulasan dari 2.640 menjadi 2.458 karena sejumlah ulasan menjadi kosong (seluruh katanya merupakan *stopword*).

2.4 Identifikasi Aspek Menggunakan LDA

Identifikasi aspek menggunakan *Latent Dirichlet Allocation* (LDA) [5] yang diimplementasikan dengan *Gensim LdaMulticore*. Parameter pemodelan dirinci pada Tabel 2. Pemilihan jumlah topik dilakukan melalui iterasi *k* pada rentang 2 sampai 10. Untuk setiap nilai *k*, dihitung *coherence score* tipe *C_v* menggunakan *CoherenceModel* karena tipe ini terbukti paling konsisten dengan penilaian manusia pada teks pendek seperti ulasan aplikasi [13]. Selain *coherence score*, evaluasi tumpang tindih antar topik juga dilakukan secara visual menggunakan *pyLDAvis* (visualisasi *intertopic distance map*). Parameter konfigurasi tercantum pada Tabel 3.

Tabel 3. Parameter Konfigurasi LDA

<i>Parameter</i>	<i>Nilai</i>
<i>num_topics</i>	2 – 10

<i>passes</i>	10
<i>workers</i>	2
<i>random_state</i>	42
<i>filter_extremes no_below</i>	10 (token muncul minimal di 10 dokumen)
<i>filter_extremes no_above</i>	0,5 (token muncul maksimum di 50% dokumen)
<i>Coherence type</i>	C_v (paling konsisten dengan penilaian manusia)
<i>Library</i>	Gensim (<i>LdaMulticore</i>) dan pyLDAvis (visualisasi)

2.5 Pelabelan Sentimen Menggunakan IndoBERT

Pelabelan sentimen dilakukan secara *semi-otomatis* menggunakan IndoBERT [6] *pre-trained model* *mdhugol/indonesia-bert-sentiment-classification* yang telah *fine-tuned* untuk klasifikasi sentimen bahasa Indonesia. Setiap ulasan *preprocessed* dilewatkan ke *HuggingFace Transformers pipeline* untuk memperoleh label sentimen (positif, netral, negatif). Pemilihan varian IndoBERT-BASE didasarkan pada pertimbangan efisiensi memori, meskipun IndoBERT-LARGE menawarkan akurasi yang sedikit lebih tinggi. Sebagai pemeriksaan kewajaran (*sanity check*), label IndoBERT dibandingkan dengan rating numerik 1–5 yang dikelompokkan menjadi negatif (1–2), netral (3), dan positif (4–5). Hasil perbandingan menunjukkan ulasan dengan rating tinggi tetapi isinya keluhan serta sebaliknya, sehingga label IndoBERT dianggap lebih merefleksikan isi tekstual dibandingkan rating. Pada kasus konflik label, *default* mengikuti label IndoBERT.

2.6 Penilaian Prioritas Perbaikan Menggunakan Net Reputation Score (NRS)

Net Reputation Score (NRS) adalah metrik kuantitatif yang mengukur reputasi suatu entitas berdasarkan selisih relatif antara ulasan positif dan negatif [8], [17]. NRS dipilih karena memberikan satu angka tunggal yang mudah dibandingkan antar aspek, dapat memberi bobot eksplisit pada selisih *positif* dan *negatif* sehingga lebih sensitif terhadap polarisasi ulasan dibandingkan sekadar persentase positif, serta mampu menerjemahkan distribusi sentimen menjadi peringkat prioritas perbaikan yang *actionable*. Rumus NRS yang digunakan dalam penelitian ini disajikan pada Persamaan (1).

$$NRS_a = \frac{P_a - N_a}{T_a} \quad (1)$$

dengan P_a adalah jumlah ulasan positif pada aspek a , N_a adalah jumlah ulasan negatif pada aspek a , dan T_a adalah total ulasan pada aspek a (positif + netral + negatif). Nilai NRS positif menandakan dominasi sentimen positif, nilai negatif menandakan dominasi sentimen negatif. Aspek dengan NRS terendah ditetapkan sebagai prioritas perbaikan utama.

2.7 Evaluasi Klasifikasi dengan LSTM

Untuk mengevaluasi sejauh mana label aspek (dari LDA) dan label sentimen (dari IndoBERT) dapat dipelajari oleh model *supervised*, dibangun dua model *LSTM* [7], [16] terpisah: satu untuk klasifikasi topik (4 kelas) dan satu untuk klasifikasi sentimen (3 kelas). *LSTM* tidak menerima *embedding* IndoBERT sebagai input, melainkan dilatih dari awal menggunakan *trainable embedding layer* Keras, sehingga *LSTM* berfungsi sebagai *validator* independen terhadap label yang dihasilkan *pipeline*.

Pembagian data dilakukan dengan *scikit-learn train_test_split* menggunakan rasio 80:20 (1.969 data latih + 489 data uji), dengan *random_state=42* dan *stratify* pada label kelas untuk menjaga proporsi antar kelas konsisten. Stratifikasi ini bertujuan mengurangi bias terhadap kelas mayoritas dan menjaga validitas evaluasi pada kelas minoritas (*tebutams sentimen netral*). *Tokenizer* Keras dilatih pada seluruh data sebelum *split* untuk membangun vokabuler, namun pelatihan model dan evaluasi dilakukan setelah *split* model tidak pernah melihat label data uji

selama pelatihan, sehingga bebas dari *data leakage*. *Embedding layer* Keras berperan sebagai representasi vektor kata yang dilatih dari awal [18], [19]. Arsitektur dan *hyperparameter* LSTM dirinci pada Tabel 4.

Tabel 4. Hyperparameter dan Arsitektur Model LSTM

<i>Parameter</i>	<i>Nilai</i>
<i>Framework</i>	TensorFlow / Keras (Python)
<i>Vocabulary size</i>	10.000 (<i>Tokenizer</i> dengan OOV token <unk>)
<i>Sequence length (maxlen)</i>	persentil ke-90 dari panjang token; <i>padding</i> ='post', <i>truncating</i> ='post'
<i>Embedding dimension</i>	128 (trainable embedding layer)
<i>LSTM unit</i>	64 unit
<i>LSTM dropout</i>	0,3 (dropout & recurrent_dropout)
<i>Dropout layer tambahan</i>	0,5 sebelum lapisan output
<i>Fungsi aktivasi output</i>	<i>softmax</i> (<i>multi-class</i>)
<i>Loss function</i>	<i>categorical_crossentropy</i>
<i>Optimizer</i>	Adam (<i>learning rate</i> default 1e-3)
<i>Batch size</i>	32
<i>Epoch</i>	14 (klasifikasi topik); 10 (klasifikasi sentimen)
<i>Early stopping</i>	Direkomendasikan pada epoch ke-7 s.d. ke-9 (topik) dan ke-4 s.d. ke-5 (sentimen)

Evaluasi kinerja model dilakukan menggunakan *confusion matrix* dan *classification report* (akurasi, *precision*, *recall*, *F1-score*) pada *data testing*.

3. HASIL DAN PEMBAHASAN

Bagian ini membahas hasil penerapan metode *aspect-based sentiment analysis* pada ulasan pengguna Aplikasi Polri Presisi. Analisis mencakup identifikasi aspek layanan, klasifikasi sentimen, penentuan prioritas perbaikan aplikasi berdasarkan *net reputation score*, serta evaluasi kinerja model.

3.1 Hasil Identifikasi Aspek Menggunakan LDA

Pemodelan LDA dilakukan dengan iterasi jumlah topik $k = 2$ sampai $k = 10$. Nilai *coherence score* C_v tertinggi diperoleh pada $k = 8$ sebesar 0,3984. Namun, evaluasi visual menggunakan *intertopic distance map* (LDavis) menunjukkan bahwa pada $k = 8$ terdapat beberapa topik yang saling berdekatan dengan distribusi kata mirip, yang mengindikasikan fragmentasi dan redundansi. Pemodelan dengan $k = 4$ menghasilkan *coherence score* 0,3341 lebih rendah dari $k = 8$ tetapi menunjukkan pemisahan topik yang lebih jelas pada LDavis dan interpretabilitas yang lebih tinggi. Berdasarkan trade-off antara *coherence* dan interpretabilitas, konfigurasi empat topik dipilih sebagai konfigurasi akhir. Ringkasan *coherence score* disajikan pada Tabel 5.

Tabel 5. Ringkasan Coherence Score (C_v) pada Variasi Jumlah Topik

<i>Jumlah Topik (k)</i>	<i>Coherence (C_v)</i>	<i>Keterangan</i>
2	Rendah	Topik tumpang tindih signifikan (<i>underfitting</i>)
4	0,3341	Pemisahan topik jelas (dipilih)
8	0,3984 (tertinggi)	<i>Coherence</i> puncak tetapi terdapat fragmentasi & redundansi topik
10	Menurun	Topik terfragmentasi



Tabel 6. *Top Words*, Interpretasi Topik, dan Contoh Ulasan Nyata (k = 4)

Persebaran ulasan ke dalam empat aspek adalah Efisiensi Pendaftaran 700 ulasan (28,6%), Isu Teknis 628 ulasan (25,7%), Pengurusan SKCK 573 ulasan (23,4%), dan Kepuasan Layanan 544 ulasan (22,2%). Dominasi Efisiensi Pendaftaran dan Isu Teknis menunjukkan bahwa proses awal interaksi pengguna dengan aplikasi (registrasi dan stabilitas sistem) menjadi pusat perhatian pengguna. Hasil ini menjawab rumusan masalah pertama, yaitu empat aspek layanan utama berhasil diidentifikasi dengan pemisahan visual yang baik pada LDAvis.

3.2.1 Hasil Klasifikasi Sentimen Menggunakan IndoBERT

Tabel 6. Distribusi Sentimen Pengguna per Aspek

282 | IJAI (Indonesian Journal of Applied Informatics)

1	Pengurusan SKCK	81,2%	13,4%	5,4%
2	Efisiensi Pendaftaran	71,8%	7,5%	20,7%
3	Isu Teknis	35,9%	9,3%	54,8%
4	Kepuasan Layanan	84,7%	8,1%	7,2%

Aspek Pengurusan SKCK (81,2%) dan Kepuasan Layanan (84,7%) didominasi sentimen positif, mengindikasikan persepsi pengguna yang positif terhadap manfaat dan proses administratif utama. Aspek Efisiensi Pendaftaran masih didominasi sentimen positif (71,8%), namun proporsi negatifnya (20,7%) cukup signifikan dan menunjuk pada kendala registrasi awal seperti kode OTP yang lambat dan verifikasi yang berulang konsisten dengan temuan analisis *N-gram* pada ulasan negatif yang memunculkan frasa seperti “skck online”, “daftar online”, dan “aplikasi lambat”. Aspek Isu Teknis menjadi pengecualian dengan sentimen negatif dominan (54,8%), frasa dominan pada *N-gram* ulasan negatif adalah “aplikasi error”, “buka aplikasi”, “masuk aplikasi”, dan “pakai wifi”, yang mengindikasikan gangguan saat menjalankan aplikasi serta kendala konektivitas. Pola ini sejalan dengan temuan Hidayat dan Gata [9] yang juga melaporkan isu teknis sebagai sumber utama keluhan pada aplikasi layanan publik digital di Indonesia.

3.2.2 Hasil Penilaian Prioritas Perbaikan Aplikasi

Net Reputation Score (NRS) digunakan dalam penelitian ini untuk menentukan prioritas perbaikan Aplikasi Polri Presisi berdasarkan tingkat kepuasan pengguna pada setiap aspek layanan. Perhitungan NRS dilakukan dengan mempertimbangkan proporsi ulasan positif, netral, dan negatif pada masing-masing topik, sehingga mampu memberikan gambaran reputasi layanan secara lebih komprehensif dan terukur. Nilai NRS untuk setiap aspek dihitung menggunakan Persamaan (1) dan disajikan pada Tabel 7.

Tabel 7. Net Reputation Score per Aspek dan Urutan Prioritas Perbaikan

Topik	Aspek	Positif	Netral	Negatif	NRS (%)	Prioritas
1	Pengurusan SKCK	466	77	31	75,78%	3
2	Efisiensi Pendaftaran	502	52	143	51,51%	2
3	Isu Teknis	223	59	348	-19,84%	1 (utama)
4	Kepuasan Layanan	461	44	39	77,57%	4

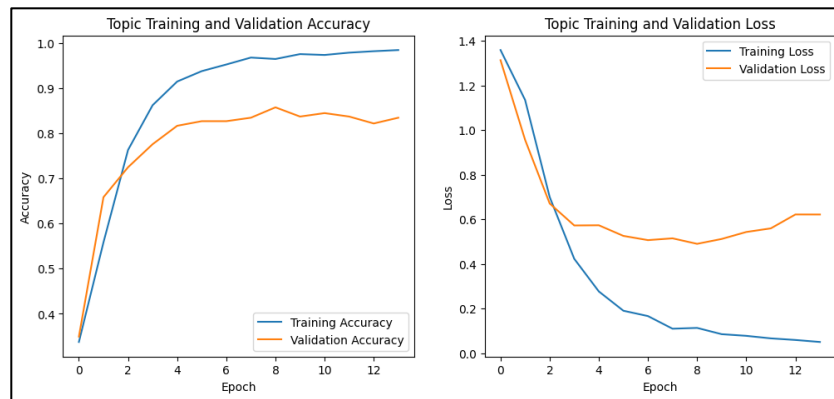
Aspek Isu Teknis memiliki NRS terendah (-19,84%) dan menjadi prioritas perbaikan utama. Nilai negatif menandakan bahwa ulasan negatif lebih banyak dibanding positif, sehingga aspek ini secara langsung menurunkan reputasi keseluruhan aplikasi. Dari sisi rekomendasi praktis yang operasional bagi pengelola aplikasi, perbaikan yang dapat ditindaklanjuti sebagai berikut. Pertama, penanganan *error* saat *login* dan verifikasi data yang menjadi keluhan paling sering muncul. Kedua, perbaikan mekanisme pengiriman OTP yang sering lambat (terbukti pada contoh ulasan “masa hanya mengirim kode otp untuk verifikasi email saja lama nya minta ampun”). Ketiga, pengujian *regression* setelah *update* untuk mencegah hilangnya data pengguna (terbukti pada contoh ulasan “setelah update gak bisa masuk lagi, data yang tersimpan jadi hilang semua”). Terakhir, optimasi kestabilan aplikasi untuk mengurangi *crash* saat dibuka. Aspek Efisiensi Pendaftaran (NRS = 51,51%) menjadi prioritas kedua karena masih memiliki keluhan signifikan pada proses verifikasi data dan *input* tanggal lahir. Aspek Pengurusan SKCK (NRS = 75,78%) dan Kepuasan Layanan (NRS = 77,57%) dapat dipertahankan sebagai *benchmark* pengembangan fitur lainnya.

3.3 Membangun Model LSTM

Pada tahap ini, dibangun model *Long Short-Term Memory* (LSTM) untuk melakukan klasifikasi topik dan sentimen ulasan pengguna Aplikasi Polri Presisi. Model dilatih menggunakan data pelatihan (*training set*) dan divalidasi menggunakan data validasi (*validation set*) untuk memantau kinerja model secara berkelanjutan selama proses pelatihan. Evaluasi performa model dilakukan menggunakan dua metrik utama, yaitu akurasi dan loss, baik pada data pelatihan maupun data validasi.

3.3.1 Pelatihan Model LSTM untuk Klasifikasi Topik

Model LSTM dilatih menggunakan data pelatihan dan divalidasi dengan data validasi selama 14 epoch. Evaluasi kinerja model dilakukan menggunakan dua metrik utama, yaitu akurasi dan *loss*, baik pada data *training* maupun validasi.

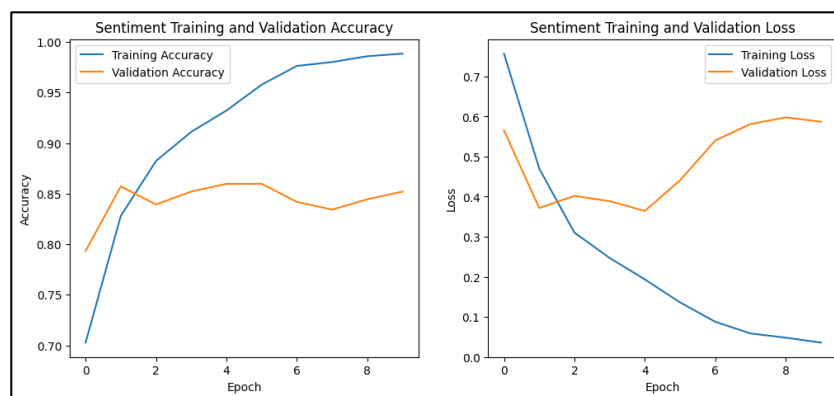


Gambar 2. Grafik Akurasi & Loss Model LSTM Pada Topik

Berdasarkan grafik pada Gambar 2, model *LSTM* untuk klasifikasi topik dilatih selama 14 *epoch*. Akurasi *training* meningkat tajam dari 29,07% (*epoch* 1) menjadi 86,48% (*epoch* 4), dan terus naik hingga 97–98% pada *epoch* 10–14. Akurasi validasi tertinggi sebesar 85,71% dicapai pada *epoch* 9 dan kemudian relatif stabil pada kisaran 82–84%. *Training loss* menurun konsisten dari 1,3760 menjadi 0,0478, sementara *validation loss* mencapai nilai terendah 0,4898 pada *epoch* 9 lalu meningkat hingga 0,6213 pada *epoch* 13–14, yang mengindikasikan kecenderungan *overfitting*. Karena itu, performa optimal berada pada *epoch* 7–9 di mana keseimbangan akurasi generalisasi masih terjaga.

3.3.2 Pelatihan Model LSTM untuk Klasifikasi Sentimen

Model LSTM untuk klasifikasi sentimen dilatih hingga 10 epoch. Evaluasi performa model dilakukan dengan mengamati akurasi dan loss pada data *training* serta data validasi untuk mengetahui kemampuan pembelajaran dan generalisasi model.



Gambar 3. Grafik Akurasi & Loss Model LSTM Pada Sentimen

Berdasarkan hasil pelatihan pada Gambar 3, model *LSTM* untuk klasifikasi sentimen dilatih selama 10 *epoch*. Akurasi *training* meningkat dari 64,13% (*epoch* 1) menjadi 91,19% (*epoch* 4), dan terus naik hingga 98,67% pada *epoch* 9–10. Akurasi validasi tertinggi sebesar 86,48% dicapai pada *epoch* 5 dan kemudian stagnan pada kisaran 83–85%. *Training loss* menurun tajam dari 0,8577 menjadi 0,0335, sementara *validation loss* mencapai nilai terendah 0,3645 pada *epoch* 5 lalu meningkat hingga 0,5979, yang juga mengindikasikan *overfitting*. Penerapan *early stopping* pada *epoch* 4–5 disarankan untuk menjaga generalisasi model.

3.3.3 Evaluasi Model Menggunakan Data Testing Pada Klasifikasi Topik

Evaluasi model LSTM untuk klasifikasi topik diuji menggunakan *confusion matrix* ditampilkan pada Tabel 8 dan *classification report* pada table 9. Dari total 489 *data testing*, model mencapai akurasi keseluruhan sebesar 85,07%, yang menunjukkan performa klasifikasi yang baik.

Tabel 8. *Confusion Matrix* Klasifikasi Topik (Akurasi 85,07%)

<i>Topik</i>	<i>Pengurusan SKCK</i>	<i>Efisiensi Pendaftaran</i>	<i>Isu Teknis</i>	<i>Kepuasan Layanan</i>
Pengurusan SKCK	93	8	8	6
Efisiensi Pendaftaran	5	118	11	5
Isu Teknis	4	11	109	1
Kepuasan Layanan	4	5	3	98

Tabel 9. *Classification Report* Klasifikasi Topik

<i>Topik</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Pengurusan SKCK	0,88	0,81	0,84
Efisiensi Pendaftaran	0,84	0,85	0,84
Isu Teknis	0,83	0,87	0,85
Kepuasan Layanan	0,87	0,88	0,87

Hasil *confusion matrix* menunjukkan bahwa kesalahan klasifikasi terjadi terutama antara aspek yang memiliki kemiripan konteks Efisiensi Pendaftaran sering tertukar dengan Isu Teknis (11 kesalahan), karena keduanya dapat melibatkan permasalahan saat awal penggunaan aplikasi. Demikian pula, Pengurusan SKCK kadang tertukar dengan Efisiensi Pendaftaran karena keduanya mencakup proses administratif. Performa cukup seimbang antar kelas dengan *F1-score* 0,84–0,87

3.3.4 Evaluasi Model Menggunakan Data Testing Pada Klasifikasi Sentimen

Evaluasi model LSTM untuk klasifikasi sentimen menghasilkan akurasi keseluruhan 86,91% dan disajikan pada Tabel 10 dan Tabel 11.

Tabel 10. *Confusion Matrix* Klasifikasi Sentimen (Akurasi 86,91%)

<i>Sentiment</i>	<i>Negatif</i>	<i>Netral</i>	<i>Positif</i>
Negatif	93	1	21
Netral	9	8	29
Positif	9	9	310

Tabel 11. *Classification Report* Klasifikasi Sentimen

<i>Sentiment</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Negatif	0,84	0,81	0,83
Netral	0,44	0,17	0,25
Positif	0,90	0,98	0,94

Model menunjukkan performa yang sangat baik dalam mengenali sentimen positif dan negatif, namun masih mengalami keterbatasan dalam mengklasifikasikan sentimen netral. Rendahnya nilai *recall* pada sentimen netral mengindikasikan adanya ambiguitas bahasa serta kemiripan konteks dengan sentimen lain. Secara keseluruhan, hasil evaluasi menunjukkan bahwa model LSTM memiliki performa yang andal dalam mendukung analisis sentimen berbasis aspek. Keterbatasan pada klasifikasi sentimen netral membuka peluang pengembangan lebih lanjut melalui penambahan data latih atau penyempurnaan proses pelabelan, tanpa mengurangi validitas model dalam mendukung analisis *Net Reputation Score* (NRS).

3.4 Pembahasan Ketidakseimbangan Kelas dan Analisis Kesalahan

Akurasi 85,07% (topik) dan 86,91% (sentimen) sebanding dengan studi serupa. Hammami [6] memperoleh 84% (aspek) dan 83% (sentimen) menggunakan kombinasi LDA + IndoBERT + CNN pada aplikasi IKD, Hidayat dan Gata [5] memperoleh 97% pada klasifikasi sentimen aplikasi Depok *Single Window* tanpa identifikasi aspek, dan Tyas dkk. [7] memperoleh 81% untuk BERT pada konteks yang berbeda (berita saham). *Benchmark* paling sebanding adalah [10] karena objek (aplikasi layanan publik) dan kerangka analisis (ABSA) serupa: hasil penelitian ini lebih unggul +1,07% pada aspek dan +3,91% pada sentimen. Hal ini mengindikasikan bahwa kombinasi *LSTM* dengan *dropout* yang lebih tinggi memberikan generalisasi yang lebih baik dibanding CNN pada konteks ulasan aplikasi berbahasa Indonesia.

Namun, performa kelas netral sangat rendah (*recall* = 0,17; *F1-score* = 0,25). Analisis menunjukkan tiga penyebab. *Pertama*, ketidakseimbangan kelas dari 2.458 ulasan, kelas netral hanya 232 (9,4%), sehingga model cenderung bias ke kelas mayoritas. Pada *data testing* 46 ulasan netral, hanya 8 yang berhasil diklasifikasikan dengan benar, sementara 29 diprediksi sebagai positif dan 9 sebagai negatif pola yang konsisten dengan *majority class bias*. *Kedua*, ambiguitas linguistik label netral ulasan seperti “coba perbanyak metode pembayarannya” mengandung saran sekaligus pengakuan terhadap layanan, sehingga sulit dipisahkan dari sentimen positif/negatif. *Ketiga*, efek *preprocessing* agresif: *stopword removal* yang menghapus kata-kata pemodal (mis. “mungkin”, “sebaiknya”) justru menghilangkan penanda kuat sentimen netral. Implikasinya, klaim performa baik perlu dibatasi pada kelas positif dan negative, peningkatan klasifikasi netral menjadi salah satu arah penelitian lanjutan.

Sebagai contoh *error analysis*, ulasan “aplikasi eror tidak dapat di gunakan” dengan rating 5 diberi label negatif oleh IndoBERT (sesuai isi tekstual yang berisi keluhan), namun LSTM memprediksi positif mengindikasikan bahwa *trainable embedding* LSTM belum menangkap konteks negasi sebaik IndoBERT. Sebaliknya, ulasan singkat seperti “bagus, cepat, yg pasti puas” diklasifikasikan dengan benar sebagai positif oleh kedua model. Hal ini menegaskan keunggulan *pre-trained transformer* pada teks pendek dengan konteks ambigu, dan memberi justifikasi untuk peran IndoBERT sebagai *annotator* utama pada *pipeline* ini. Hasil ini menjawab LSTM mampu mengevaluasi performa klasifikasi dengan akurasi yang baik, meskipun masih memiliki keterbatasan pada kelas netral

3.5 Demo dan Pengujian Model

Untuk memvalidasi *end-to-end* pipeline secara kualitatif, dilakukan pengujian model menggunakan input ulasan baru yang dimasukkan langsung ke dalam sistem. Demo model dideploy sebagai aplikasi berbasis web menggunakan [Streamlit](#) untuk memudahkan eksplorasi oleh pemangku kepentingan. Tabel 12 menyajikan beberapa hasil pengujian yang menunjukkan kemampuan model dalam mengklasifikasikan topik dan sentimen secara konsisten dengan isi tekstual ulasan.

Tabel 12. Contoh Hasil Pengujian Model pada Ulasan Baru

No	Input Ulasan	Topik (LDA)	Sentimen (IndoBERT)
1	pembayaran SKCK sudah dilakukan, tapi setelah 2 hari masih saja status nya belum terbayar. mau bayar lagi dengan BRIVA yang muncul, keterangannya sudah dibayar...	Pengurusan SKCK	Negatif
2	coba perbanyak metode pembayarannya, terus kode bayar sangat lama munculnya. terus dilakukan secara online menggunakan aplikasi sangat keren menurut saya...	Isu Teknis	Negatif
3	sangat membantu, irit juga ongkos transportasi mudah dan aman bisa diisi di mana saja sambil santai, kumpul sama keluarga...	Kepuasan Layanan	Positif
4	aplikasi jelek ribet setiap tes harus bikin akun ribet bgt pokoknya	Efisiensi Pendaftaran	Negatif

Hasil pengujian menunjukkan bahwa model mampu menyesuaikan klasifikasi topik dan sentimen dengan isi keluhan atau pujian yang disampaikan pengguna. Pada input ke-1, keluhan mengenai pembayaran SKCK dan layanan bantuan yang tidak solutif diklasifikasikan ke topik Pengurusan SKCK dengan sentimen negatif. Pada input ke-2, kritik mengenai metode pembayaran terbatas dan kode bayar yang lambat diklasifikasikan sebagai Isu Teknis dengan sentimen negatif. Pada input ke-3, apresiasi terhadap kemudahan, efisiensi waktu, dan penghematan biaya diklasifikasikan sebagai Kepuasan Layanan dengan sentimen positif. Pada input ke-4, keluhan tentang proses

pendaftaran yang rumit dikenali sebagai Efisiensi Pendaftaran dengan sentimen negatif. Pengujian ini memperkuat temuan kuantitatif pada Bagian 3.4 bahwa *pipeline* terpadu LDA - IndoBERT - LSTM mampu mengaitkan aspek dan sentimen secara konsisten dengan konteks ulasan pengguna

4. KESIMPULAN

Penelitian ini berhasil menerapkan pendekatan *Aspect-Based Sentiment Analysis* (ABSA) untuk menganalisis ulasan pengguna Aplikasi Polri Presisi secara sistematis. penelitian ini memberikan tiga kontribusi ilmiah utama. Pertama, penelitian ini mengusulkan dan mengimplementasikan pipeline ABSA terpadu (LDA - IndoBERT - NRS - LSTM) yang sebelumnya belum tersedia dalam satu kerangka pada konteks layanan publik digital Indonesia. Kedua, penelitian ini memperkaya literatur text mining dan NLP bahasa Indonesia dengan menunjukkan bahwa kombinasi topic modeling tanpa pelabelan awal dan klasifikasi sentimen *transformer-based* mampu memberikan akurasi kompetitif (85,07% aspek, 86,91% sentimen) sekaligus menjaga interpretabilitas. Ketiga, penelitian ini menempatkan NRS sebagai metrik prioritas perbaikan yang menerjemahkan distribusi sentimen menjadi peringkat kuantitatif, di mana aspek Isu Teknis (NRS = -19,84%) menjadi prioritas perbaikan utama, diikuti Efisiensi Pendaftaran (51,51%), Pengurusan SKCK (75,78%), dan Kepuasan Layanan (77,57%).

Penelitian ini memiliki sejumlah keterbatasan yang perlu diakui. Salah satunya adalah rendahnya nilai recall pada klasifikasi sentimen netral, yang hanya mencapai 0,17 akibat ketidakseimbangan jumlah data antar kelas. Selain itu, proses pelabelan sentimen yang dilakukan secara semi-otomatis masih belum dilengkapi dengan validasi antar-annotator manusia, sehingga berpotensi menimbulkan bias subjektivitas pada data latih. Kelemahan berikutnya terletak pada cakupan data yang terbatas hanya pada satu versi aplikasi (2.0.27) dalam rentang waktu tertentu, sehingga generalisasi hasil belum dapat sepenuhnya dijamin. Tidak hanya itu, pemilihan jumlah topik LDA sebanyak empat ($k = 4$) yang lebih mengutamakan kemudahan interpretasi dibandingkan koherensi statistik puncak pada $k = 8$, serta belum dilakukannya perbandingan langsung dengan metode baseline klasik seperti SVM, *Naive Bayes*, dan *Random Forest*, maupun dengan model IndoBERT yang di-*fine-tune* secara langsung sebagai classifier, juga menjadi celah yang membatasi kokohnya klaim performa dalam penelitian ini.

Untuk menjawab keterbatasan tersebut, sejumlah arah penelitian lanjutan sangat direkomendasikan. Peningkatan performa pada kelas netral dapat diupayakan melalui penerapan teknik penanganan ketidakseimbangan data seperti SMOTE [21], *class weighting*, atau *focal loss* [20], yang sekaligus dapat dikombinasikan dengan pelabelan manual ulang pada sebagian subset data oleh annotator manusia disertai perhitungan inter-annotator agreement guna menjamin kualitas label. Selain itu, penguatan klaim keunggulan pipeline perlu didukung oleh perbandingan komprehensif dengan arsitektur *transformer* lain (IndoBERT-LARGE, mBERT, XLM-RoBERTa, serta IndoBERT yang di-*fine-tune* langsung) dan baseline machine learning klasik, sehingga kontribusi metodologis dapat diukur secara lebih objektif. Tidak kalah penting, validasi data yang lebih luas pada versi aplikasi Polri Presisi lainnya maupun pada aplikasi layanan publik sejenis (seperti PeduliLindungi, IKD, dan JMO) sangat diperlukan untuk menguji portabilitas pipeline yang telah dibangun. Terakhir, pengembangan dashboard interaktif berbasis hasil *pipeline* dinilai krusial guna mendukung pemantauan *Net Reputation Score* secara berkelanjutan oleh para pengelola aplikasi pemerintah dalam pengambilan keputusan strategis.

DAFTAR PUSTAKA

- [1] R. Fatmasari, W. Gata, N. K. Wardhani, K. Prayogi, and M. B. Husna, "Klasifikasi sentimen terhadap kualitas aplikasi bahan ajar digital akademik Universitas Terbuka di *Google Play*," Jurnal SAINTEKOM, vol. 14, no. 1, pp. 48–60, 2024.
- [2] D. Fristtikasari, S. Alam, and I. Kurniawan, "Analisis sentimen pengguna aplikasi Kitalulus pada ulasan *Google Play Store* menggunakan metode *Naïve Bayes*," Jurnal Teknologi Informasi dan Komputer, vol. 10, no. 2, pp. 458–473, 2024.
- [3] I. K. S. A. Rizaldi and S. Alam, "Analisis sentimen pengguna aplikasi JMO (Jamsostek Mobile) pada *Google Play Store* menggunakan metode *Naïve Bayes*," STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer, vol. 3, no. 1, pp. 45–58, 2024.
- [4] S. Poria, E. Cambria, D. Hazarika, and P. Vij, "*Aspect-based sentiment analysis: A survey*," IEEE Computational Intelligence Magazine, vol. 11, no. 2, pp. 21–35, May 2016.
- [5] D. M. Blei, A. Y. Ng, and M. I. Jordan, "*Latent Dirichlet allocation*," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, Jan. 2003.

- [6] B. Wilie, K. Vincentio, G. I. Winata, S. Cahyawijaya, X. Li, Z. Y. Lim, S. Soleman, R. Mahendra, P. Fung, S. Bahar, and A. Purwarianti, "IndoNLU: *Benchmark and resources for evaluating Indonesian natural language understanding*," in Proc. 1st Conf. Asia-Pacific Chapter Assoc. Comput. Linguistics and 10th Int. Joint Conf. Nat. Lang. Process. (AACL-IJCNLP), 2020, pp. 843–857.
- [7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
- [8] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. 10th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), Seattle, WA, USA, 2004, pp. 168–177.
- [9] R. Hidayat and W. Gata, "Pemanfaatan IndoBERT dan Long Short-Term Memory (LSTM) pada analisis sentimen ulasan aplikasi Depok Single Window," *Information System for Educators and Professionals*, vol. 10, no. 1, pp. 13–24, 2025.
- [10] A. Hammami, "Aspect-based sentiment analysis ulasan pengguna aplikasi Identitas Kependudukan Digital Indonesia," *Jurnal Statistika dan Aplikasinya*, vol. 8, no. 2, pp. 121–134, 2024.
- [11] H. A. Tyas, R. Sarno, and B. S. Rintyarna, "Analisis perbandingan metode klasifikasi sentimen berita saham: Pendekatan *machine learning*, *deep learning*, *transfer learning*, dan *graf*," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 10, no. 2, pp. 145–158, 2024.
- [12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. 2019 Conf. North Am. Chapter Assoc. Comput. Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2019, pp. 4171–4186.
- [13] H. Jelodar, Y. Wang, R. Orji, and X. Zeng, "Latent Dirichlet allocation (LDA) for topic modeling: A review," *Neurocomputing*, vol. 358, pp. 290–305, Sep. 2019.
- [14] D. Ma, S. Li, Y. Wang, and B. Li, "Interactive attention networks for aspect-level sentiment classification," in Proc. 26th Int. Joint Conf. Artificial Intelligence (IJCAI), Melbourne, Australia, 2017, pp. 4068–4074.
- [15] A. Khan, B. Baharudin, and A. Khan, "Text preprocessing and feature engineering for sentiment analysis: A survey," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, pp. 160–170, 2020.
- [16] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.
- [17] F. Reichheld, "The one number you need to grow," *Harvard Business Review*, vol. 81, no. 12, pp. 46–54, Dec. 2003.
- [18] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in Proc. Int. Conf. Learning Representations (ICLR) Workshop, Scottsdale, AZ, USA, 2013, pp. 1–12.
- [19] Y. Goldberg, "A primer on neural network models for natural language processing," *Journal of Artificial Intelligence Research*, vol. 57, pp. 345–420, 2016.
- [20] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proc. IEEE Int. Conf. Computer Vision (ICCV), Venice, Italy, 2017, pp. 2980–2988.
- [21] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.