INFOTECH journal
ISSN : 2460-1861 (Print), 2615-4250 (Online)
Vol. 12 No. 1 2026

PENERAPAN INDOBERT UNTUK ANALISIS SENTIMEN BERBASIS ASPEK PADA EVALUASI APLIKASI LAYANAN PUBLIK

Muhammad Rafli[1], Badieah Assegaf[2]
[1,2]Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Sutan Agung, Semarang, Indonesia
Penulis Korespondensi: muhammadrafli09.ml@gmail.com

ABSTRAK

Ulasan pengguna pada aplikasi layanan publik digital seperti Mobile JKN, M-Paspor, dan SIGNAL menyimpan wawasan penting bagi peningkatan kualitas layanan. Namun, analisis sentimen konvensional belum mampu mengidentifikasi aspek spesifik yang menjadi sumber kepuasan maupun keluhan pengguna. Penelitian ini mengembangkan sistem *Aspect-Based Sentiment Analysis* (ABSA) berbasis IndoBERT dengan pendekatan *token classification*. Sistem dirancang untuk mengekstraksi tiga aspek utama, yaitu *Usability*, *Reliability*, dan *Efficiency*, sekaligus mengklasifikasikan kategori sentimen. Metode penelitian meliputi pengumpulan data *scraping*, pra-pemrosesan, dan *fine-tuning* model IndoBERT. Hasil pengujian menunjukkan kinerja yang sangat baik dengan akurasi pada aspek *Usability* 0.98, *Reliability* 0.97, *Efficiency* 0.99 untuk Mobile JKN. Untuk M-Paspor mendapatkan akurasi aspek *Usability* 0.99, *Reliability* 0.98, *Efficiency* 0.99. Dan untuk SIGNAL mendapatkan akurasi aspek *Usability* 0.99, *Reliability* 0.98, *Efficiency* 0.97. Analisis kurva *loss* mengungkap adanya indikasi *overfitting* ringan pada model Mobile JKN yang disebabkan oleh kompleksitas dan keberagaman domain layanan kesehatan, sementara model M-Paspor dan SIGNAL juga mengalami *overfitting* namun dengan tingkat yang lebih rendah berkat karakteristik domain yang lebih spesifik dan jumlah data pelatihan yang lebih besar. Analisis lebih lanjut mengungkap bahwa aspek *Efficiency* secara konsisten menjadi yang paling mudah diklasifikasi dengan F1-Score di atas 0.94, sementara aspek *Reliability* menjadi tantangan utama dengan pola kesulitan berbeda pada setiap dataset. Temuan ini menunjukkan bahwa sistem yang dibangun efektif dalam memberikan wawasan terperinci guna mendukung peningkatan berkelanjutan layanan publik digital.

***Kata Kunci**: Analisis Sentimen, Multi Aspek, IndoBERT, Token Classification, Fine-Tuning, Layanan Publik, Pemrosesan Bahasa Alami.*

Riwayat Artikel :

Tanggal diterima : 23-02-2026 Tanggal terbit : 26-04-2026

Kutipan :

Rafli, M., & Assegaf, B. (2026). PENERAPAN INDOBERT UNTUK ANALISIS SENTIMEN BERBASIS ASPEK PADA EVALUASI APLIKASI LAYANAN PUBLIK. INFOTECH Journal, 12(1), 93–101. https://doi.org/10.31949/infotech.v12i1.17555

1. PENDAHULUAN

Evolusi teknologi digital dan komunikasi telah mendorong perubahan mendasar dalam cara masyarakat mengakses berbagai layanan, termasuk layanan publik. Menanggapi perubahan ini, pemerintah Indonesia menghadirkan sejumlah aplikasi digital seperti Mobile JKN, M-Paspor, dan Samsat Digital Nasional (SIGNAL) untuk meningkatkan efisiensi dan efektivitas pelayanan . Aplikasi-aplikasi ini memberikan kemudahan nyata mulai dari pelayanan kesehatan, pengurusan dokumen keimigrasian, hingga pembayaran pajak kendaraan, yang secara signifikan menghemat waktu dan sumber daya masyarakat (Filemon Haganta Kaban & Yudistira, 2021). Tingginya tingkat adopsi platform digital ini berbanding lurus dengan meluasnya penetrasi internet di Indonesia yang tercatat mencapai 69,21% pada tahun 2023, menandakan bahwa platform digital telah menjadi jalur interaksi utama antara warga dan pemerintah (Badan Pusat Statistik, 2024).

Analisis sentimen merupakan proses memahami dan mengekstrak data tekstual untuk memperoleh informasi mengenai sentimen yang terkandung di dalamnya, baik positif maupun negatif (Rahman et al., 2024). Beberapa penelitian terdahulu telah mengeksplorasi analisis sentimen pada aplikasi layanan publik menggunakan berbagai pendekatan. Husada & Paramita (2021) mengimplementasikan *Support Vector Machine* (SVM) untuk analisis sentimen maskapai penerbangan di *platform twitter* dan akurasi terbaik mencapai 84,37% (Husada & Paramita, 2021). Implementasi model SmallBERT pada ulasan hotel berbahasa Indonesia dan berhasil mencapai akurasi 91,40% setelah *fine-tuning* (Chandradev et al., 2023). Penerapan *Aspect-Based Sentiment Analysis* (ABSA) pada aplikasi KAI *Access* menggunakan SVM untuk menganalisis aspek Learnability, *Efficiency*, *Errors*, dan *Satisfaction*, dengan akurasi terbaik mencapai 94,73% (Radiena & Nugroho, 2023). Penggunaan *Latent Dirichlet Allocation* (LDA) untuk ekstraksi aspek dan IndoBERT untuk klasifikasi sentimen pada aplikasi M-Paspor, mencapai akurasi 94% (Widiansyah et al., 2024). Dalam survey komprehensif menyatakan bahwa pendekatan *deep learning* berbasis transformer menunjukkan performa superior dibandingkan metode klasik untuk tugas ABSA (Zhou et al., 2019). pendekatan *end-to-end* dalam ABSA memberikan hasil yang lebih konsisten dibandingkan pendekatan multi-tahap (Truşcă & Frasincar, 2023). Implementasi IndoBERT yang dilatih khusus dengan korpus bahasa Indonesia mengungguli Multilingual BERT pada berbagai tugas *Natural Language Processing* (NLP) bahasa Indonesia (Nuryadi et al., 2025).

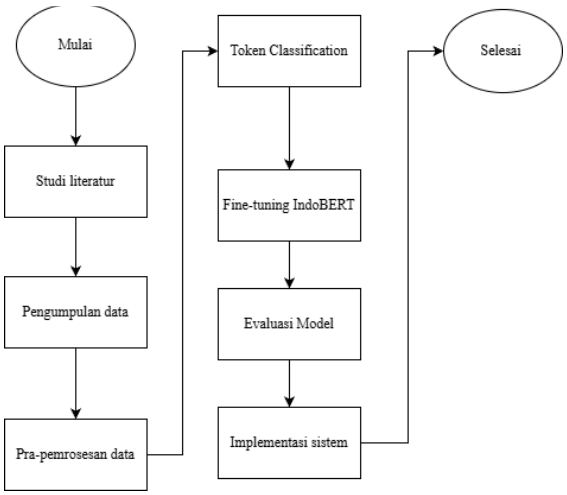
Meskipun penelitian-penelitian tersebut telah memberikan kontribusi signifikan, terdapat beberapa gap yang perlu diatasi. Pertama, mayoritas pendekatan yang ada menggunakan metode multi-tahap dimana ekstraksi aspek dan klasifikasi sentimen dilakukan secara terpisah, yang berpotensi mengurangi efisiensi dan konsistensi hasil analisis. Kedua, penerapan ABSA pada layanan publik di Indonesia masih terbatas pada domain tunggal dan belum mencakup evaluasi komparatif antar aplikasi yang berbeda karakteristiknya. Ketiga, sebagian besar penelitian masih menggunakan metode klasik seperti SVM yang memiliki keterbatasan dalam memahami konteks semantik bahasa Indonesia secara mendalam.

Penelitian ini bertujuan mengembangkan sistem berbasis ABSA pada aplikasi pelayanan masyarakat dengan mengimplementasikan arsitektur model IndoBERT untuk mendeteksi aspek dan polaritas sentimen secara bersamaan menggunakan pendekatan *token classification*. Secara spesifik, penelitian ini mengidentifikasi performa model IndoBERT dalam melakukan klasifikasi sentimen berbasis aspek pada ulasan aplikasi Mobile JKN, M-Paspor, dan SIGNAL. Selain itu, penelitian ini menganalisis aspek layanan yang mengacu pada standar ISO/IEC 25010 yaitu *Usability*, *Reliability*, dan *Efficiency*, yang dominan menerima sentimen positif dan negatif pada setiap aplikasi berdasarkan hasil model ABSA yang dikembangkan.

Kontribusi utama penelitian ini adalah pengembangan sistem ABSA *end-to-end* berbasis IndoBERT yang mampu mendeteksi multi-aspek dalam satu kalimat secara simultan, berbeda dengan pendekatan konvensional yang hanya mendeteksi satu aspek dominan. Penelitian ini juga menyediakan framework evaluasi berbasis data yang spesifik dan dapat ditindaklanjuti bagi penyedia layanan publik dalam mengidentifikasi aspek kualitas yang memerlukan perbaikan prioritas. Selain itu, penelitian ini berkontribusi pada pengayaan riset di bidang NLP bahasa Indonesia, khususnya dalam penggunaan model *transformer* untuk analisis sentimen berbasis aspek pada domain layanan publik.

2. METODE

Penelitian ini dimulai dari studi literatur untuk membangun landasan teori, diikuti dengan pengumpulan dan persiapan data, perancangan dan pelatihan model, hingga implementasi model ke dalam sebuah sistem. Alur metode penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur metode penelitian

Tahapan penelitian diawali dengan studi literatur guna membangun dasar teori yang komprehensif. Selanjutnya dilakukan pengumpulan data berupa ulasan aplikasi yang diperoleh dari *Google Play Store*. Data mentah yang terkumpul kemudian diproses pada tahap pra-pemrosesan, yang mencakup dua kegiatan utama, yaitu pembersihan data (*preprocessing*) serta pelabelan data (*labeling*) secara manual, sehingga dihasilkan dataset yang bersih dan teranotasi dengan baik. Dataset tersebut selanjutnya digunakan pada tahap token *classification* untuk merumuskan *Aspect-Based Sentiment Analysis* (ABSA) pada tingkat token. Dilakukan *sequence labeling* agar sistem mampu mengidentifikasi lebih dari satu aspek dalam satu kalimat ulasan. Tahap berikutnya adalah proses *fine-tuning* model dengan menggunakan hyperparameter yang telah ditetapkan sebelumnya. Kinerja model hasil pelatihan kemudian dievaluasi secara kuantitatif melalui tahap evaluasi model dengan memanfaatkan dataset pelatihan, dataset validasi, dan dataset pengujian. Pada tahap akhir model terbaik diimplementasikan ke dalam sebuah sistem berbasis web menggunakan *framework Streamlit*.

2.1. Studi Literatur

Tahap awal dalam penelitian ini diawali dengan pelaksanaan studi literatur. Pada tahap tersebut dilakukan pengumpulan, penelaahan, serta analisis terhadap berbagai sumber ilmiah, seperti buku dan artikel jurnal, yang berkaitan dengan topik *Aspect-Based Sentiment Analysis* (ABSA). Fokus kajian diarahkan pada penelitian-penelitian sebelumnya yang mengimplementasikan model berbasis *Transformer*, khususnya BERT, dalam pengolahan teks berbahasa Indonesia. Hasil kajian literatur menjadi landasan dalam menentukan pendekatan, arsitektur model, serta metode evaluasi yang digunakan dalam penelitian ini.

2.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa ulasan pengguna yang diperoleh dari aplikasi layanan publik di Google Play Store. Aplikasi yang menjadi objek penelitian meliputi Mobile JKN, M-Paspor, dan SIGNAL. Masing-masing aplikasi diambil sebanyak 1.700 ulasan sehingga total data awal yang terkumpul berjumlah 5.100 ulasan sebelum dilakukan *preprocessing*. Pemilihan ketiga aplikasi tersebut tidak dilakukan secara acak melainkan dengan mempertimbangkan terhadap tiga sektor layanan publik esensial, yaitu kesehatan pada Mobile JKN, keimigrasian pada M-Paspor, dan administrasi kendaraan bermotor SIGNAL . Proses pengumpulan data dilakukan secara otomatis menggunakan teknik *web scraping* dengan pustaka Python *google_play_scraper*. Setelah melalui tahap *preprocessing* dan labeling manual ditemukan adanya ketidakseimbangan distribusi pada class aspek-sentimen di setiap aplikasi. Untuk mengatasi masalah imbalanced class tersebut, dilakukan teknik *oversampling* pada masing-masing dataset aplikasi. *Oversampling* difokuskan pada class aspek-sentimen yang minoritas sehingga distribusi class menjadi lebih seimbang. Teknik ini menyebabkan perbedaan jumlah data akhir pada setiap aplikasi karena tingkat ketidakseimbangan *class* yang berbeda-beda. Dataset dibagi dengan rasio 80:10:10 untuk data *training*, *validation*, dan *testing* seperti yang ditunjukkan pada Tabel 1.

Tabel 1. *Splitting* data

Aplikasi	*Data Training*	*Data Validation*	*Data Testing*
Mobile JKN	2452	307	307
M-Paspor	3398	500	500
Signal	3796	475	475

2.3. Pra-pemrosesan Data

Pada tahap pra-pemrosesan data dilakukan dua langkah utama, yaitu *text processing* dan *labeling data*.

a. *Text Preprocessing*

Tahap *text preprocessing* merupakan proses persiapan data teks dari kondisi mentah hingga siap digunakan pada tahapan analisis selanjutnya. Proses ini meliputi beberapa tahapan utama, yaitu *case folding*, *cleaning text*, dan *normalization*.

1. *Case Folding*

Case folding dilakukan dengan mengubah seluruh teks menjadi huruf kecil guna menjaga konsistensi data, sehingga perbedaan penggunaan huruf kapital tidak menyebabkan kata yang sama dianggap berbeda oleh sistem(Sari et al., 2025).

2. *Cleaning Text*

Setelah itu, dilakukan tahap cleaning text untuk menghilangkan elemen-elemen yang tidak relevan dalam analisis, seperti simbol, tanda baca, emoticon, angka, hashtag, dan URL. Proses ini bertujuan menghasilkan teks yang lebih bersih dan mudah diproses oleh model(Sari et al., 2025).

3. *Normalization*

Tahap selanjutnya adalah *normalization*, yaitu proses penggantian kata tidak baku, singkatan, atau bahasa informal menjadi kata baku menggunakan kamus normalisasi (Safitri et al., 2023). Tahap ini bertujuan mengurangi variasi kata dan meningkatkan konsistensi representasi teks.

b. *Labeling Data*

Pelabelan data dilakukan setelah tahap *preprocessing*, dan proses anotasi dikerjakan secara manual berdasarkan aspek, sentimen, serta kriteria yang mangacu pada ISO/IEC 25010. Aspek *usability* (kemudahan penggunaan) mengacu pada sejauh mana pengguna dapat menggunakan aplikasi tanpa

memerlukan pembelajaran tambahan. Aspek *reliability* (keandalan) menilai kemampuan sistem dalam mempertahankan kinerja secara konsisten pada kondisi tertentu dalam rentang waktu tertentu. Sementara itu, aspek *efficiency* (efisiensi) berkaitan dengan kemampuan sistem dalam memberikan performa yang optimal, terutama yang berkaitan dengan perilaku waktu. Proses pelabelan ini dilakukan dengan bantuan perangkat lunak Label Studio versi 1.20.

2.4. *Token Classification*

Pada tahap *Token Classification* dalam ABSA, setiap kalimat terlebih dahulu melalui proses tokenisasi menggunakan *tokenizer* dari model IndoBERT (indobenchmark/indobert-base-p1). Proses tokenisasi ini bertujuan untuk memecah kalimat menjadi unit-unit token yang dapat diproses oleh model. Selanjutnya, setiap token yang dihasilkan disejajarkan dengan label yang telah ditentukan berdasarkan skema anotasi aspek.Token yang merupakan bagian dari kata atau frasa yang merepresentasikan aspek tertentu diberi *sequence labeling* sesuai dengan skema BIO (*Begin*, *Inside*, *Outside*), sedangkan token yang tidak berkaitan dengan aspek diberikan label O (Hisyam Pradhana et al., 2025). Tahap ini menghasilkan dataset siap latih yang terdiri dari pasangan token dan label aspek. Dataset tersebut kemudian digunakan oleh model untuk mempelajari pola kemunculan aspek dalam teks, sehingga model mampu melakukan identifikasi aspek secara otomatis dari ulasan sebagai langkah awal dalam proses analisis sentimen berbasis aspek. Pada Tabel 2 merupakan contoh hasil penerapan *token classification*.

Tabel 2 *Token Classification*

Label	*Before*	*After*
{"tidak berguna verifikasi wajah dari pagi malam pagi ulang tetap gagal terus": ["*efficiency* ", "negatif"]}	['tidak', 'berguna', 'verifikasi', 'wajah', 'dari', 'pagi', 'malam', 'pagi', 'ulang', 'tetap', 'gagal', 'terus']	['O', 'O', 'B-*efficiency*-negatif', 'I-*efficiency*-negatif', 'I-*efficiency*-negatif', 'I-*efficiency*-negatif', 'I-*efficiency*-negatif', 'I-*efficiency*-negatif', 'I-*efficiency*-negatif', 'I-*efficiency*-negatif']

2.5. *Fine-tuning IndoBERT*

Fine-tuning merupakan proses adaptasi model *neural network* yang telah dilatih sebelumnya agar sesuai dan optimal dalam menyelesaikan tugas spesifik yang baru (Nyoman Saputra Wahyu Wijaya et al., 2025). Pada konteks tersebut Model IndoBERT dilatih melalui proses *fine-tuning* untuk tujuan ABSA dengan menetapkan jumlah *epoch* sebanyak 5. Pemilihan jumlah *epoch* ini didasarkan pada hasil pengamatan terhadap data validasi, yang menunjukkan bahwa kinerja model mencapai kondisi optimal dan mulai menunjukkan kecenderungan stagnasi setelah *epoch* kelima. *Learning rate* sebesar 3e-5 digunakan karena nilai tersebut umum direkomendasikan pada proses *fine-tuning* model berbasis BERT dan terbukti mampu menghasilkan pelatihan yang stabil (Nur Karimah & Anna Baita, 2024). Proses pelatihan model diimplementasikan menggunakan pustaka *HuggingFace Transformers*. Tabel 3 menyajikan rincian *hyperparameter* yang digunakan.

Tabel 3. *Hyperparamater*

Hyperparameter	***Nilai***
Max Sequence Length	256
Batch Size	8
Epoch	5
Learning Rate	3e-5 (0.00003)
Weight Decay	0.01

2.6. Evaluasi Model

Evaluasi akhir model dilakukan pada data uji yang belum pernah dilihat oleh model sebelumnya. Performa model diukur menggunakan metrik standar untuk tugas *sequence labeling* dari pustaka seqeval, yang meliputi *accuracy* merupakan proses pengukuran pada seberapa banyak prediksi yang benar dibandingkan dengan seluruh data, *precision* proses mengukur sejauh mana prediksi porsitif yang dihasilkan pada model benar, *recall* proses mengukur sejauh mana model dapat menangkap semua data positif yang sebenarnya, *f1-score* didefinisikan sebagai rata-rata harmonik dari *precision* dan *recall*. yang bertujuan untuk menyeimbangkan antara *precision* dan *recall* (Hakim et al., 2024). Perhitungan metrik dilakukan secara mikro (*micro average*) dan makro (*macro average*) untuk mendapatkan gambaran performa model secara keseluruhan dan per kategori label. Perhitungan nilai *accuracy* ,*recall*, *precision*, dan *f1-score* masing-masing ditunjukkan pada Persamaan (1), Persamaan (2), Persamaan (3), dan Persamaan (4).

$$accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \quad (1)$$

$$Precision = \frac{Tp}{Tp + Fp} \quad (2)$$

$$Recall = \frac{Tp}{Tp + Fn} \quad (3)$$

$$F1 - Score = \frac{2\ x\ Precision\ Recall}{Precision + Recall} \quad (4)$$

Keterangan:

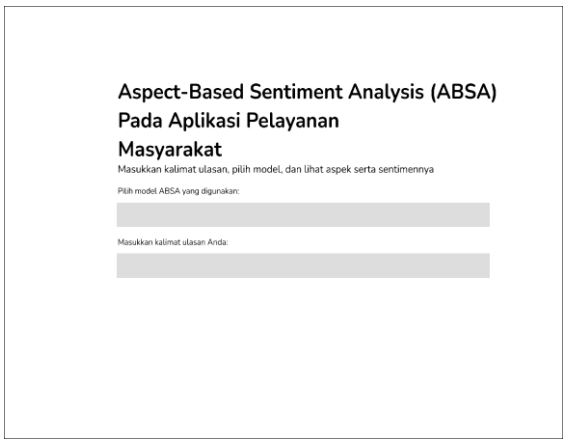
(TP) = *True Positive*
(TN) = *True Negative*
(FP) = *False Positive*
(FN) = *False Negative*

2.7. Implementasi Sistem

Pada Implementasi sistem melakukan perancangan interface dan pembuatan sistem dengan rincian sebagai berikut.

1. Perancangan *User Interface*

Pada tahap perancangan *interface*, dilakukan proses desain dan pembuatan rancangan visual yang akan menjadi jembatan interaksi antara pengguna dengan sistem. Gambar 2 merupakan tampilan dari *user interface*.

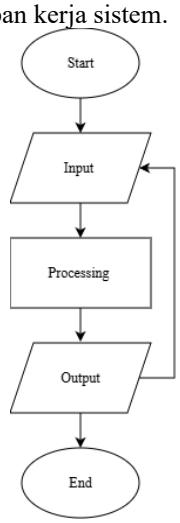


Gambar 2. *User interface*

Gambar 2 merupakan tampilan *interface* ketika pertama kali diakses. Pada halaman ini, pengguna dapat melihat dua komponen input utama yang dirancang untuk memudahkan interaksi dengan sistem. Komponen pertama adalah sebuah *dropdown menu* yang memungkinkan pengguna untuk memilih model klasifikasi yang diinginkan dari beberapa opsi model yang telah tersedia. Komponen kedua adalah sebuah text area input yang didesain khusus untuk memungkinkan pengguna memasukkan teks ulasan yang akan dianalisis oleh sistem.

2. Tahapan Kerja Sistem

Pembuatan sistem dibangun menggunakan *framework Streamlit* dan dihubungkan melalui file utama app.py, yang di dalamnya telah dimuat model IndoBERT-ABSA yang telah dilatih. Gambar 3 merupakan tahapan kerja sistem.



Gambar 3. Tahapan kerja sistem

Berikut ini penjelasan dari masing-masing tahapan berdasarkan alur sistem yang telah diterapkan pada antarmuka *Streamlit*:

a. Mulai
   Sistem aktif dan siap digunakan ketika pengguna mengakses aplikasi melalui antarmuka *Streamlit*. Pada tahap ini, halaman analisis menampilkan dua komponen input utama, yaitu *dropdown* menu untuk memilih model aplikasi (Mobile JKN, M-Paspor, atau SIGNAL) dan *text* area untuk memasukkan teks ulasan yang akan dianalisis. Tahap ini merupakan awal dari interaksi antara pengguna dengan sistem analisis sentimen berbasis aspek.
b. Input Data
   Pada tahap input, pengguna Memilih model aplikasi melalui dropdown menu sesuai dengan jenis ulasan yang akan dianalisis Mobile JKN, M-Paspor, atau SIGNAL. Pemilihan ini menentukan model spesifik yang akan digunakan untuk prediksi. Memasukkan teks ulasan pada *text* area yang tersedia. Teks dapat berupa satu atau beberapa kalimat yang mengandung opini pengguna terhadap aplikasi.
c. *Processing*
   Pada Tahap processing data yang telah diinput *preprocessing*, kemudia teks yang telah bersih dianalisis menggunakan model IndoBERT-ABSA yang telah di*fine-tuning* sesuai dengan aplikasi yang dipilih pada tahap prediksi model. Proses prediksi dimulai dengan token *classification*, di mana model melakukan klasifikasi pada setiap token menggunakan pendekatan sequence labeling. Kemudian, sistem mengidentifikasi span token yang merepresentasikan aspek-aspek spesifik seperti *Usability*, *Reliability*, dan *Efficiency* melalui proses ekstraksi aspek. Terakhir, setiap aspek yang berhasil terdeteksi akan diklasifikasikan sentimennya menjadi Positif atau Negatif.
d. *Output*
   Setelah selesai melakukan hasil analisis akan ditampilan yang mencakup aspek-aspek yang terdeteksi beserta label sentimennya, seperti *Usability*-Negatif atau *Reliability*-Negatif. Selain itu, sistem juga menampilkan teks yang di *highlight* untuk menunjukkan bagian ulasan mana yang teridentifikasi sebagai aspek tertentu.

3. PEMBAHASAN

3.1. Analisis Performa Model Mobile JKN

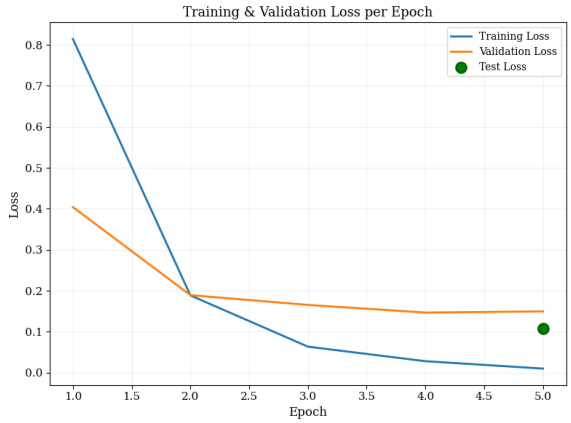
Evaluasi model pada dataset Mobile JKN berfokus pada kemampuan generalisasi model dalam mengklasifikasikan aspek dan sentimen pada data uji. Pada Tabel 4 merupakan rincian performa per kelas.

Tabel 4. Matrik evaluasi Mobile JKN

Aspek	Sentimen	*Precicion*	*Recall*	*F1-Score*	*Accurarcy* (Aspek)
Usability	Positif	0.87	0.96	0.91	0.98
Usability	Negatif	0.90	0.97	0.94	0.98
Reliability	Positif	0.95	0.98	0.97	0.97
Reliability	Negatif	0.74	0.86	0.80	0.97
Efficiency	Positif	0.92	0.95	0.94	0.99
Efficiency	Negatif	0.97	1.00	0.99	0.99

Berdasarkan Tabel 4 hasil pengujian pada dataset Mobile JKN menunjukkan aspek *Efficiency* menunjukkan kinerja terbaik dengan *F1-Score* mencapai 0.99 pada sentimen negatif dan akurasi aspek sebesar 0.99, mengindikasikan bahwa model sangat andal dalam mengenali ulasan yang berkaitan dengan efisiensi aplikasi, baik positif maupun negatif. Sementara itu, aspek *Reliability* pada sentimen negatif mencatatkan performa paling rendah dibandingkan kelas lainnya, dengan nilai *Precision* sebesar 0.74, *Recall* 0.86, dan *F1-Score* 0.80. Hal ini mengindikasikan bahwa model terkadang mengalami kesulitan dalam membedakan keluhan terkait *reliability* dengan keluhan pada aspek lain, karena kemiripan konteks bahasa yang

digunakan pengguna dalam menyampaikan ketidakpuasan terhadap keandalan layanan. Aspek *Usability* menunjukkan performa yang cukup baik dengan akurasi aspek sebesar 0.98 meskipun nilai *Precision* pada sentimen positif 0.87 sedikit lebih rendah dibandingkan kelas lainnya. Secara keseluruhan model mampu mengklasifikasikan sebagian besar aspek dan sentimen dengan baik, dengan *nilai F1-Score* yang bervariasi antara 0.80 hingga 0.99 di seluruh kelas. Gambar 4 merupakan visualisasi grafik *loss* model.



Gambar 4. Grafik *Loss* Mobile JKN

Berdasarkan Gambar 4, grafik menunjukkan penurunan *Training Loss* yang tajam mendekati 0.0, sementara *Validation Loss* melandai di angka 0.15. Celah yang cukup signifikan antara *Training Loss* dan *Validation Loss* ini mengindikasikan adanya gejala *overfitting* ringan, yang berarti model cenderung terlalu menyesuaikan diri dengan pola data latih sehingga kemampuannya dalam menggeneralisasi pada data baru menjadi sedikit terbatas. Kondisi ini dapat dijelaskan oleh beberapa faktor. Domain layanan kesehatan pada Mobile JKN bersifat lebih luas dan variatif, mencakup berbagai topik seperti keluhan antrian, akses faskes, klaim BPJS, hingga fitur aplikasi, sehingga model lebih sulit membangun representasi yang konsisten. Teknik *oversampling* yang diterapkan untuk menyeimbangkan kelas minoritas berpotensi mengakibatkan model terlalu familiar dengan sampel yang direplikasi. Meskipun demikian, nilai *Test Loss* yang rendah (0.1) mengonfirmasi bahwa model tetap memiliki kemampuan generalisasi yang cukup baik terhadap data baru, dan akurasi *testing* sebesar 97% membuktikan bahwa dampak *overfitting* tersebut masih dalam batas yang dapat ditoleransi.

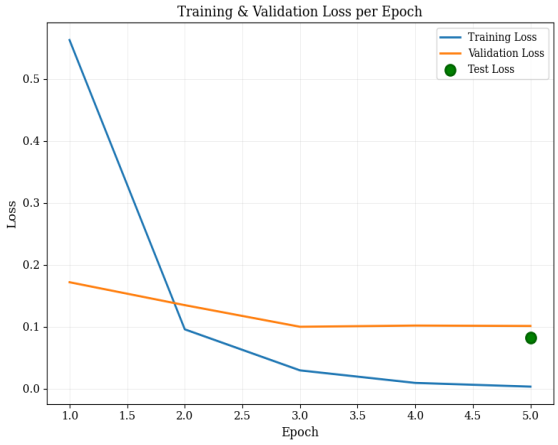
3.2. Analisis Performa Model M-Paspor

Pada dataset M-Paspor, model menunjukkan performa yang sedikit lebih unggul dengan akurasi pengujian mencapai 98%. Pada Tabel 5 merupakan hasil evaluasi mendetail.

Tabel 5. Matrik evaluasi M-Paspor

Aspek	Sentimen	*Precicion*	*Recall*	*F1-Score*	*Accurarcy* (Aspek)
Usability	Positif	0.94	0.97	0.96	0.99
Usability	Negatif	0.92	0.96	0.94	0.99
Reliability	Positif	0.98	0.97	0.96	0.98
Reliability	Negatif	0.86	0.92	0.89	0.98
Efficiency	Positif	0.99	0.99	0.99	0.99
Efficiency	Negatif	0.94	1.00	0.97	0.99

Hasil pada Tabel 5 menunjukkan bahwa aspek *Efficiency* kembali menjadi aspek yang paling mudah diklasifikasikan, dengan F1-Score yang konsisten tinggi, yaitu 0.99 pada sentimen positif dan 0.97 pada sentimen negatif, serta akurasi aspek secara keseluruhan mencapai 0.99. Aspek *Usability* juga menunjukkan performa yang sangat baik dengan akurasi aspek mencapai 0.99. Sentimen positif mencatatkan Precision sebesar 0.94, Recall 0.97, dan F1-Score 0.96, sementara sentimen negatif memperoleh *Precision* 0.92, Recall 0.96, dan F1-Score 0.94. Hal ini menunjukkan bahwa ulasan pada M-Paspor memiliki struktur kalimat yang lebih spesifik terkait kemudahan penggunaan, sehingga pola fitur bahasa lebih mudah dikenali oleh model IndoBERT. Adapun aspek *Reliability* mencatatkan akurasi aspek sebesar 0.98, namun pada sentimen negatif performa relatif lebih rendah dibandingkan kelas lainnya, dengan *Precision* 0.86, *Recall* 0.92, dan F1-Score 0.89. Secara keseluruhan, performa model pada dataset M-Paspor lebih merata dibandingkan Mobile JKN, dengan nilai F1-Score yang berkisar antara 0.89 hingga 0.99 di seluruh kelas, mengindikasikan kemampuan generalisasi model yang lebih baik pada dataset ini. Pada Gambar 5 merupakan visualisasi grafik loss model.



Gambar 5. Grafik *Loss* M-Paspor

visualisasi kurva *Loss* menunjukkan proses pembelajaran model selama 5 *epoch*. Dari angka 0.55 pada *epoch* pertama hingga menyentuh angka mendekati 0.0 pada *epoch* kelima. Hal ini menunjukkan bahwa model sangat cepat beradaptasi dan mengenali pola pada data latih. *Validation Loss* dimulai dari angka yang sudah cukup rendah sekitar 0.17 dan terus menurun secara perlahan hingga stabil di kisaran 0.10. Meskipun kedua kurva masih menunjukkan celah yang mengindikasikan gejala *overfitting*, celah tersebut jauh lebih kecil dibandingkan Mobile JKN sehingga dampaknya lebih minimal. Kondisi ini didukung oleh domain layanan keimigrasian M-Paspor yang lebih spesifik dan terfokus sehingga variasi bahasa ulasan lebih

terprediksi. Titik *Test Loss* yang berada di posisi rendah 0.08 mengonfirmasi bahwa performa model pada data uji selaras dengan data validasi, membuktikan konsistensi model dalam menangani data baru.

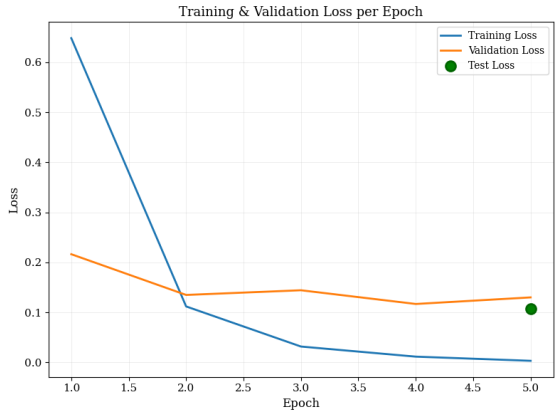
3.3. Analisis Performa Model SIGNAL

Model SIGNAL juga mencatatkan akurasi pengujian sebesar 98%, dengan karakteristik performa yang unik terutama dalam mendeteksi sentimen negatif. Pada Tabel 6 merupakan rincian evaluasi.

Tabel 6. Matrik evaluasi SIGNAL

Aspek	Sentimen	*Precicion*	*Recall*	*F1-Score*	*Accuracy* (aspek)
Usability	Positif	0.92	0.94	0.93	0.99
Usability	Negatif	0.97	1.00	0.99	0.99
Reliability	Positif	0.89	0.98	0.93	0.98
Reliability	Negatif	0.96	0.98	0.97	0.98
Efficiency	Positif	0.93	0.95	0.94	0.97
Efficiency	Negatif	0.99	1.00	0.99	0.97

Berdasarkan Tabel 6, menunjukkan kemampuan model yang sangat unggul dalam mendeteksi sentimen negatif. Aspek *Usability* sentimen negatif mencatatkan *Precision* 0.97, *Recall* 1.00, dan F1-Score 0.99, sementara aspek *Efficiency* sentimen negatif mencapai *Precision* 0.99, *Recall* 1.00, dan *F1-Score* 0.99. Hal ini mengindikasikan pengguna SIGNAL cenderung menyampaikan keluhan dengan bahasa yang tegas dan eksplisit, sehingga meminimalisir ambiguitas bagi model. Aspek *Usability* mencatatkan akurasi aspek tertinggi sebesar 0.99, diikuti aspek *Reliability* sebesar 0.98, dan aspek *Efficiency* sebesar 0.97. Meskipun akurasi aspek *Efficiency* sedikit lebih rendah dibandingkan aspek lainnya, model tetap mampu mendeteksi kedua sentimen pada aspek tersebut dengan sangat baik. Adapun nilai *Precision* terendah tercatat pada aspek *Reliability* sentimen positif sebesar 0.89, namun model tetap mampu mempertahankan performa keseluruhan yang tinggi dengan nilai F1-Score yang berkisar antara 0.93 hingga 0.99 di seluruh kelas. Pada Gambar 6 merupakan visualisasi grafik *loss* model.



Gambar 6. Grafik *Loss* SIGNAL

visualisasi kurva *loss* menunjukkan proses pembelajaran model selama 5 *epoch*. Dimulai dari angka sekitar 0.65 pada *epoch* pertama dan turun drastis hingga mendekati 0.0 pada *epoch* kelima. Penurunan ini mengindikasikan bahwa model dapat menyerap informasi dan pola dari data latih SIGNAL dengan sangat cepat. *Validation Loss* dimulai dari posisi yang cukup rendah yaitu 0.22, kemudian menurun perlahan dan stabil di kisaran angka 0.12. Meskipun selisih antara kedua kurva mengindikasikan adanya gejala *overfitting*, tingkat keparahannya tergolong ringan karena kedua kurva bergerak relatif berdekatan dan tidak menunjukkan divergensi yang signifikan. Hal ini dapat dikaitkan dengan domain administrasi kendaraan SIGNAL yang sangat spesifik sehingga pola bahasa ulasan lebih konsisten dan mudah dipelajari model. Titik *Test Loss* yang berada di posisi rendah 0.10. Posisi titik ini sedikit lebih rendah dibandingkan akhir garis validasi mengonfirmasi bahwa performa model justru bekerja lebih baik atau setidaknya sama baiknya saat menghadapi data *testing* dibandingkan saat fase validasi.

3.4. Analisis Performa Model

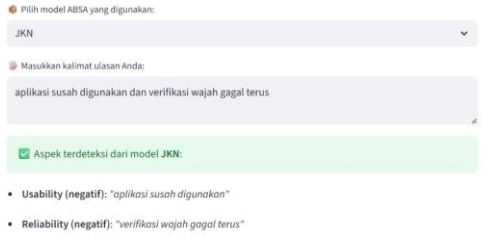
Secara keseluruhan, penerapan IndoBERT pada ketiga dataset menunjukkan hasil yang sangat memuaskan dengan rata-rata akurasi di atas 97%. Analisis perbandingan menyoroti beberapa temuan.

1. Dominasi Aspek *efficiency* pada ketiga aplikasi, aspek *Efficiency* secara konsisten menjadi aspek yang paling mudah diklasifikasikan dengan F1-*Score* lebih dari 0.97. Kata kunci seperti "lambat", "lemot", atau "cepat" memiliki makna semantik yang kuat dan minim ambiguitas.
2. Tantangan pada *reliability*: Aspek *reliability* konsisten menjadi tantangan tersulit, namun dengan pola yang berbeda. Pada Mobile JKN dan M-Paspor, kesulitan terletak pada sentimen negatif karena keluhan teknis yang seringkali memiliki kemiripan konteks dengan aspek *usability*, sedangkan pada SIGNAL kesulitan justru pada sentimen positif karena pujian umum yang seringkali dianggap sebagai indikator keandalan.
3. Dataset M-Paspor dan SIGNAL menghasilkan akurasi *testing* keseluruhan model sebesar 98%, lebih tinggi dibandingkan Mobile JKN yang mencapai 97%. Ditinjau dari akurasi per aspek, Mobile JKN mencatatkan akurasi *Usability* 0.98, *Reliability* 0.97, dan *Efficiency* 0.99; M-Paspor mencatatkan *Usability* 0.99, *Reliability* 0.98, dan *Efficiency* 0.99; sementara SIGNAL mencatatkan *Usability* 0.99, *Reliability* 0.98, dan *Efficiency* 0.97. Perbedaan ini dipengaruhi oleh variasi topik layanan kesehatan pada Mobile JKN yang lebih luas dan kompleks dibandingkan layanan paspor atau pajak kendaraan yang lebih spesifik.

4. Pola *overfitting* berbeda antar dataset pada analisis kurva *loss* mengungkap bahwa Mobile JKN menunjukkan indikasi *overfitting* ringan yang disebabkan oleh kompleksitas domain kesehatan yang luas, jumlah data pelatihan yang paling kecil, serta pengaruh *oversampling* pada kelas minoritas. Sementara itu, M-Paspor dan SIGNAL juga mengalami *overfitting* namun dengan derajat yang lebih rendah, karena domain keduanya lebih spesifik dan dataset pelatihannya lebih besar sehingga model memperoleh representasi fitur bahasa yang lebih kaya. Dengan demikian, ketiga model mengalami *overfitting* namun dengan tingkat keparahan yang berbeda-beda paling signifikan pada Mobile JKN, sedang pada SIGNAL, dan paling minimal pada M-Paspor. Meskipun demikian, nilai *Test Loss* yang rendah pada ketiga model 0.1, 0.08, dan 0.10 serta akurasi per aspek yang secara konsisten tinggi pada ketiga dataset Mobile JKN *Usability* 0.98, *Reliability* 0.97, *Efficiency* 0.99, M-Paspor *Usability* 0.99, *Reliability* 0.98, *Efficiency* 0.99, dan SIGNAL *Usability* 0.99, *Reliability* 0.98, *Efficiency* 0.97 hal ini menunjukkan bahwa kemampuan generalisasi model secara keseluruhan tetap terjaga dengan baik meskipun terjadi *overfitting* pada seluruh model.
5. Perbandingan pada penelitian terdahulu yang disebutkan pada bagian Pendahuluan, pendekatan *end-to-end* berbasis IndoBERT dalam penelitian ini secara konsisten menghasilkan akurasi yang lebih unggul. Penelitian oleh Radiena & Nugroho, 2023yang menggunakan SVM pada aplikasi KAI *Access* memperoleh akurasi tertinggi 94,73%, sedangkan Widiansyah et al., 2024 yang mengombinasikan LDA dengan IndoBERT pada M-Paspor mencapai akurasi 94%. Sementara itu, model yang dikembangkan dalam penelitian ini mencatatkan akurasi rata-rata di atas 97% pada ketiga dataset (Mobile JKN 97%, M-Paspor 98%, SIGNAL 98%). Keunggulan ini dapat dijelaskan oleh tiga faktor utama. Pertama, arsitektur transformer pada IndoBERT mampu menangkap dependensi kontekstual jarak jauh antar token melalui mekanisme *self-attention*, sesuatu yang tidak dapat dilakukan oleh SVM yang bergantung pada fitur permukaan. Kedua, *pendekatan end-to-end* yang mengintegrasikan deteksi aspek dan klasifikasi sentimen dalam satu alur menghilangkan akumulasi kesalahan yang kerap terjadi pada pendekatan multi-tahap seperti LDA dengan kombinasi IndoBERT. Ketiga, IndoBERT yang telah *melakukan pre-training* pada korpus bahasa Indonesia yang besar memiliki representasi semantik yang lebih kaya dibandingkan model umum, sehingga proses *fine-tuning* pada domain layanan publik dapat dilakukan secara lebih efisien dan akurat.

3.5. Implementasi Model

Model terbaik diimplementasikan pada *Streamlit* pengujian fungsionalitas sistem dilakukan dengan memasukkan ulasan yang mengandung sentimen multi-aspek. Analisis teks memungkinkan pengguna untuk menganalisis sentimen berbasis aspek dari teks yang dimasukkan secara langsung. Terdapat sebuah kotak input tempat pengguna dapat mengetik atau menempelkan teks yang ingin dianalisis. Setelah teks dimasukkan, pengguna dapat memilih aplikasi pelayanan masyarakat terlebih dahulu dengan opsi Mobile JKN, M-Paspor, dan SIGNAL untuk memproses teks dan mendapatkan hasil analisis sentimen, untuk mengkategorikan sentimen dalam teks menjadi berdasarkan aspek dan sentimen yang telah ditentukan. Pada Gambar 8 merupakan contoh implementasi sistem.



Gambar 7. Implementasi Sistem

Berdasarkan Gambar 8 pengujian sistem dengan ulasan "aplikasi susah digunakan dan verifikasi wajah gagal terus" menunjukkan keberhasilan implementasi model. Sistem yang dibangun menggunakan *Streamlit* ini mampu melakukan deteksi multi-aspek dari satu kalimat ulasan untuk aplikasi Mobile JKN. Hasil deteksi menunjukkan bahwa model berhasil mengidentifikasi dua aspek berbeda beserta sentimen negatifnya secara akurat yaitu *Usability* (negatif) terdeteksi dari frasa "aplikasi susah digunakan" dan *Reliability* (negatif) terdeteksi dari frasa "verifikasi wajah gagal terus".

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa *Aspect-Based Sentiment Analysis* (ABSA) telah berhasil dirancang dan diimplementasikan melalui kombinasi arsitektur model IndoBERT dengan pendekatan *token classification*. Alur sistem yang dibangun mulai dari pengumpulan data, pra-pemrosesan, pelabelan, penyeimbangan data, hingga *fine-tuning* terbukti efektif dalam mengekstraksi aspek dan sentimen dari ulasan pengguna secara simultan. Performa model yang dikembangkan menunjukkan kinerja yang sangat solid, dibuktikan dengan capaian *testing accuracy* yang tinggi pada ketiga aplikasi, yaitu 97% untuk Mobile JKN, serta 98% untuk M-Paspor dan SIGNAL. Tingginya nilai metrik precision, recall,

dan F1-score turut menegaskan kemampuan model dalam mengklasifikasikan kelas aspek sentimen dengan baik. Meskipun analisis kurva *loss* mengungkap bahwa ketiga model mengalami *overfitting* dengan tingkat keparahan berbeda—paling signifikan pada Mobile JKN akibat kompleksitas domain dan keterbatasan data, serta lebih ringan pada M-Paspor dan SIGNAL—nilai *Test Loss* yang rendah pada ketiga model membuktikan bahwa kemampuan generalisasi terhadap data baru tetap terjaga dengan baik. Analisis perbandingan antar dataset mengungkapkan karakteristik unik pada setiap aplikasi. Aspek *Efficiency* secara konsisten menjadi aspek yang paling mudah diklasifikasi dengan F1-Score di atas 0.94, sedangkan aspek *Reliability* menjadi tantangan tersulit dengan pola kesalahan yang bervariasi akibat kompleksitas domain dan variasi ekspresi pengguna. Dari sisi substansi ulasan, model berhasil mengidentifikasi bahwa aspek *Reliability* secara konsisten menjadi sumber keluhan utama dengan dominasi sentimen negatif di ketiga aplikasi. Sebaliknya, aspek *Usability* teridentifikasi sebagai kekuatan utama layanan yang paling banyak mendapatkan sentimen positif dari pengguna.

PUSTAKA

Badan Pusat Statistik. (2024). *Statistik Telekomunikasi Indonesia 2023*.

Chandradev, V., Made, I., Dwi Suarjaya, A., Putu, I., & Bayupati, A. (2023). *Chandradev, Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning Bert 107 Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning Bert*.

Filemon Haganta Kaban, A., & Yudistira, N. (2021). *Analisis Sentimen Aplikasi E-Goverment Berdasarkan Ulasan Pengguna Menggunakan Metode Maximum Entropy Dan Seleksi Fitur Mutual Information* (Vol. 5, Number 4). http://j-ptiik.ub.ac.id

Hakim, G., Fatyanosa, T. N., & Widodo, A. W. (2024). *Analisis Sentimen Masyarakat Terhadap Kereta Cepat Whoosh Pada Platform X Menggunakan Indobert* (Vol. 8, Number 10). http://j-ptiik.ub.ac.id

Hisyam Pradhana, A., Daniati, E., Muzaki, M. N., & Informasi, S. (2025). Penerapan Bi-Lstm Untuk Named Entity Recognition Pada Teks Bahasa Indonesia. *Ijcsr: The Indonesian Journal Of Computer Science Research E*. https://doi.org/10.37905

Husada, H. C., & Paramita, A. S. (2021). Analisis Sentimen Pada Maskapai Penerbangan Di Platform Twitter Menggunakan Algoritma Support Vector Machine (Svm). *Teknika*, *10*(1), 18–26. https://doi.org/10.34148/teknika.v10i1.311

Nur Karimah, & Anna Baita. (2024). Multi-Aspect Sentiment Analysis Pada Review Film Menggunakan Metode Bidirectional Encoder Representations From Transformers (Bert). *Komputika : Jurnal Sistem Komputer*, *13*(1), 63–72. https://doi.org/10.34010/komputika.v13i1.11098

Nuryadi, D., Metandi, F., Alam Hadiwijaya, N., Zainul Rohman, M., Hartanto, S., Syafrizal, A., Yadie Teknologi Informasi, E., Negeri Samarinda Jl Cipto Mangunkusumo, P., Seberang, S., & Timur, K. (2025). Fine Tuning Indobert Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Tiket.Com Di Google Play Store. In *Jurnal Mahasiswa Teknik Informatika)* (Vol. 9, Number 2).

Nyoman Saputra Wahyu Wijaya, I., Agus Seputra, K., & Putu Novita Puspa Dewi, N. (2025). Fine Tunning Model Indobert Untuk Analisis Sentimen Berita Pariwisata Indonesia. *Jurnal Pendidikan Teknologi Dan Kejuruan*, *22*(2). https://www.detik.com/search/searchall?query=Wisata&Siteid=3&Sortby=Time&Fromdatex=01/01/2022&

Radiena, G., & Nugroho, A. (2023). Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Kai Access Menggunakan Metode Support Vector Machine. In *Jurnal Pendidikan Teknologi Informasi (Jukanti)* (Number 1).

Rahman, I. F., Hasanah, A. N., & Heryana, N. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi Samsat Digiital Nasional (Signal) Dengan Menggunakan Metode Naïve Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(2). https://doi.org/10.23960/jitet.v12i2.4073

Safitri, T., Umaidah, Y., & Maulana, I. (2023). Analisis Sentimen Pengguna Twitter Terhadap Bts Menggunakan Algoritma Support Vector Machine. In *Journal Of Applied Informatics And Computing (Jaic)* (Vol. 7, Number 1). http://jurnal.polibatam.ac.id/index.php/jaic

Sari, P. W. S., Firmansyah, F., & Kadafi, A. R. K. (2025). Perbandingan Algoritma Random Forest Dan Naïve Bayes Dalam Menganalisis Sentimen Ulasan Pada Produk Skincare Lokal Di Media Sosial Tiktok. *Jurnal Informatika Dan Teknik Elektro Terapan*, *13*(3s1). https://doi.org/10.23960/jitet.v13i3s1.8150

Truşcă, M. M., & Frasincar, F. (2023). Survey On Aspect Detection For Aspect-Based Sentiment Analysis. *Artificial Intelligence Review*, *56*(5), 3797–3846. https://doi.org/10.1007/s10462-022-10252-y

Widiansyah, M., Frazna Az-Zahra, F., & Pambudi, A. (2024). Fine-Tuning Model Indobert (Indonesian Bidirectional Encoder Representations From Transformers) Untuk Analisis Sentimen Berbasis Aspek Pada Aplikasi M-Paspor. In *Journal Of Informatic Engineering (Joutica)*. https://doi.org/10.30736/informatika.v9i2.1310

Zhou, J., Huang, J. X., Chen, Q., Hu, Q. V., Wang, T., & He, L. (2019). Deep Learning For Aspect-Level Sentiment Classification:

Survey, Vision, And Challenges. In *Ieee Access* (Vol. 7, Pp. 78454–78483). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/access.2019.2920075