

Evaluasi IndoBERTweet untuk Klasifikasi Sentimen Ulasan Wisata melalui Inferensi Langsung Lintas Domain: Studi Kasus Kota Baubau

Evaluation of IndoBERTweet for Tourism Review Sentiment Classification via Cross-Domain Direct Inference: A Case Study in Baubau City

Nurul Hidayah*¹, Dodiman², Asniati³, La Ode Dwiyan Pramono Darmin⁴

^{1,2,3}Program Studi Teknik Informatika, ⁴Program Studi Administrasi Negara,

Universitas Dayanu Ikhsanuddin, Baubau Indonesia

Email: ¹nurul.hyh@gmail.com

Article	Received	Revised	Accepted	Published:
Info:	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. Penelitian ini menganalisis kemampuan IndoBERTweet dalam mengklasifikasikan sentimen ulasan wisatawan melalui pendekatan cross-domain direct inference tanpa fine-tuning tambahan, dengan studi kasus pada Benteng Keraton Buton dan Pantai Nirwana di Kota Baubau. Sebanyak 896 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. Evaluasi menggunakan confusion matrix menghasilkan akurasi 0,67 dan weighted F1-Score 0,71, dengan kinerja terbaik pada kelas positif (F1: 0,80) dan kinerja lebih rendah pada kelas negatif (F1: 0,41) dan netral (F1: 0,44). Temuan ini mengindikasikan bahwa IndoBERTweet berpotensi digunakan sebagai instrumen awal analisis sentimen wisata lokal pada kondisi ketiadaan dataset berlabel domain spesifik, dan hasilnya dapat menjadi pertimbangan awal bagi prioritas kebijakan pengelolaan pariwisata Kota Baubau, namun peningkatan performa lebih lanjut memerlukan fine-tuning dengan data berlabel dari domain wisata yang relevan.

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

Abstract

Tourist reviews on digital platforms represent a valuable source of public perception data to support evidence-based tourism management. This study analyzes the capability of IndoBERTweet in classifying tourist review sentiment through a cross-domain direct inference approach without additional fine-tuning, with a case study on Benteng Keraton Buton and Pantai Nirwana in Baubau City. A total of 896 reviews were collected from Google Maps via web scraping and processed through case folding, cleaning, text normalization, and negation word concatenation to classify sentiment into positive, negative, and neutral classes. Evaluation using a confusion matrix yielded an accuracy of 0.67 and a weighted F1-Score of 0.71, with the best performance on the positive class (F1: 0.80) and lower performance on the negative (F1: 0.41) and neutral (F1: 0.44) classes. These findings indicate that IndoBERTweet has potential as an initial instrument for local tourism sentiment analysis in the absence of domain-specific labeled datasets, and the results may serve as a preliminary consideration for tourism management policy priorities in Baubau City; however, further performance improvement requires fine-tuning with labeled data from a relevant tourism domain.

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

This is an open access article under the CC BY-SA license.



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 konteks 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

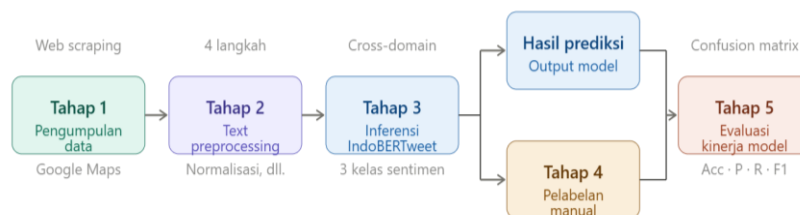
analisis sentimen terhadap data Twitter [21], [22], dengan fokus yang lebih banyak pada performa klasifikasi dibanding pemanfaatan hasilnya untuk pengelolaan pariwisata daerah. Penggunaan ulasan Google Maps dari destinasi lokal seperti Kota Baubau juga belum banyak dieksplorasi.

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]. Studi pada destinasi wisata Indonesia menunjukkan bahwa IndoBERT memerlukan *fine-tuning* domain spesifik untuk mencapai performa optimal pada ulasan Google Maps [23], [24], sehingga penerapan *cross-domain direct inference* menggunakan model IndoBERTweet yang telah di-*fine-tune* untuk sentimen bahasa Indonesia dan diterapkan langsung pada domain ulasan wisata tanpa *fine-tuning* tambahan menjadi alternatif yang praktis sebelum investasi anotasi data untuk *fine-tuning* dilakukan [23], [25]. Dengan demikian, penelitian ini sekaligus mengisi kekosongan studi serupa pada destinasi wisata lokal di luar Jawa yang belum banyak dieksplorasi.

Oleh karena itu, penelitian ini bertujuan untuk (1) mengevaluasi kemampuan IndoBERTweet untuk klasifikasi sentimen ulasan wisatawan di Kota Baubau melalui *cross-domain direct inference* tanpa *fine-tuning* tambahan; (2) menganalisis distribusi sentimen positif, negatif, dan netral; serta (3) mengukur performa klasifikasi melalui akurasi, *precision*, *recall*, dan *F1-Score*. Penelitian ini mengkaji sejauh mana IndoBERTweet yang diterapkan secara lintas domain mampu mengklasifikasikan sentimen ulasan wisata lokal serta implikasinya bagi analisis sentimen pariwisata daerah berbasis data.

2. METODE

2.1. Tahapan Penelitian

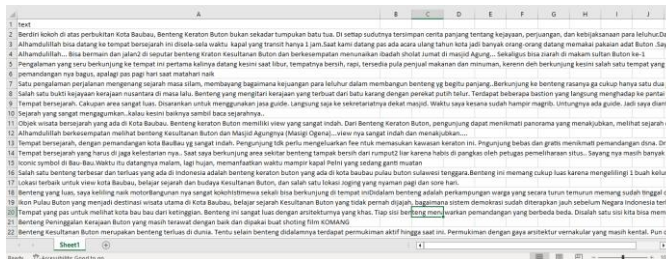


Gambar 1. Alur Penelitian

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 IndoBERTweet melalui pendekatan *cross-domain direct inference* untuk mengklasifikasikan sentimen ke dalam tiga kelas yaitu 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 896 ulasan digunakan sebagai input inferensi IndoBERTtweet tanpa *fine-tuning* tambahan pada domain wisata, mengingat model telah di-*fine-tune* untuk klasifikasi sentimen bahasa Indonesia pada domain Twitter sehingga dapat diterapkan langsung secara lintas domain (*cross-domain direct inference*) [23]. Distribusi data ditampilkan pada Tabel 1.

Tabel 1. Distribusi Data Berdasarkan Label Kelas Sentimen

Label Kelas	Jumlah Data	Proporsi
Positif	663	74.0%
Netral	164	18.3%
Negatif	69	7.7%
Total	896	100%

Data menunjukkan ketidakseimbangan kelas (*class imbalance*) yang bersifat alami, dengan kelas positif mendominasi sebesar 663 ulasan (74,0%), diikuti netral 164 ulasan (18,3%), dan negatif 69 ulasan (7,7%). Kondisi ini mencerminkan kecenderungan alami wisatawan yang merasa puas untuk lebih aktif menuliskan pengalamannya pada platform digital [20].

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 *cross-domain direct inference* di mana model IndoBERTtweet yang telah di-*fine-tune* pada domain Twitter diterapkan langsung pada domain ulasan wisata tanpa memerlukan data berlabel dari domain target sebagai input pelatihan tambahan [23]. 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 [26]. 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 normalisasi teks informal terbatas (*custom normalization lexicon*) yang dikurasi secara manual sebanyak 20 entri kata tidak baku yang dominan muncul pada data [27]; dan (e) penggabungan kata negasi seperti “tidak”, “ga”, dan “nggak” dengan kata berikutnya menggunakan *underscore* (contoh: “tidak_bagus”) agar konteks negatif dapat ditangkap model secara akurat [28], [29].

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

Tabel 3. Contoh Hasil *Text Preprocessing*

Tahap	Teks
Sebelum <i>preprocessing</i>	“Tempatnya BAGUS banget!! Recommended bgt, gak nyesel kesini 😊 https://t.co/xxx ”
Setelah <i>case folding</i>	“tempatny bagus banget!! Recommended bgt, gak nyesel kesini 😊”
Setelah <i>cleaning</i>	“tempatny bagus banget recommended bgt gak nyesel kesini”
Setelah normalisasi	“tempatny bagus banget recommended banget tidak_nyesel kesini”

2.5. Model IndoBERT untuk Klasifikasi Sentimen

Penelitian ini menggunakan model IndoBERTweet (Aardiiiy/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 [11]. Pendekatan *cross-domain direct inference* diterapkan dengan menggunakan model tersebut secara langsung pada data ulasan wisata Kota Baubau tanpa *fine-tuning* tambahan, dengan tujuan mengevaluasi kemampuan generalisasi model dalam mentransfer pengetahuan linguistik dan sentimen dari domain Twitter ke domain ulasan wisata lokal yang memiliki karakteristik tekstual berbeda.

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
<i>Framework</i>	PyTorch + HuggingFace Transformers
Jumlah Kelas	3 (Positif, Negatif, Netral)
Panjang Maksimum Token	128 token
Mode Inferensi	<i>Cross-domain direct inference</i>

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]:

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

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

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

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

yang merupakan rata-rata harmonik *precision* dan *recall* sehingga lebih representatif pada data tidak seimbang.

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

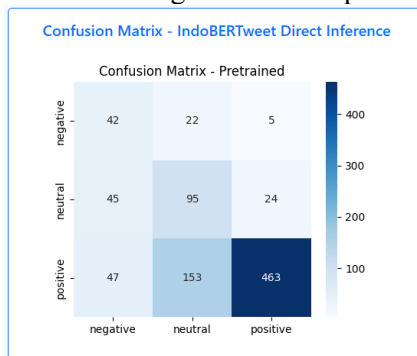
Melalui tahapan *preprocessing*, data mentah berhasil ditransformasi menjadi teks yang bersih dan terstandarisasi. *Case folding* menyeragamkan kapitalisasi, *cleaning* menghilangkan *noise*, normalisasi menggunakan *custom normalization lexicon* menyeragamkan penulisan informal (contoh: "gak" menjadi "tidak"), dan penggabungan kata negasi (contoh: "tidak_bagus") diterapkan sebagai langkah eksploratif untuk mempertahankan konteks semantik negatif dalam teks. 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

Inferensi dilakukan terhadap seluruh 896 ulasan menggunakan *pipeline IndoBERTtweet* (Aardiiiy/indoberttweet-base-Indonesian-sentiment-analysis). Model mengklasifikasikan setiap ulasan ke dalam tiga kelas sentimen: positif, negatif, dan netral. Hasil prediksi kemudian dibandingkan dengan label manual (*ground truth*) menggunakan *confusion matrix 3x3* 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 42 dari 69 ulasan negatif, 95 dari 164 ulasan netral, dan 463 dari 663 ulasan positif. Kelas positif memperoleh prediksi terbanyak namun juga menghasilkan kesalahan terbesar, dengan 153 ulasan positif salah diprediksi sebagai netral dan

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. [31]	<i>Zero-shot</i> (dua tahap)	<i>Zero-shot</i> (3 kelas)	63%	—
Penelitian ini	IndoBERTweet	<i>Cross-domain direct inference</i> (3 kelas)	67%	71%
Akhdan et al. [14]	IndoBERT + <i>Confident Learning</i>	<i>Fine-tuning</i>	86,03%	—
Ahlul et al. [15]	IndoBERT	<i>Fine-tuning</i>	97,71%	—

Tabel 5 menunjukkan bahwa hasil penelitian ini berada pada rentang yang wajar untuk setting *cross-domain direct inference* 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. [31] melaporkan akurasi 63%, yang berada pada rentang serupa dengan penelitian ini (67%), meskipun keduanya menggunakan pendekatan yang berbeda secara konseptual yaitu *Zero-shot* pada model generik versus *cross-domain direct inference* pada model yang telah di-*fine-tune* untuk sentimen. Nilai akurasi ini mengindikasikan bahwa model memiliki kemampuan klasifikasi yang kompetitif dan berada di atas performa model berbasis *machine learning* konvensional maupun pendekatan *zero-shot*, meskipun diaplikasikan langsung tanpa adanya proses *fine-tuning* tambahan. Hasil tersebut dinilai wajar dan memuaskan mengingat IndoBERTweet digunakan secara lintas domain (*cross-domain*) tanpa adaptasi khusus pada karakteristik teks ulasan wisata.

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 *cross-domain direct inference* memiliki batas performa inheren akibat *domain gap* antara korpus pelatihan dan domain target, dan menjadi justifikasi kuat untuk pengembangan *dataset* berlabel spesifik Kota Baubau sebagai langkah lanjutan penelitian.

4.2. Pembahasan Word Cloud

Tabel 5. Top-10 Kata Dominan per Kelas Sentimen (setelah *preprocessing*)

No	Positif	Frek	%	Netral	Frek	%	Negatif	Frek	%
1	pantai	214	32,3%	benteng	71	43,3%	sampah	26	37,7%
2	indah	156	23,5%	Buton	53	32,3%	pantai	22	31,9%
3	benteng	138	20,8%	Pantai	34	20,7%	bagus	16	23,2%
4	tempat	132	19,9%	Tempat	27	16,5%	sayang	13	18,8%
5	buton	99	14,9%	Keraton	26	15,9%	kurang	12	17,4%
6	putih	95	14,3%	Dunia	25	15,2%	kotor	8	11,6%
7	kota	93	14,0%	Terluas	23	14,0%	indah	10	14,5%
8	bersih	86	13,0%	Baubau	23	14,0%	agak	9	13,0%
9	pasir	84	12,7%	Kesultanan	20	12,2%	perlu	7	10,1%
10	pemandangan	65	9,8%	Sejarah	18	11,0%	beach	10	14,5%

Analisis *word cloud* mengungkap pola tematik yang memiliki implikasi praktis bagi masing-masing destinasi. Untuk memperkuat temuan *word cloud* secara kualitatif, Tabel 6

menyajikan contoh ulasan representatif dari masing-masing kelas sentimen yang memuat kata kunci dominan sebagaimana teridentifikasi pada Tabel 5.

Tabel 6. Contoh Ulasan Representatif per Kelas Sentimen

No	Kelas	Contoh Ulasan	Kata Kunci
1	Positif	Pantainya bersih, terawat, banyak spot foto, fasilitas terawat	Bersih, terawat
2	Positif	Tempatnya sangat bersih, asrih, hijau, indah & nyaman	Bersih, indah, nyaman
3	Netral	Benteng terluas di dunia, banyak unsur-unsur sejarah yang masih kental akan kesultanan Buton	Benteng, sejarah, kesultanan
4	Netral	Bagus pemandangan... hanya kuliner yang kurang	Bagus, pemandangan, kurang
5	Negatif	Pantainya bagus sayang banyak sampah	Sampah, sayang
6	Negatif	Pantainya jernih...namun masih banyak sampah plastik	Jernih, sampah

Pada kelas positif, dominasi kata “indah”, “bersih”, dan “pemandangan” 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, kemunculan kata “sampah” dan “kotor” yang mendominasi lebih relevan bagi Pantai Nirwana sebagai destinasi bahari terbuka yang rentan terhadap akumulasi sampah, sehingga pengelola perlu memprioritaskan sistem pengelolaan sampah berbasis jadwal rutin dan keterlibatan komunitas lokal.

Menariknya, kemunculan kata “bagus” pada kelas netral dan kata “indah” pada kelas negatif mengindikasikan bahwa banyak ulasan menggunakan kata positif yang diikuti penyangkalan, seperti “bagus tapi kurang terawat” atau “indah, sayang banyak sampah”, sehingga sentimen keseluruhan tidak cukup kuat untuk diklasifikasikan sebagai positif. Hal ini sekaligus mencerminkan keterbatasan model dalam menangkap nuansa kontekstual antarklausa sebagai konsekuensi dari penggunaan *direct inference* tanpa *fine-tuning*.

Secara kuantitatif, kata “sampah” dan “kotor” muncul pada 28 dari 69 ulasan negatif (40,6%), menjadikannya tema keluhan paling dominan. Sementara itu, kata “kurang”, “perlu”, dan “fasilitas” muncul pada 21 ulasan negatif (30,4%), mengindikasikan ketidakpuasan terhadap kelengkapan sarana pendukung. Temuan ini memberikan dasar empiris bagi pengelola untuk memprioritaskan dua intervensi, yaitu (1) pengelolaan kebersihan di Pantai Nirwana, dan (2) peningkatan fasilitas pendukung di kedua destinasi.

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 dan negatif. Kedua, pemotongan teks pada 128 token berpotensi menghilangkan informasi sentimen pada ulasan panjang. Ketiga, model IndoBERTweet yang dilatih pada korpus Twitter memiliki *domain gap* terhadap 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 IndoBERTweet melalui pendekatan *cross-domain direct inference* untuk mengklasifikasikan sentimen 896 ulasan wisatawan Google Maps pada destinasi Benteng Keraton Buton dan Pantai Nirwana di Kota Baubau. Model memperoleh akurasi 0,67, weighted F1-score 0,71, dengan kinerja terbaik pada kelas positif (F1: 0,80) dan terendah pada kelas netral (F1: 0,44), mengindikasikan kemampuan klasifikasi yang kompetitif tanpa *fine-tuning* tambahan pada domain wisata. Rendahnya performa kelas netral dan negatif mencerminkan adanya *domain gap* antara korpus pelatihan berbasis Twitter dan karakteristik bahasa ulasan wisata lokal, kemudian *class imbalance* yang menjadikan kelas positif jauh lebih dominan (663 ulasan, 74,0%) dibanding kelas netral (164 ulasan, 18,3%) dan negatif (69 ulasan, 7,7%). Keterbatasan lain mencakup pemotongan teks pada 128 token yang berpotensi menghilangkan informasi sentimen pada ulasan panjang, serta perbedaan karakteristik tekstual antara teks Twitter dan ulasan Google Maps. Analisis *word cloud* mengungkap bahwa keluhan kebersihan (“sampah”, “kotor”) mendominasi ulasan negatif Pantai Nirwana, sementara kata “kurang” dan “fasilitas” mengindikasikan ketidakpuasan terhadap sarana pendukung di kedua destinasi. Temuan ini merekomendasikan dua prioritas intervensi bagi pengelola, yaitu penguatan sistem kebersihan di Pantai Nirwana dan peningkatan fasilitas pendukung di kedua destinasi.

Penelitian selanjutnya disarankan untuk (1) melakukan *fine-tuning* IndoBERTweet 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 IndoBERTweet 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 Sistematis 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: 10.38043/jids.v4i1.2219.
- [4] T. Sutadi and E. P. Marsongko, “Studi Kebijakan Pengembangan Kawasan Pangandaran Sebagai Kawasan Strategis Pariwisata Nasional,” *Jurnal Kepariwisata: 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: 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. Sudiarmika, 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] 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.
- [12] M. A. N. Hibrizi and D. A. Fatah, "Dampak Pra-pemrosesan Teks pada Akurasi Analisis Sentimen Multi-Aspek Menggunakan IndoBERT," *Sains Data Jurnal Studi Matematika dan Teknologi*, vol. 4, no. 1, pp. 42–48, Mar. 2026, doi: 10.52620/sainsdata.v4i1.301.
- [13] S. Nur, S. Sahibu, and M. Razak, "Aspect-Based Sentiment Analysis of Tourist Attractions in Labuanbajo Using The Transformer Model As A Recommendation for Improving Service Quality," *Journal of Applied Informatics and Computing (JAIC)*, vol. 10, no. 1, p. 496, Feb. 2026, doi: 10.30871/jaic.v10i1.11565.
- [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. Y. Pratama, G. A. Sanjaya, N. K. Lubis, M. R. Aditya, and Y. Yennimar, "Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada Platform Media Sosial," *Metik Jurnal*, vol. 9, no. 1, p. 2025, Jun. 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," *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] C. A. Bahri and L. H. Suadaa, “Aspect-Based Sentiment Analysis in Bromo Tengger Semeru National Park Indonesia Based on Google Maps User Reviews,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 17, no. 1, p. 79, Feb. 2023, doi: 10.22146/ijccs.77354.
- [25] D. D. Purwanto, “Empirical Evaluation of IndoBERT and LSTM for Sentiment Analysis of Tourism Reviews: A Data-Driven Study on Kenjeran Park,” *Jurnal Teknik Informatika (Jutif)*, vol. 7, no. 1, pp. 463–474, Feb. 2026, doi: 10.52436/1.jutif.2026.7.1.4901.
- [26] M. A. Rosid, A. S. Fitriani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, “Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 874, no. 1, Art. no. 012017, pp. 1–6, Jul. 2020, doi: 10.1088/1757-899X/874/1/012017.
- [27] 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.
- [28] R. Amalia, M. A. Bijaksana, and D. Darmantoro, “Negation Handling in Sentiment Classification Using Rule-Based Adapted from Indonesian Language Syntactic for Indonesian Text in Twitter,” in *Journal of Physics: Conference Series*, Bandung: Institute of Physics Publishing, Apr. 2018. doi: 10.1088/1742-6596/971/1/012039.
- [29] T. G. Prahasiwi and R. Kusumaningrum, “Implementation of Negation Handling Techniques Using Modified Syntactic Rule in Indonesian Sentiment Analysis,” in *Journal of Physics: Conference Series*, Semarang: Institute of Physics Publishing, Jun. 2019. doi: 10.1088/1742-6596/1217/1/012115.
- [30] V. Ganganwar and R. Rajalakshmi, “Employing Synthetic Data for Addressing the Class Imbalance in Aspect-Based Sentiment Classification,” *Journal of Information and Telecommunication*, vol. 8, no. 2, pp. 167–188, Apr. 2024, doi: 10.1080/24751839.2023.2270824.
- [31] 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.