Jurnal Informatika
E-ISSN: 2528-0090
P-ISSN: 2252-9160
Vol. 15, No. 1, Juni 2026, Page. 60~72
https://ejournal.unidayan.ac.id/index.php/JIU
DOI: https://doi.org/10.55340/jiu.v15i1.2683

Evaluasi IndoBERT Pra-Latih untuk Klasifikasi Sentimen Ulasan Wisata dengan Pendekatan *Zero-Shot Inference*: Studi Kasus Kota Baubau

Evaluation of Pre-trained IndoBERT for Tourism Review Sentiment Classification Using Zero-Shot Inference: A Case Study in Baubau City

Nurul Hidayah*[1], Dodiman[2], Asniati[3], La Ode Dwiyan Pramono Darmin[4]

[1,2,3]Program Studi Teknik Informatika, Universitas Dayanu Ikhsanuddin, Indonesia
[4]Program Studi Administrasi Negara, Universitas Dayanu Ikhsanuddin, Indonesia
Email: [1]nurul.hyh@gmail.com

Article Info:	Received	Revised	Accepted	Published:
	09 Mei 2026	09 Mei 2026	05 Juni 2026	05 Juni 2026

Abstrak

Ulasan wisatawan di platform digital merupakan sumber data persepsi publik yang berharga untuk mendukung pengelolaan pariwisata berbasis bukti. Namun, ketiadaan dataset berlabel domain spesifik menjadi kendala utama penerapan model klasifikasi sentimen pada destinasi lokal. Penelitian ini bertujuan untuk mengevaluasi kemampuan model IndoBERTweet dalam mengklasifikasikan sentimen ulasan wisatawan tanpa fine-tuning tambahan pada domain target (direct inference), dengan studi kasus pada Benteng Keraton Buton dan Pantai Nirwana di Kota Baubau. Sebanyak 918 ulasan dikumpulkan dari Google Maps melalui web scraping dan diproses melalui case folding, cleaning, normalisasi teks, dan penggabungan kata negasi untuk mengklasifikasikan sentimen positif, negatif, dan netral. Kinerja model dievaluasi menggunakan confusion matrix dengan metrik akurasi, weighted precision, weighted recall, dan weighted F1-Score. Hasil penelitian menunjukkan akurasi 0,58, weighted precision 0,79, weighted recall 0,58, dan weighted F1-Score 0,63 (fair), dengan performa terendah pada kelas netral (F1: 0,27) yang mencerminkan kesulitan model dalam mengenali ekspresi sentimen ambigu di luar domain pelatihannya. Analisis tematik menunjukkan bahwa sentimen negatif didominasi isu kebersihan di Pantai Nirwana dan ketidakpuasan tarif di Benteng Keraton Buton. Model IndoBERTweet layak digunakan sebagai instrumen awal analisis sentimen wisata lokal tanpa anotasi data domain, namun peningkatan performa terutama pada kelas netral dan negatif memerlukan fine-tuning dengan data ulasan wisata berbahasa Indonesia dari destinasi serupa. Temuan tematik penelitian ini selanjutnya dapat dimanfaatkan sebagai dasar rekomendasi kebijakan pengelolaan pariwisata Kota Baubau berbasis bukti.

Kata Kunci: *Analisis sentimen; Deep learning; Google Maps; IndoBERT; Klasifikasi teks; Pariwisata.*

Jurnal Informatika
E-ISSN: 2528-0090
P-ISSN: 2252-9160
Vol. 15, No. 1, Juni 2026, Page. 60~72
https://ejournal.unidayan.ac.id/index.php/JIU
DOI: https://doi.org/10.55340/jiu.v15i1.2683

Abstract

Tourist reviews on digital platforms provide valuable public perception data for evidence-based tourism management. However, the absence of domain-specific labeled datasets constrains the deployment of sentiment classification models for local destinations. This study evaluates the IndoBERTweet model in classifying tourist review sentiment without additional fine-tuning on the target domain (direct inference), with a case study of Benteng Keraton Buton and Pantai Nirwana in Baubau City. A total of 918 Google Maps reviews were collected via web scraping and preprocessed through case folding, cleaning, text normalization, and negation merging to classify positive, negative, and neutral sentiment. Performance was evaluated using a confusion matrix measuring accuracy, weighted precision, weighted recall, and weighted F1-Score. Results indicate an accuracy of 0.58, weighted precision of 0.79, weighted recall of 0.58, and weighted F1-Score of 0.63 (fair). The neutral class yielded the lowest performance (F1: 0.27), reflecting the model's limited capacity to recognize ambiguous expressions beyond its training domain. Thematic analysis reveals that negative sentiment is associated with sanitation concerns at Pantai Nirwana and dissatisfaction with admission fees at Benteng Keraton Buton. These findings confirm that IndoBERTweet is applicable as a preliminary sentiment analysis tool without domain annotation, though performance improvement, particularly for neutral and negative classes, requires fine-tuning on Indonesian tourism review datasets. These thematic findings provide an empirical basis for evidence-based tourism management policy recommendations in Baubau City.

Keywords: *Deep learning; Google Maps; IndoBERT; Sentiment analysis; Text classification; Tourism*

1. PENDAHULUAN

Pariwisata merupakan sektor strategis dalam pembangunan daerah melalui peningkatan ekonomi, pelestarian budaya, dan penguatan identitas wilayah [1], [2], yang diakui sebagai bagian integral dari kebijakan pembangunan lokal hingga nasional [3], [4]. Keberhasilan pembangunan pariwisata dapat ditentukan oleh kualitas kawasan wisata dan persepsi wisatawan terhadap pelayanan yang diterima sebagai indikator penting bagi pengelola dan pemerintah daerah [5]. Dalam kontek ini, Kota Baubau memiliki potensi wisata alam, sejarah, dan budaya yang beragam, mulai dari wisata bahari, situs Kesultanan Buton, hingga tradisi lokal yang menjadikannya menarik dikaji dalam konteks pembangunan pariwisata daerah.

Seiring berkembangnya pariwisata digital, persepsi wisatawan semakin terdokumentasi melalui ulasan daring di platform seperti Google Maps sebagai bentuk *electronic word of mouth* [6], ulasan ini berpotensi menjadi dasar pengelolaan pariwisata yang responsif dan berbasis bukti (*evidence-based policy*) [7]. Namun, ulasan dari destinasi lokal seperti Kota Baubau menghadirkan tantangan tersendiri karena nuansa bahasa daerah, *slang* wisatawan, dan konteks budaya yang sering membingungkan model NLP konvensional [8].

Perkembangan *Natural Language Processing* (NLP) mendorong penggunaan model *deep learning* seperti IndoBERT yang dilatih pada korpus besar bahasa Indonesia, sehingga mampu menangkap struktur bahasa, konteks semantik, dan variasi linguistik sehari-hari [8], [9], [10],

[11], [12]. Model berbasis *transformer* ini juga dipandang sebagai instrumen pendukung pengambilan keputusan berbasis data dalam tata kelola pariwisata daerah [13].

Beberapa penelitian membuktikan efektivitas IndoBERT, di antaranya mencapai akurasi 85,15% yang meningkat menjadi 86,03% dengan kombinasi *confident learning* [14], serta akurasi 97,71% pada analisis sentimen publik [15]. Penelitian lain juga telah membuktikan bahwa model BERT dan variannya efektif dalam mencapai hasil terkini (*state-of-the-art*) pada berbagai tugas NLP [16], [17], [18], [19]. Sedangkan penelitian analisis sentimen umumnya masih mengandalkan metode konvensional yang belum disesuaikan dengan karakteristik bahasa Indonesia yang memiliki akurasi 83,8 % untuk analisis sentimen wisata [20], 61%-74% untuk analisis sentimen terhadap data Twitter [21], [22], dengan fokus yang lebih banyak pada performa klasifikasi dibanding pemanfaatan hasilnya untuk pengelolaan pariwisata daerah. Penggunaan ulasan Google Maps dari destinasi lokal seperti Kota Baubau juga belum banyak dieksplorasi.

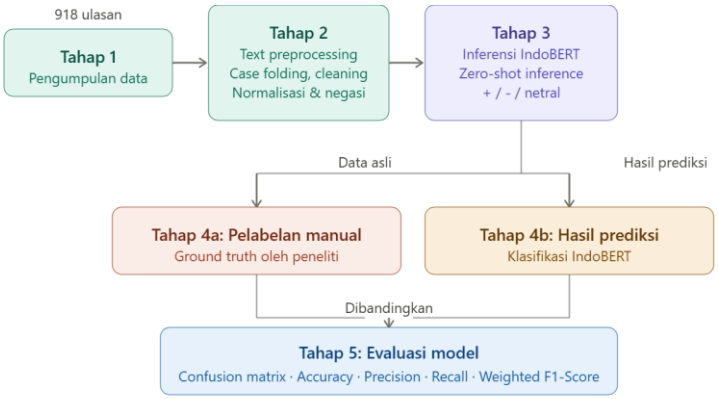
Penerapan *fine-tuning* pada destinasi wisata lokal terkendala oleh ketiadaan *dataset* berlabel, yang merupakan hambatan umum dalam pengembangan model klasifikasi sentimen berbasis ulasan pariwisata [23]. Kondisi ini mendorong eksplorasi pendekatan *direct inference*, yakni penggunaan model yang telah di-*fine-tune* pada satu domain dan diterapkan langsung pada domain berbeda tanpa adaptasi tambahan. Meskipun pendekatan serupa telah menunjukkan hasil yang menjanjikan pada konteks ulasan wisata [25], [26], keterbatasannya pada bahasa Indonesia informal perlu diakui secara eksplisit. Bahasa Indonesia pada ulasan wisata lokal mengandung kosakata daerah, singkatan tidak baku, dan ekspresi ambigu yang tidak sepenuhnya terwakili dalam korpus pelatihan berbasis media sosial [8], [23]. Studi adaptasi domain pada IndoBERT menunjukkan bahwa kesenjangan antara domain pelatihan dan domain target berkontribusi signifikan terhadap penurunan performa, terutama pada kelas sentimen dengan frekuensi rendah [12], [23]. Penelitian ini karena itu memposisikan *direct inference* bukan sebagai solusi optimal, melainkan sebagai *baseline* empiris yang terukur untuk menentukan sejauh mana adaptasi domain diperlukan dalam sebuah kontribusi metodologis yang belum dikaji secara eksplisit pada konteks ulasan pariwisata berbahasa Indonesia.

Oleh karena itu, penelitian ini bertujuan untuk: (1) mengevaluasi kemampuan model IndoBERTweet dalam mengklasifikasikan sentimen ulasan wisatawan Kota Baubau melalui *direct inference* tanpa adaptasi domain tambahan; (2) mengidentifikasi faktor linguistik yang menjadi batas kemampuan generalisasi model antar domain; serta (3) mengekstrak pola sentimen tematik dari ulasan wisatawan sebagai dasar rekomendasi kebijakan pengelolaan pariwisata berbasis bukti. Penelitian ini menjawab dua pertanyaan utama: (1) Sejauh mana model IndoBERTweet yang telah di-*fine-tune* pada domain media sosial dapat digeneralisasi ke domain ulasan wisata lokal berbahasa Indonesia tanpa adaptasi tambahan, dan faktor apa yang membatasi kemampuan generalisasinya? (2) Pola sentimen tematik apa yang dapat diekstrak dari ulasan wisatawan Kota Baubau, dan bagaimana temuan tersebut dapat menginformasikan kebijakan pengelolaan pariwisata berbasis bukti?

2. METODE

2.1. Tahapan Penelitian

Vol. 15, No. 1, Juni 2026, Page. 60~72
https://ejournal.unidayan.ac.id/index.php/JIU
DOI: https://doi.org/10.55340/jiu.v15i1.2683

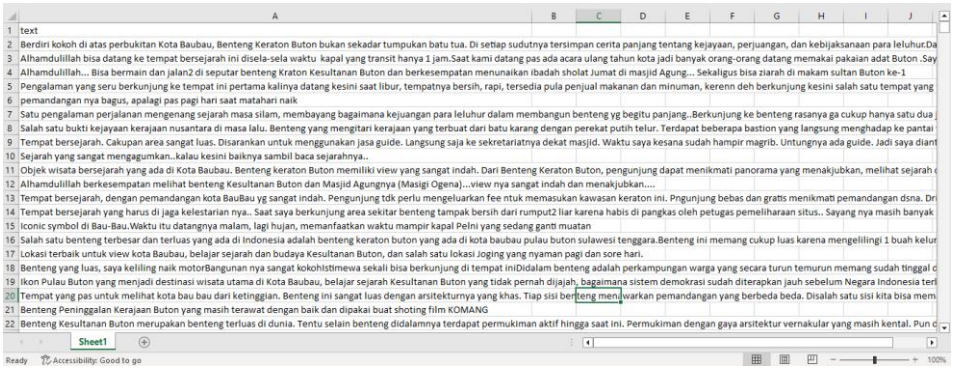


Gambar 1. Alur Penelitian

Penelitian ini terdiri dari lima tahapan sebagaimana digambarkan pada Gambar 1. Tahap pertama adalah pengumpulan data melalui *web scraping* ulasan wisatawan dari Google Maps. Tahap kedua adalah *text preprocessing* meliputi *case folding*, *cleaning*, normalisasi teks, dan penggabungan kata negasi. Tahap ketiga adalah inferensi menggunakan model IndoBERT pra-latih untuk mengklasifikasikan sentimen ke dalam tiga kelas: positif, negatif, dan netral. Tahap keempat adalah pelabelan manual oleh peneliti sebagai *ground truth* untuk keperluan evaluasi. Tahap kelima adalah evaluasi kinerja model menggunakan *confusion matrix* dengan mengukur akurasi, *precision*, *recall*, dan *F1-Score*.

2.2. Sumber Data

Data bersumber dari ulasan wisatawan Google Maps pada destinasi Benteng Keraton Buton dan Pantai Nirwana di Kota Baubau, dikumpulkan menggunakan *web scraper* Apify dengan format Excel dan ditampilkan pada Gambar 2.



Gambar 2. Data Hasil Scraping Review Google Maps

Total 918 ulasan digunakan sebagai input inferensi IndoBERT tanpa *fine-tuning*, mengingat model merupakan pra-latih yang telah dilatih pada korpus besar bahasa Indonesia [12]. Distribusi data ditampilkan pada Tabel 1.

Tabel 1. Distribusi Data Berdasarkan Label Kelas Sentimen

Label Kelas	Jumlah Data	Proporsi
Positif	682	74,3%
Netral	109	11,9%
Negatif	127	13,8%
Total	918	100%

Data menunjukkan ketidakseimbangan kelas (*class imbalance*) yang bersifat alami, dengan kelas positif mendominasi sebesar 682 ulasan (74,3%), diikuti negatif 127 ulasan (13,8%), dan netral 109 ulasan (11,9%). Kondisi ini mencerminkan kecenderungan alami wisatawan yang merasa puas untuk lebih aktif menuliskan pengalamannya pada platform digital [20]. Mengingat ketidakseimbangan kelas ini berpotensi memengaruhi performa klasifikasi, penelitian ini tidak menerapkan teknik penyeimbangan data seperti *oversampling* atau *undersampling* karena tujuan penelitian adalah mengevaluasi kemampuan generalisasi model pada kondisi data riil yang tidak dimanipulasi. Oleh karena itu, *weighted* F1-*Score* dipilih sebagai metrik evaluasi utama karena memperhitungkan proporsi tiap kelas dalam perhitungannya, sehingga lebih representatif dibandingkan akurasi pada kondisi data tidak seimbang.

2.3. Pelabelan Data (*Manual Labeling*)

Pelabelan data dilakukan secara manual oleh peneliti ke dalam tiga kelas yaitu positif (mengandung kepuasan atau pujian), negatif (mengandung kekecewaan atau kritik), dan netral (informatif tanpa polaritas yang jelas). Hasil pelabelan digunakan semata-mata sebagai *ground truth* untuk evaluasi, bukan sebagai data pelatihan, konsisten dengan pendekatan *zero-shot inference* yang tidak memerlukan data berlabel sebagai input pelatihan [12]. Contoh hasil pelabelan terhadap data ulasan ditampilkan pada Tabel 2 berikut.

Tabel 2. Contoh Hasil Pelabelan Data Ulasan

No	Teks Ulasan	Label
1	“Tempatnya bagus, pemandangannya indah dan bersih”	Positif
2	“Fasilitasnya kurang memadai, toilet kotor dan tidak terawat”	Negatif
3	“Benteng bersejarah yang terletak di pusat kota Baubau”	Netral

2.4. *Text Preprocessing*

Data mentah hasil *scraping* mengandung berbagai *noise* seperti URL, karakter khusus, dan singkatan tidak baku, sehingga diperlukan tahap *text preprocessing* sebelum dimasukkan ke dalam pipeline model [27]. Tahapan *preprocessing* yang dilakukan dalam penelitian ini meliputi (a) *case folding* yaitu konversi seluruh teks menjadi huruf kecil; (b) penghapusan URL, angka, simbol, dan karakter tidak relevan menggunakan *regex*; (c) penghapusan spasi berlebih; (d) normalisasi kata tidak baku menggunakan kamus alay [ref baru] yang terdiri dari 3.592 entri pasangan kata tidak baku dan padanan bakunya, bersumber dari repositori nasalsabila/kamus-alay (https://github.com/nasalsabila/kamus-alay), untuk meningkatkan konsistensi teks informal; dan (e) penggabungan kata negasi seperti "tidak", "ga", dan "nggak" dengan kata berikutnya menggunakan *underscore* (contoh: "tidak_bagus") sebagai perlakuan eksplisit terhadap ekspresi negasi. Perlu dicatat bahwa efektivitas teknik ini pada model berbasis transformer yang sudah memiliki representasi kontekstual bidireksional belum diuji secara empiris dalam penelitian ini; teknik ini diterapkan sebagai langkah konservatif mengikuti praktik preprocessing pada studi sentimen berbahasa Indonesia sebelumnya, dan dampak spesifiknya terhadap performa model memerlukan *ablation study* pada penelitian lanjutan.

Berikut ini merupakan Tabel 3 yang menggambarkan contoh proses *Preprocessing* dalam teks ulasan wisata:

Vol. 15, No. 1, Juni 2026, Page. 60~72
https://ejournal.unidayan.ac.id/index.php/JIU
DOI: https://doi.org/10.55340/jiu.v15i1.2683

Tabel 3. Contoh Hasil *Text Preprocessing*

Tahap	Teks
Sebelum *preprocessing*	"Tempatnya BAGUS banget!! Recommended bgt, gak nyesel kesini 😊 https://t.co/xxx"
Setelah *case folding*	"tempatnya bagus banget!! Recommended bgt, gak nyesel kesini 😊"
Setelah *cleaning*	"tempatnya bagus banget recommended bgt gak nyesel kesini"
Setelah normalisasi	"tempatnya bagus banget recommended banget tidak_nyesel kesini"

2.5. Model IndoBERT untuk Klasifikasi Sentimen

Penelitian ini menggunakan model IndoBERTweet (Aardiiiiy/indobertweet-base-Indonesian-sentiment-analysis) melalui *framework* HuggingFace Transformers. Model ini merupakan turunan IndoBERT dengan *domain-adaptive pretraining* pada korpus besar bahasa Indonesia dari media sosial, sehingga mampu memahami variasi bahasa informal dan ekspresi sehari-hari [12]. Pendekatan *zero-shot inference* diterapkan tanpa *fine-tuning* pada data pariwisata Kota Baubau, dengan tujuan mengevaluasi kemampuan generalisasi model pra-latih dalam mentransfer pengetahuan linguistik ke domain baru tanpa adaptasi tambahan.

Pipeline klasifikasi diimplementasikan menggunakan library transformers dari HuggingFace dan *framework* PyTorch. Konfigurasi model ditampilkan pada Tabel 4.

Tabel 4. Konfigurasi Model IndoBERT

Parameter	Nilai
Model	`indobertweet-base-Indonesian-sentiment-analysis`
Framework	PyTorch + HuggingFace Transformers
Jumlah Kelas	3 (Positif, Negatif, Netral)
Panjang Maksimum Token	128 token
Mode Inferensi	*Zero-shot* (tanpa *fine-tuning*)

Panjang maksimum token ditetapkan sebesar 128 mengacu pada kapasitas optimal model *indobertweet-base* yang efisien secara komputasi sekaligus memadai untuk sebagian besar ulasan wisata pendek hingga menengah. Ulasan yang melebihi 128 token akan mengalami pemotongan (*truncation*) pada bagian akhir teks, berpotensi menghilangkan informasi sentimen di segmen penutup ulasan khususnya pada kelas negatif dan netral yang cenderung menggunakan kalimat panjang dengan nuansa kontekstual [16]. Hal ini menjadi salah satu faktor yang dapat memengaruhi performa klasifikasi, terutama pada kedua kelas tersebut.

2.6. Evaluasi Model

Evaluasi dilakukan dengan membandingkan hasil prediksi IndoBERT terhadap label manual menggunakan *confusion matrix* 3×3. Empat metrik dihitung dari *confusion matrix* dengan persamaan (1-4) berikut [21]:

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

(b) $Precision = \frac{TP}{(TP+FP)}$ (2)

(c) $Recall = \frac{TP}{(TP+FN)}$ (3)

(d) $F1 - score = 2x\frac{(PrecisionxRecall)}{(Precision+Recall)}$ (4)

Vol. 15, No. 1, Juni 2026, Page. 60~72
https://ejournal.unidayan.ac.id/index.php/JIU
DOI: https://doi.org/10.55340/jiu.v15i1.2683

Mengingat adanya ketidakseimbangan kelas pada data, *weighted F1-Score* digunakan sebagai metrik evaluasi utama karena memperhitungkan proporsi masing-masing kelas dalam perhitungannya. Selain itu, visualisasi *word cloud* digunakan untuk memperlihatkan kata-kata dominan pada kelas negatif dan positif secara kualitatif.

3. HASIL

3.1. Hasil *Text Preprocessing*

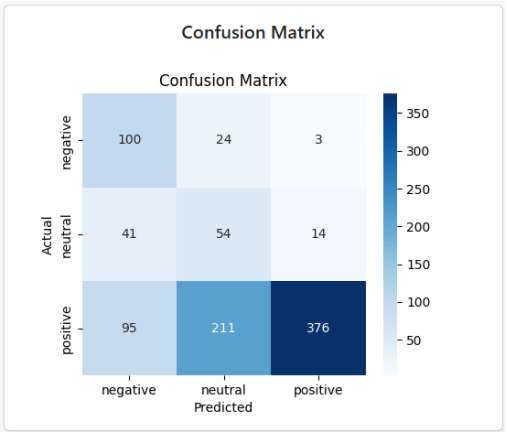
Melalui tahapan *preprocessing*, data mentah berhasil ditransformasi menjadi teks yang bersih dan terstandarisasi. *Case folding* menyeragamkan kapitalisasi, *cleaning* menghilangkan *noise*, normalisasi *alay dictionary* menyeragamkan penulisan informal (contoh: "bgtt" → "bagus"), dan penggabungan kata negasi (contoh: "tidak_bagus") mempertahankan konteks semantik negatif. Contoh hasil *preprocessing* ditampilkan pada Gambar 3.

serba bayar pokoknya buat kalian jangan pernah datang kesini lagi	serba bayar pokoknya buat kalian jangan pernah datang kesini lagi
kepada pemerintah kota tolong ditata object wisata ini sampah banyak tenda pedagang tidak teratur wc dan tempat bilas tidak memadai air nya tidak bersih kami ingin object wisata kampung halaman kami bisa lebih dioptimalkan agar bisa dikunjungi oleh wisatawan diluar sumatera barat	kepada pemerintah kota tolong ditata object wisata ini sampah banyak tenda pedagang tidak_teratur wc dan tempat bilas tidak_memadai air nya tidak_bersih kami ingin object wisata kampung halaman kami bisa lebih dioptimalkan agar bisa dikunjungi oleh wisatawan diluar sumatera barat

Gambar 3. Hasil Text Preprocessing pada Sistem

3.2. Hasil Klasifikasi Model IndoBERT

Hasil prediksi model terhadap 918 ulasan dibandingkan dengan label manual (*ground truth*) menggunakan *confusion matrix* 3×3 sebagaimana ditampilkan pada Gambar 4.



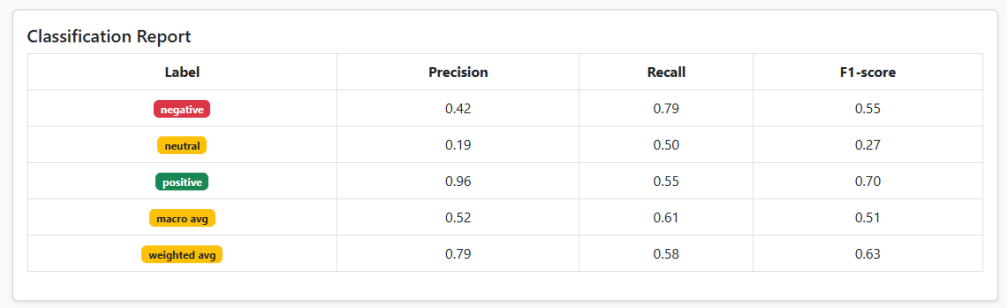
Gambar 4. Hasil *Confusion Matrix* Klasifikasi Sentimen IndoBERT

Berdasarkan Gambar 4, hasil *confusion matrix* menunjukkan bahwa model mengklasifikasikan dengan benar 100 dari 127 ulasan negatif, 54 dari 109 ulasan netral, dan 376 dari 682 ulasan positif. Kelas positif memperoleh prediksi terbanyak namun juga menghasilkan kesalahan terbesar, dengan 211 ulasan positif salah diprediksi sebagai netral dan 95 sebagai negatif. Kelas netral merupakan kelas dengan performa terlemah. Hasil ini mengindikasikan kesulitan model dalam membedakan sentimen ambigu tanpa konteks domain spesifik.

Vol. 15, No. 1, Juni 2026, Page. 60~72
https://ejournal.unidayan.ac.id/index.php/JIU
DOI: https://doi.org/10.55340/jiu.v15i1.2683

3.3. Hasil Evaluasi Kinerja Model

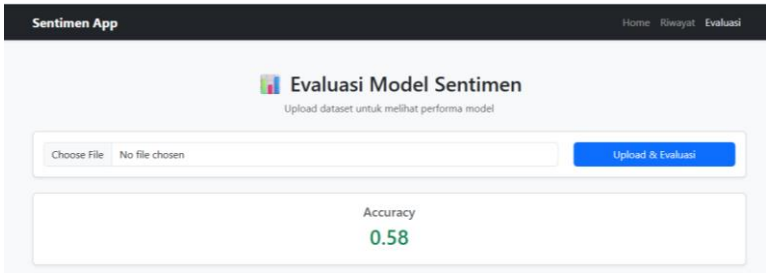
Hasil evaluasi kinerja model IndoBERT menggunakan metrik *precision*, *recall*, dan *F1-Score* per kelas ditampilkan pada Gambar 4 berikut.



Gambar 5. *Classification Report* Model IndoBERT pada Sistem

Berdasarkan Gambar 5, kinerja model bervariasi antarkelas. Kelas positif memiliki *precision* tertinggi (0,96) namun *recall* rendah (0,55), mengindikasikan model jarang salah saat memprediksi positif tetapi melewatkan banyak ulasan positif yang sebenarnya. Kelas negatif menunjukkan pola sebaliknya *recall* tinggi (0,79) namun *precision* rendah (0,42) menunjukkan model terlalu agresif dalam mendeteksi sentimen negatif. Kelas netral memiliki performa terrendah (F1: 0,27) akibat kesulitan model mengenali ekspresi sentimen yang ambigu. *Weighted F1-Score* 0,63 mengindikasikan kemampuan klasifikasi cukup (*fair*) tanpa *fine-tuning*, konsisten dengan temuan sebelumnya bahwa model pra-latih tanpa adaptasi domain cenderung menghasilkan performa lebih rendah dibanding model yang telah di-*fine-tuning* [12].

Kemudian, terdapat hasil akurasi keseluruhan sebesar 0,58 (58%) yang ditunjukkan pada Gambar 6.

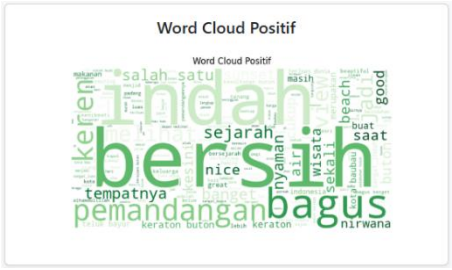


Gambar 6. Nilai Akurasi Model

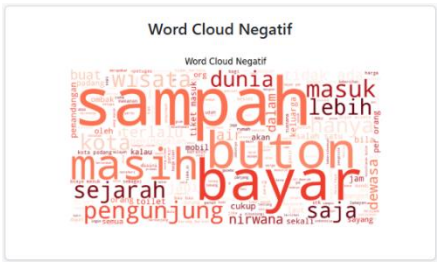
Nilai akurasi 0,58 tidak dapat diinterpretasikan secara terpisah karena pada data yang tidak seimbang, akurasi cenderung bias terhadap kelas mayoritas [21]. Oleh karena itu, *weighted F1-Score* 0,63 digunakan sebagai acuan utama, yang mencerminkan kinerja model secara lebih proporsional dan mengindikasikan kemampuan klasifikasi pada tingkatan cukup (*fair*) tanpa *fine-tuning* tambahan.

Vol. 15, No. 1, Juni 2026, Page. 60~72
https://ejournal.unidayan.ac.id/index.php/JIU
DOI: https://doi.org/10.55340/jiu.v15i1.2683

3.4. Analisis *Word Cloud*



Gambar 7. Word Cloud Positif



Gambar 8. Word Cloud Negatif

Gambar 9. Word Cloud Netral

Visualisasi *word cloud* dibuat berdasarkankata dominan pada masing-masing kelas sentimen setelah *preprocessing*. Pada kelas positif, kata dominan "indah", "bersih", "bagus", "pemandangan", dan "sejarah" mencerminkan apresiasi wisatawan terhadap keindahan alam, kebersihan, dan nilai historis destinasi. Kelas Netral memiliki dominan kata Keraton, "kesultanan" dan "cukup". Pada kelas negatif, kata dominan "sampah", "bayar", dan "masuk" mengindikasikan keluhan utama terkait kebersihan dan biaya kunjungan temuan yang berharga bagi pengelola pariwisata dalam memprioritaskan perbaikan layanan [20].

4. PEMBAHASAN

Model IndoBERT pra-latih mampu mengklasifikasikan sentimen ulasan wisata dengan kinerja cukup (*weighted* F1-*Score* = 0.63) tanpa *fine-tuning* pada domain pariwisata. Hasil ini sejalan dengan karakteristik umum pendekatan *zero-shot inference*, di mana model memanfaatkan representasi semantik general yang diperoleh selama pra-pelatihan tanpa adaptasi terhadap domain target [25]. Secara teoretis, kinerja yang diperoleh mencerminkan keterbatasan model pra-latih ketika diterapkan di luar domain pelatihannya, sebagaimana ditunjukkan bahwa adaptasi kosakata dan data domain spesifik berkontribusi signifikan terhadap peningkatan performa model berbasis IndoBERT [12], termasuk pada domain ulasan wisata berbahasa Indonesia [23].

Perbedaan kinerja antarkelas, terutama antara kelas positif (F1: 0.70) dan netral (F1: 0.27), dipengaruhi oleh dua faktor utama. Pertama, *class imbalance*, di mana kelas positif mendominasi dengan 682 data (74,3%) sehingga lebih mudah dikenali model, sementara kelas netral dengan hanya 109 data (11,9%) menjadi yang paling sulit konsisten dengan temuan bahwa kelas minoritas cenderung menghasilkan F1-*Score* lebih rendah pada klasifikasi multi-kelas [20]. Kedua, ambiguitas linguistik kelas netral. Ulasan netral umumnya tidak mengandung kata bermuatan sentimen yang kuat, diperparah oleh penggunaan bahasa informal dan ekspresi lokal wisatawan Indonesia yang belum sepenuhnya terwakili dalam korpus pelatihan model [8].

Perbandingan dengan studi sejenis disajikan pada Tabel 5 untuk memberikan konteks kuantitatif yang lebih terukur.

Tabel 5. Komparasi Model

Studi	Model/Metode	Pendekatan	Akurasi	F1
Yuyun et al. [21]	Multinomial Naïve Bayes	Konvensional (3 kelas)	74%	74%
Asniati et al. [22]	Multinomial Naïve Bayes	Konvensional (3 kelas)	61%	61%
Tesfagergish et al. [26]	*Zero-shot* (dua tahap)	*Zero-shot* (3 kelas)	63%	—
Penelitian ini	IndoBERT pra-latih	*Zero-shot inference* (3 kelas)	58%	63%
Akhdan et al. [14]	IndoBERT + *Confident Learning*	*Fine-tuning*	86,03%	—
Ahlul et al. [15]	IndoBERT	*Fine-tuning*	97,71%	—

Tabel 5 menunjukkan bahwa hasil penelitian ini berada pada rentang yang wajar untuk setting *zero-shot* tiga kelas. Pendekatan konvensional berbasis Naïve Bayes pada domain bahasa Indonesia melaporkan akurasi 74% [21] dan 61% [22] menunjukkan bahwa metode konvensional pun tidak selalu unggul secara konsisten, terutama pada klasifikasi tiga kelas dengan data tidak seimbang. Studi *zero-shot* tiga kelas oleh Tesfagergish et al. [26] melaporkan akurasi 63%, yang berada pada rentang serupa dengan penelitian ini (58%), mengindikasikan bahwa keterbatasan pada kelas netral dan negatif merupakan tantangan umum pendekatan *zero-shot* tanpa adaptasi domain. Sebaliknya, studi dengan *fine-tuning* pada domain spesifik melaporkan akurasi jauh lebih tinggi, seperti IndoBERT dengan *confident learning* yang mencapai 86,03% [14] dan 97,71% pada sentimen publik [15]. Kesenjangan ini menegaskan bahwa *zero-shot inference* memang memiliki batas performa inheren tanpa adaptasi domain, dan menjadi justifikasi kuat untuk pengembangan *dataset* berlabel spesifik Kota Baubau sebagai langkah lanjutan penelitian.

Analisis *word cloud* mengungkap pola tematik yang memiliki implikasi praktis berbeda untuk masing-masing destinasi. Pada kelas positif, dominasi kata "indah" dan "bagus" mencerminkan apresiasi wisatawan terhadap nilai estetika dan daya tarik visual kedua destinasi, sementara kemunculan kata "bersih" mengindikasikan bahwa kebersihan lingkungan menjadi faktor kepuasan utama yang perlu dipertahankan secara konsisten oleh pengelola. Pada kelas negatif, pola kata dominan menunjukkan permasalahan yang berbeda antara kedua destinasi. Kemunculan kata "sampah" secara dominan lebih relevan bagi Pantai Nirwana sebagai destinasi bahari terbuka yang rentan terhadap akumulasi sampah, baik dari pengunjung maupun kiriman laut, sehingga pengelola perlu memprioritaskan sistem pengelolaan sampah berbasis jadwal rutin dan keterlibatan komunitas lokal. Sementara itu, kata "bayar" yang muncul pada kelas negatif lebih relevan bagi Benteng Keraton Buton sebagai situs bersejarah dengan sistem tiket masuk, mengindikasikan ketidakpuasan wisatawan terhadap struktur biaya yang dianggap tidak sepadan dengan fasilitas atau pengalaman yang diterima, sinyal bagi pengelola untuk mengevaluasi kebijakan penetapan harga atau meningkatkan nilai tambah layanan seperti pemandu wisata dan informasi sejarah interaktif. Secara tematik, penelitian ini menunjukkan bahwa citra digital (e-WOM) Kota Baubau sangat bergantung pada manajemen operasional harian. Kegagalan menangani masalah sampah dan transparansi tarif dapat mendegradasi nilai kompetitif destinasi ini di platform digital. Secara keseluruhan, temuan ini memberikan panduan berbasis data bagi Dinas Pariwisata Kota Baubau untuk memprioritaskan dua intervensi utama: (1) penguatan sistem kebersihan dan pengelolaan lingkungan di Pantai Nirwana, dan (2) evaluasi kebijakan tarif dan peningkatan kualitas layanan interpretatif di Benteng Keraton Buton.

Adapun keterbatasan penelitian ini mencakup tiga aspek. Pertama, ketiadaan *dataset* berlabel domain spesifik menjadikan *fine-tuning* sebagai langkah lanjutan yang tidak dapat dihindari untuk meningkatkan performa, khususnya pada kelas netral. Kedua, pemotongan teks

pada 128 token berpotensi menghilangkan informasi sentimen pada ulasan panjang. Ketiga, model *indobertweet-base* yang dioptimalkan untuk teks Twitter mungkin kurang optimal untuk ulasan Google Maps yang memiliki karakteristik tekstual berbeda. Keterbatasan ini sekaligus membuka peluang penelitian lanjutan berupa pengembangan *dataset* anotasi ulasan wisata lokal Kota Baubau dan eksplorasi model berbasis *fine-tuning* domain spesifik.

5. KESIMPULAN

Penelitian ini mengevaluasi kemampuan IndoBERT pra-latih dengan pendekatan *zero-shot inference* untuk mengklasifikasikan sentimen 918 ulasan wisatawan dari Google Maps pada destinasi Benteng Keraton Buton dan Pantai Nirwana di Kota Baubau. Model memperoleh akurasi 0,58, *weighted precision* 0,79, *weighted recall* 0,58, dan *weighted* F1-*Score* 0,63, mengindikasikan kemampuan klasifikasi pada tingkatan cukup (*fair*) tanpa *fine-tuning* tambahan. Rendahnya performa kelas netral (F1: 0.27) mencerminkan tantangan model terhadap ekspresi sentimen ambigu dan *domain gap* antara korpus pelatihan berbasis Twitter dengan karakteristik bahasa ulasan wisata lokal yang merupakan salah satu *research gap* yang diidentifikasi di awal penelitian ini.

IndoBERT pra-latih dapat dipertimbangkan sebagai titik awal (*baseline*) analisis sentimen wisata lokal melalui *zero-shot inference* dalam kondisi ketiadaan *dataset* berlabel domain spesifik, dengan catatan bahwa akurasi 0,58 masih tergolong terbatas dan terdapat kesenjangan substansial dibandingkan pendekatan *fine-tuning* yang menjadi keterbatasan utama penelitian ini. Secara praktis, hasil ini dapat menjadi dasar rekomendasi kebijakan berbasis data bagi Dinas Pariwisata Kota Baubau dalam menyusun strategi peningkatan layanan yang lebih responsif, khususnya pada aspek yang teridentifikasi melalui sentimen negatif pengunjung.

Penelitian selanjutnya disarankan untuk: (1) melakukan *fine-tuning* IndoBERT dengan *dataset* ulasan wisata lokal berlabel untuk meningkatkan performa khususnya pada kelas netral dan negatif; (2) memperluas cakupan dataset ke destinasi wisata lain di Sulawesi Tenggara; serta (3) membandingkan kinerja IndoBERT dengan model multilingual seperti XLM-RoBERTa atau mBERT untuk menilai efektivitas lintas model pada konteks bahasa Indonesia lokal.

KONFLIK KEPENTINGAN

Para penulis menyatakan bahwa tidak terdapat konflik kepentingan antara para penulis maupun dengan objek penelitian dalam makalah ini.

DAFTAR PUSTAKA

[1] A. Mukti, "Dinamika Pengembangan Desa Wisata di Indonesia: Analisis Sistematik tentang Pendorong, Tantangan, dan Dampak," *Jurnal Pembangunan Nagari*, vol. 10, no. 1, pp. 20–37, Jun. 2025, doi: 10.30559/jpn.v10i1.529.

[2] Y. Okdamaiyanti, I. Muda, and N. Angelia, "Implementasi Kebijakan Pengembangan Daerah Wisata oleh Pemerintah Kabupaten Karo (Studi Deskriptif Terhadap Wisata Danau Lau Kawar)," *Jurnal Ilmu Pemerintahan, Administrasi Publik, dan Ilmu Komunikasi (JIPIKOM)*, vol. 7, no. 2, pp. 179–185, May 2025, doi: 10.31289/jipikom.v7i2.6070.

[3] P. N. Yasintha, "Collaborative Governance Dalam Kebijakan Pembangunan Pariwisata di Kabupaten Gianyar," *Jurnal Ilmiah Dinamika Sosial*, vol. 4, no. 1, pp. 1–23, Jan. 2020, doi: https://doi.org/10.38043/jids.v4i1.2219.

[4] T. Sutadi and E. P. Marsongko, "Studi Kebijakan Pengembangan Kawasan Pangandaran Sebagai Kawasan Strategis Pariwisata Nasional," *Jurnal Kepariwisataan: Destinasi, Hospitalitas dan Perjalanan*, vol. 1, no. 1, pp. 1–9, Jun. 2017, doi: 10.34013/jk.v1i1.1.

[5] L. F. Nago, S. N. Hamzah, and C. Panigoro, "Persepsi Wisatawan terhadap Destinasi Wisata Pantai Tilalohe, Kabupaten Gorontalo," *Buletin Ilmiah Marina Sosial Ekonomi Kelautan dan Perikanan*, vol. 10, no. 1, pp. 49–58, Feb. 2024, doi: http://dx.doi.org/10.15578/marina.v10i1.13130.

[6] A. I. P. Nugraheni, L. Prihanti Putri, and N. Pancawati, "Penggunaan Electronic Word of Mouth (eWOM) untuk Berbagi Pengalaman Kuliner oleh Wisatawan," *Tourism Scientific Journal*, vol. 7, no. 1, pp. 15–30, Mar. 2022, doi: 10.32659/tsj.v7i1.144.

[7] I. P. G. A. Sudiatmika, P. S. Saputra, R. L. Rahardian, and K. H. S. Dewi, "Sentiment Analysis of Tourist Reviews on Google Maps For Pura Besakih Using Machine Learning Algorithms," *Jurnal Mandiri IT*, vol. 14, no. 1, pp. 149–158, Jul. 2025, doi: 10.35335/mandiri.v14i1.449.

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

[9] Ardiansyah, A. S. Widagdo, K. N. Qodri, F. E. N. Saputro, and N. A. Rizky P, "Analisis Sentimen Terhadap Pelayanan Kesehatan Berdasarkan Ulasan Google Maps Menggunakan BERT," *Jurnal Fasilkom*, vol. 13, no. 2, pp. 326–333, Aug. 2023, doi: 10.37859/jf.v13i02.5170.

[10] F. W. Atmojo, V. Atina, and H. Permatasari, "Analisis Sentimen Pelanggan Pada Ulasan Google Maps Restoran Al-Ghiff Steak Menggunakan Model IndoBERT," *Jurnal Sistem Informasi dan Teknik Komputer*, vol. 10, no. 2, pp. 336–343, Oct. 2025, doi: 10.51876/simtek.v10i2.1602.

[11] N. Istiqomah and F. Novika, "Perbandingan Kinerja Model NER IndoBERT dan IndoLEM dalam Ekstraksi Informasi Kesehatan Pascabencana dari Berita Daring di Indonesia," *Journal of Computer Science and Informatics Engineering*, vol. 04, no. 3, pp. 158–174, Jul. 2025, doi: 10.55537/cosie.v4i3.1173.

[12] F. Koto, J. H. Lau, and T. Baldwin, "INDOBERTWEET: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Melbourne: Association for Computational Linguistics, Nov. 2021, pp. 10660–10668. doi: 10.18653/v1/2021.emnlp-main.833.

[13] L. R. Andhika, "Public Service Management: An Emerging Research Trend," *Jurnal Borneo Administrator*, vol. 21, no. 1, pp. 61–74, Apr. 2025, doi: 10.24258/jba.v21i1.1581.

[14] D. Al Akhdaan, T. E. Sutanto, and M. Liebenlito, "Confident Learning pada IndoBERT: Peningkatan Kinerja Klasifikasi Sentimen," *The Indonesian Journal of Computer Science*, vol. 13, no. 5, Oct. 2024, doi: 10.33022/ijcs.v13i5.4401.

[15] A. Yoga Pratama, G. Ananda Sanjaya, N. Khairunisa Lubis, and M. Rangga Aditya, "Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada Platform Media Sosial," vol. 9, p. 2025, doi: 10.47002/metik.v9i1.1055.

[16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of North American Chapter of the Association for Computational Linguistics*, Minneapolis, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[17] N. M. Gardazi, A. Daud, M. K. Malik, A. Bukhari, T. Alsahfi, and B. Alshemaimri, "BERT applications in natural language processing: a review," *Artif. Intell. Rev.*, vol. 58, no. 6, Jun. 2025, doi: 10.1007/s10462-025-11162-5.

[18] A. Rogers, O. Kovaleva, and A. Rumshisky, "A Primer in BERTology: What We Know About How BERT Works," *Trans. Assoc. Comput. Linguist.*, vol. 8, pp. 842–866, Dec. 2020, doi: 10.1162/tacl_a_00349.

[19] R. Gupta, "Bidirectional encoders to state-of-the-art: a review of BERT and its transformative impact on natural language processing," *Информатика. Экономика.*

Управление - Informatics. Economics. Management, vol. 3, no. 1, pp. 0311–0320, Mar. 2024, doi: 10.47813/2782-5280-2024-3-1-0311-0320.

[20] J. Ipmawati, S. Saifulloh, and K. Kusnawi, "Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 247–256, Jan. 2024, doi: 10.57152/malcom.v4i1.1066.

[21] Yuyun, N. Hidayah, and S. Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820–826, Aug. 2021, doi: 10.29207/resti.v5i4.3146.

[22] A. Asniati, S. Hady, and I. N. Tolanto, "Analisis Sentimen Netizen Twitter Terhadap Program Makan Siang Gratis Menggunakan Algoritma Naïve Bayes," *JURNAL INFORMATIKA*, vol. 14, no. 2, pp. 19–28, Dec. 2025, doi: 10.55340/jiu.v14i2.2582.

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

[24] D. Guidotti, L. Pandolfo, and L. Pulina, "Discovering Sentiment Insights: Streamlining Tourism Review Analysis with Large Language Models," *Information Technology and Tourism*, vol. 27, no. 1, pp. 227–261, Mar. 2025, doi: 10.1007/s40558-024-00309-9.

[25] I. Nawawi, K. F. Ilmawan, M. R. Maarif, and M. Syafrudin, "Exploring Tourist Experience through Online Reviews Using Aspect-Based Sentiment Analysis with Zero-Shot Learning for Hospitality Service Enhancement," *Information (Switzerland)*, vol. 15, no. 8, p. 1, Aug. 2024, doi: 10.3390/info15080499.

[26] F. Koto, T. Beck, Z. Talat, I. Gurevych, and T. Baldwin, "Zero-shot Sentiment Analysis in Low-Resource Languages Using a Multilingual Sentiment Lexicon," in *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, St. Julian's, Malta: Association for Computational Linguistics, Mar. 2024, pp. 298–320. doi: 10.18653/v1/2024.eacl-long.18.

[27] M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi," in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Jul. 2020. doi: 10.1088/1757-899X/874/1/012017.

[28] J. Asher and E. P. Rachmawati, "Analisis Sentimen Ulasan Bintang Lima Aplikasi Instagram di Google Play Store menggunakan IndoBERT," *Dinamik*, vol. 30, no. 2, pp. 318–330, Jul. 2025, doi: 10.35315/dinamik.v30i2.10192.