

ANALISIS SENTIMEN WISATA BOROBUDUR DAN PRAMBANAN MENGGUNAKAN MULTINOMIAL NAÏVE BAYES

Diterima Redaksi: 5 November 2025; Revisi Akhir: 9 Januari 2026; Diterbitkan Online: 31 Mei 2026

**Baruna Abirawa¹⁾, Irvan Alfaritz²⁾, Smertniki Javid Ahmedthian³⁾, M. Syamsuddin
Wisnubroto^{4*)} Fajri Farid⁵⁾, Meida Cahyo Untoro⁶⁾**

^{1, 2, 3, 4, 5)} Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera

⁶⁾ Program Studi Teknik Informatika, Fakultas Teknologi Produksi dan Industri, Institut Teknologi Sumatera

^{1, 2, 3, 4, 5, 6)} Jl. Terusan Ryacudu, Way Huwi, Lampung Selatan, Lampung, 35365, Indonesia

e-mail: baruna.122450097@student.itera.ac.id¹⁾, irvan.122450093@student.itera.ac.id²⁾,

smertniki.122450115@student.itera.ac.id³⁾, syamsuddin.wisnubroto@sd.itera.ac.id^{4*)}, fajri.farid@sd.itera.ac.id⁵⁾

cahyo.untoro@if.itera.ac.id⁶⁾

Abstrak: Candi Borobudur dan Candi Prambanan dikenal sebagai dua situs warisan dunia UNESCO, yang memiliki nilai sejarah dan budaya tinggi. Penelitian ini membandingkan sentimen wisatawan untuk memetakan persepsi pengunjung serta mengidentifikasi isu layanan yang perlu diprioritaskan. Data penelitian berupa 10.000 ulasan Google Maps tahun 2023 (5.000 ulasan per destinasi). Hasil menunjukkan sentimen positif lebih dominan pada Candi Prambanan (82,5%) dibandingkan Candi Borobudur (73,7%). Evaluasi model klasifikasi menunjukkan akurasi 81,9% pada dataset Prambanan dan 74,7% pada dataset Borobudur. Visualisasi wordcloud mengindikasikan keluhan negatif yang berulang pada kedua destinasi terutama berkaitan dengan kenyamanan kunjungan (panas), keterjangkauan biaya (harga tiket), dan fasilitas pendukung (parkir). Temuan ini menegaskan perbedaan pengalaman wisatawan pada kedua destinasi dan menyediakan dasar empiris bagi pengelola untuk merumuskan strategi peningkatan layanan dan penataan fasilitas secara lebih terarah berbasis umpan balik pengunjung.

Kata Kunci— Analisis Sentimen, Ulasan Google Maps, Candi Borobudur, Candi Prambanan, Multinomial Naive Bayes

Abstract: Borobudur Temple and Prambanan Temple are known as two UNESCO world heritage sites, possessing high historical and cultural value. This study compares tourist sentiment to map visitor perceptions and identify service issues that should be prioritized. The dataset consists of 10,000 Google Maps reviews from 2023 (5,000 reviews per destination). The results show that positive sentiment is more dominant for Prambanan Temple (82.5%) than for Borobudur Temple (73.7%). Model evaluation indicates classification accuracies of 81.9% for the Prambanan dataset and 74.7% for the Borobudur dataset. Wordcloud visualization highlights recurring negative complaints across both destinations, mainly related to visitor comfort (heat), affordability (ticket prices), and supporting facilities (parking). These findings confirm differences in visitor experiences between the two destinations and provide an empirical basis for managers to formulate more targeted strategies for service improvement and facility planning based on visitor feedback.

Keywords— Sentiment Analysis, Google Maps Reviews, Borobudur Temple, Prambanan Temple, Multinomial Naive Bayes

Penulis Korespondensi:

M. Syamsuddin Wisnubroto

Program Studi Sains Data

Institut Teknologi Sumatera

Email: syamsuddin.wisnubroto@sd.itera.ac.id

Orchid ID: <https://orcid.org/0000-0002-8836-9608>

I. PENDAHULUAN

KEMAJUAN teknologi digital telah mengubah secara fundamental cara masyarakat dalam menyampaikan informasi dan opini. Platform daring seperti media sosial dan situs ulasan kini menjadi sumber data berharga bagi berbagai analisis perilaku masyarakat [1]. Salah satu platform yang paling banyak digunakan adalah *Google Maps*, yang tidak hanya berfungsi sebagai peta digital, tetapi juga sebagai ruang interaktif bagi pengguna untuk memberikan ulasan (*review*) terhadap berbagai lokasi, termasuk destinasi pariwisata. Ulasan-ulasan tersebut memuat opini dan persepsi yang dapat merefleksikan pengalaman wisatawan secara autentik [2].

Candi Borobudur dan Candi Prambanan merupakan dua situs warisan budaya dunia UNESCO yang menjadi ikon utama pariwisata di Indonesia. Reputasi dan citra kedua candi ini sangat dipengaruhi oleh pengalaman yang dibagikan oleh wisatawan melalui ulasan daring. Ulasan pada *Google Maps* kini menjadi referensi penting bagi calon wisatawan lain dalam mengambil keputusan kunjungan, sekaligus menjadi sumber umpan balik langsung bagi pengelola untuk menilai kualitas layanan dan pengalaman pengunjung.

Namun, jumlah ulasan yang mencapai ribuan menjadikannya sulit untuk dianalisis secara manual. Selain itu, data teks bersifat tidak terstruktur sehingga diperlukan pendekatan komputasi yang mampu mengekstrak makna dan pola secara efisien [3]. Dalam konteks ini, *text mining* dan analisis sentimen menjadi metode penting untuk mengungkapkan opini publik. Analisis sentimen merupakan bagian dari *Natural Language Processing* (NLP) yang bertujuan untuk mengklasifikasikan opini dalam teks ke dalam kategori polaritas sentimen seperti positif, negatif, atau netral [4].

Beberapa penelitian sebelumnya telah membuktikan efektivitas *Naive Bayes* dan pendekatan leksikon seperti VADER (*Valence Aware Dictionary and sEntiment Reasoner*) dalam analisis sentimen. Namun, tantangan masih muncul dalam penerapan metode ini pada data berbahasa Indonesia yang memerlukan translasi data ke Bahasa Inggris, yang berisiko pada proses translasi sering kali menghilangkan konteks linguistik dan nuansa local [5]. Selain itu, sebagian besar penelitian terdahulu berfokus pada aplikasi seluler atau media sosial, belum pada domain pariwisata dengan pendekatan komparatif lintas destinasi [6].

Penelitian ini berupaya mengisi kesenjangan tersebut dengan menerapkan pendekatan hibrida yang menggabungkan pelabelan otomatis berbasis leksikon VADER dan model klasifikasi Multinomial Naive Bayes untuk menganalisis serta membandingkan sentimen wisatawan terhadap Candi Borobudur dan Candi Prambanan. Melalui pendekatan ini, penelitian diharapkan dapat memberikan pemahaman yang lebih mendalam mengenai persepsi wisatawan terhadap dua destinasi ikonik Indonesia, sekaligus menawarkan kontribusi praktis bagi pengelola dalam meningkatkan kualitas layanan berbasis analisis sentimen.

II. TINJAUAN PUSTAKA

Analisis sentimen terhadap data ulasan daring adalah area penelitian yang telah banyak dieksplorasi dengan berbagai jenis objek dan metode. Pengkajian terhadap studi-studi sebelumnya penting untuk menempatkan penelitian ini dan menemukan kebaruan yang dihadirkan.

Studi pertama menganalisis sentimen pada ulasan lokasi di *Google Maps* menggunakan algoritma *Naive Bayes*. Metode mereka mencakup langkah penting yaitu menerjemahkan ulasan dari Bahasa Indonesia ke Bahasa Inggris agar pustaka *TextBlob* dapat memprosesnya untuk pelabelan sentimen otomatis. Berbeda dengan itu, penelitian ini mempertahankan bahasa asli dan menggunakan VADER, yang meskipun dirancang untuk Bahasa Inggris, aturannya dapat diadaptasi dan divalidasi untuk konteks informal Bahasa Indonesia [7].

Selanjutnya, studi kedua berhasil menerapkan pendekatan hibrida dengan pelabelan otomatis menggunakan VADER dan klasifikasi *Naive Bayes* pada ulasan aplikasi PLN *Mobile*. Penelitian ini mengadopsi metodologi hibrida yang serupa, namun menerapkannya pada domain yang berbeda (pariwisata) dan dengan tujuan komparatif [8].

Terakhir, studi ketiga mengevaluasi opini masyarakat mengenai produk kendaraan listrik menggunakan VADER. Walaupun mereka tidak menggunakan *Naive Bayes* untuk pengklasifikasian, studi ini penting karena menunjukkan efektivitas pendekatan berbasis leksikon (VADER) dalam menilai

opini publik dari data media sosial X (Twitter). Studi ini juga menyoroti signifikansi visualisasi data, seperti *wordcloud*, dalam menyajikan hasil analisis agar pembaca dapat memahaminya dengan sederhana. Hal ini sejalan dengan studi ini yang juga memanfaatkan *wordcloud* sebagai salah satu alat visualisasi utama untuk mengidentifikasi kata kunci pada sentimen positif dan negatif [9].

III. METODE PENELITIAN

A. Deskripsi Data

Data dalam penelitian ini diperoleh melalui teknik web scraping menggunakan *Apify*, yaitu sebuah platform yang memfasilitasi otomatisasi pengambilan data dari berbagai sumber web [10]. Sumber data utama berasal dari ulasan pengguna pada *Google Maps* untuk dua situs bersejarah penting di Indonesia, yakni Candi Borobudur di Magelang dan Candi Prambanan di Yogyakarta.

Informasi yang diekstrak meliputi judul ulasan, tautan ke lokasi ulasan pada *Google Maps*, rating bintang yang diberikan oleh pengunjung, nama penulis ulasan, tautan langsung menuju ulasan tersebut, serta isi ulasan yang memuat pengalaman dan pendapat pengunjung terkait situs bersejarah tersebut.

B. Pengolahan Data

Pada tahap pengolahan data, ulasan yang dikumpulkan dari Google Maps terlebih dahulu diterjemahkan menggunakan *google translator* dari *deep-translator* agar konsistensi bahasa terjaga. Setelah proses penerjemahan selesai, teks hasil terjemahan dianalisis menggunakan *VADER Lexicon* untuk pelabelan sentimen. Pemilihan *VADER Lexicon* didasarkan pada tingkat akurasi dalam menganalisis sentimen teks berbahasa Inggris, sehingga ulasan dapat dikategorikan ke dalam sentimen positif, negatif, maupun netral [11], [12]. Hal ini dilakukan karena *VADER Lexicon* lebih akurat dalam melakukan analisis sentimen pada teks berbahasa Inggris [13]. Walaupun demikian, ulasan dalam bahasa Indonesia tetap diproses untuk keperluan analisis sentimen. Proses pengolahan data diawali dengan tahapan *preprocessing*, yang meliputi beberapa teknik pengolahan teks. Langkah pertama adalah *case folding*, di mana seluruh teks diubah menjadi huruf kecil untuk memastikan konsistensi. Selanjutnya dilakukan *tokenizing*, yaitu memecah teks menjadi kata-kata atau token. Proses *slang word removal* dilakukan untuk menghapus kata-kata *slang* yang tidak relevan, dilanjutkan dengan *filtering* guna menyaring kata-kata yang tidak memberikan kontribusi signifikan terhadap analisis. Tahapan terakhir dalam *preprocessing* adalah *stemming*, yang mengubah kata-kata ke bentuk dasarnya agar variasi kata dapat diminimalisir.

Setelah tahapan *preprocessing* selesai, dilakukan pemberian bobot pada kata-kata menggunakan metode TF-IDF [14]. Pembobotan ini bertujuan untuk mengidentifikasi kata-kata yang paling penting dalam dokumen ulasan, sehingga dapat digunakan dalam model klasifikasi *Naive Bayes*. Model ini kemudian digunakan untuk mengklasifikasikan sentimen ulasan menjadi kategori positif, negatif, atau netral. Akhirnya, hasil klasifikasi dievaluasi menggunakan *Confusion Matrix* untuk menilai kinerja model dalam memprediksi sentimen dengan tingkat akurasi yang optimal [15].

IV. HASIL DAN PEMBAHASAN

A. Scraping Data

Pengambilan data (*scraping data*) ulasan dilakukan melalui teknik web scraping yang mengekstraksi komentar pengguna dari *Google Maps*. Dataset terdiri atas 10.000 ulasan, masing-masing 5.000 ulasan untuk Candi Borobudur dan 5.000 ulasan untuk Candi Prambanan. Setiap entri data memiliki atribut berupa *title*, *url*, *stars*, *name*, *reviewUrl*, dan *text*. Atribut nama pengulas (*name*) telah dianonimkan sehingga tidak menampilkan identitas asli pribadi.

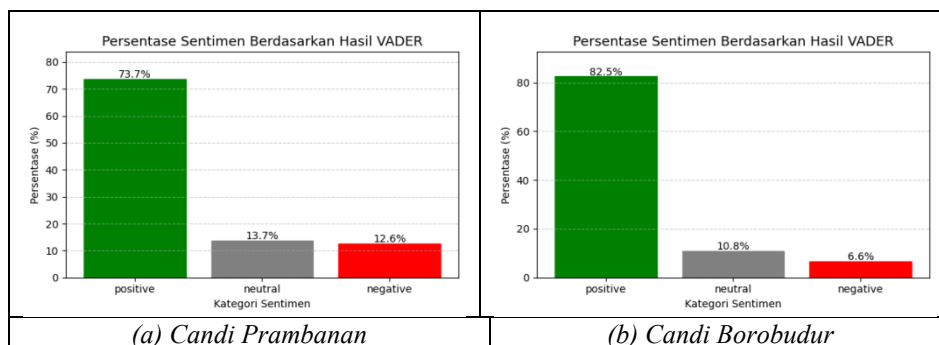
	title	url	stars	name	reviewUrl	text
0	Candi Prambanan	https://www.google.com/maps/search/?api=1&quer...	4	User_1	https://www.google.com/maps/reviews/data=14m8l...	Salah satu candi termegah di Indonesia yang wa...
1	Candi Prambanan	https://www.google.com/maps/search/?api=1&quer...	5	User_2	https://www.google.com/maps/reviews/data=14m8l...	Objek wisata super nya jogjakarta \nTerletak s...

2	Candi Borobudur	https://www.google.com/maps/search/?api=1&quer...	5	User_3	https://www.google.com/maps/reviews/data=14m8!...	Bener bener tempat ini makin indah dan merrak ...
3	Candi Borobudur	https://www.google.com/maps/search/?api=1&quer...	5	User_4	https://www.google.com/maps/reviews/data=14m8!...	Back to here after 26 yrs ago and it should be...

Gambar 1. Tampilan Data Ulasan

B. Penerjemahan dan Pelabelan Sentimen

Data yang telah dibersihkan diterjemahkan secara otomatis menggunakan GoogleTranslator dari pustaka *deep-translator* agar kompatibel dengan pustaka analisis sentimen berbahasa Inggris. Setelah proses penerjemahan selesai, dilakukan pelabelan sentimen menggunakan VADER Lexicon dari pustaka *vaderSentiment*. Setiap ulasan dikategorikan ke dalam tiga jenis sentimen, yaitu positif apabila skor compound bernilai $\geq 0,05$, negatif apabila skor $\leq -0,05$, dan netral apabila nilainya berada di antara kedua batas tersebut.



Gambar 2. Diagram Batang Hasil Analisis Sentimen VADER pada Ulasan.

Gambar 2(a) memperlihatkan distribusi sentimen hasil analisis menggunakan VADER Lexicon terhadap 5.000 ulasan pengguna Candi Prambanan. Berdasarkan visualisasi tersebut, sebagian besar ulasan tergolong positif dengan persentase sebesar 82,5% atau sekitar 4.125 ulasan. Sementara itu, kategori netral mencakup 10,8% atau sekitar 540 ulasan, dan kategori negatif sebanyak 6,6% atau sekitar 330 ulasan. Gambar 2(b) menampilkan distribusi hasil analisis sentimen menggunakan *VADER Lexicon* terhadap 5.000 ulasan pengguna Candi Borobudur. Berdasarkan grafik tersebut, kategori positif mendominasi dengan persentase sebesar 73,7% atau sekitar 3.685 ulasan, diikuti oleh kategori netral sebesar 13,7% atau sekitar 685 ulasan, serta kategori negatif sebesar 12,6% atau sekitar 630 ulasan.

C. Pra-pemrosesan Teks

Setelah proses pelabelan selesai, dilakukan tahapan text preprocessing untuk menyiapkan data agar siap diolah pada tahap pembobotan dan klasifikasi. Pra-pemrosesan bertujuan untuk menormalkan struktur teks, menghapus unsur yang tidak relevan, serta menyamakan bentuk kata agar analisis menjadi lebih konsisten. Tahapan ini terdiri atas lima langkah utama, yaitu *case folding*, *slang word normalization*, *tokenizing*, *filtering*, dan *stemming*.

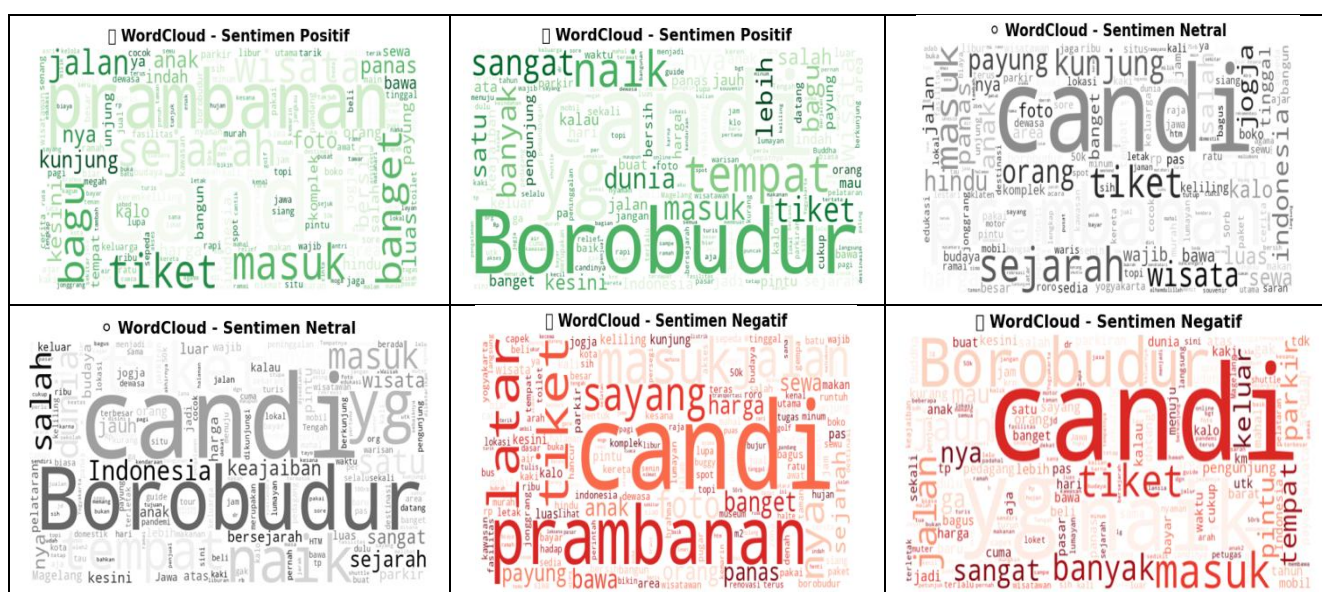
Tahap *case folding* dilakukan untuk menyeragamkan penulisan dengan mengubah seluruh huruf menjadi huruf kecil. Selanjutnya, *slang word normalization* mengganti kata tidak baku atau singkatan ke bentuk formal menggunakan pustaka *indoNLP*, agar makna teks lebih akurat. Tahapan *tokenizing* memecah kalimat menjadi potongan kata, kemudian *filtering* menghapus kata umum (*stopwords*) yang tidak berkontribusi terhadap makna utama kalimat. Langkah terakhir, *stemming*, mengubah kata ke bentuk dasarnya menggunakan pustaka *Sastrawi* agar variasi kata yang memiliki akar sama dapat dianggap identik.

Tabel 1. Contoh Hasil Proses Pra-pemrosesan Teks

Tahapan	Hasil
Teks Asli	Candi Prambanan adalah tempat wisata yang bagus bagi pecinta sejarah, budaya, dan arsitektur bangunan zaman dulu. Didalamnya juga terdapat museum. Untuk toilet jg banyak tersedia

Tahapan	Hasil
Case Folding	candi prambanan adalah tempat wisata yang bagus bagi pecinta sejarah, budaya, dan arsitektur bangunan zaman dulu. didalamnya juga terdapat museum. untuk toilet jg banyak tersedia
Slang Word Normalization	candi prambanan adalah tempat wisata yang bagus bagi pecinta sejarah, budaya, dan arsitektur bangunan zaman dulu. didalamnya juga terdapat museum. untuk toilet juga banyak tersedia
Tokenizing	['candi', 'prambanan', 'adalah', 'tempat', 'wisata', 'yang', 'bagus', 'bagi', 'pecinta', 'sejarah', 'budaya', 'dan', 'arsitektur', 'bangunan', 'zaman', 'dulu', 'didalamnya', 'juga', 'terdapat', 'museum', 'untuk', 'toilet', 'juga', 'banyak', 'tersedia']
Filtering	['candi', 'prambanan', 'wisata', 'bagus', 'pecinta', 'sejarah', 'budaya', 'arsitektur', 'bangunan', 'zaman', 'didalamnya', 'museum', 'toilet', 'tersedia']
Stemming	['candi', 'prambanan', 'wisata', 'bagus', 'cinta', 'sejarah', 'budaya', 'arsitektur', 'bangun', 'zaman', 'dalam', 'museum', 'toilet', 'sedia']

Sebagai hasil dari proses stemming dan pelabelan sentimen, dilakukan visualisasi kata menggunakan *Wordcloud* untuk menampilkan kata-kata yang paling sering muncul pada setiap kategori sentimen.



Gambar 3. Visualisasi Worcloud Sentimen pada Dataset

Pada kategori positif, kata seperti “bagus”, “indah”, “bersih”, dan “nyaman” mendominasi, yang menunjukkan bahwa mayoritas ulasan wisatawan bersifat apresiatif terhadap keindahan dan kenyamanan situs Candi Prambanan. Pada kategori netral, kata dominan seperti “candi”, “tiket”, “sejarah”, dan “wisata” merepresentasikan ulasan yang bersifat informatif tanpa ekspresi emosional yang kuat, umumnya berkaitan dengan informasi lokasi, harga tiket, atau pengalaman umum berkunjung. Sementara itu, pada kategori negatif, kata seperti “panas”, “harga”, “parkir”, dan “sayang” lebih sering muncul, menggambarkan keluhan wisatawan terhadap kondisi lingkungan, biaya, maupun fasilitas.

Analisis *wordcloud* sentimen pengunjung Candi Borobudur menunjukkan pengalaman yang terpolarisasi. Di satu sisi, sentimen positif sangat dominan, dengan pengunjung mengagumi keindahan, kemegahan, dan kebersihan candi, serta menghargai nilai sejarahnya. Namun, di sisi lain, sentimen negatif secara jelas menyoroti keluhan utama terkait harga tiket yang dianggap mahal, serta tantangan logistik seperti lokasi parkir yang jauh dan kondisi cuaca yang panas. Sementara itu, ulasan netral cenderung bersifat deskriptif, berfokus pada fakta-fakta kunjungan tanpa memberikan penilaian emosional yang kuat.

D. Pembobotan Kata dengan TF-IDF

Pembobotan dilakukan menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) dari pustaka scikit-learn. Metode ini menghitung tingkat kepentingan setiap kata berdasarkan frekuensi kemunculannya dalam dokumen. Hasilnya digunakan sebagai input numerik untuk model klasifikasi

Naive Bayes. Kata dengan bobot TF–IDF tertinggi merepresentasikan kata yang paling berpengaruh dalam menentukan polaritas sentimen.

Tabel 2. Kata dengan Bobot TF-IDF Tertinggi pada Ulasan Pengguna

Candi Prambanan		Candi Borobudur		Candi Prambanan		Candi Borobudur	
term	tfidf	term	tfidf	term	tfidf	term	tfidf
makan	0.220097	cocok	0.216206	mepet	0.146871	minum	0.145970
pas	0.206632	nya	0.212398	fyi	0.141099	sewa	0.137293
candi	0.190941	jalan	0.197347	sewa	0.134088	temple	0.131743
sebrang	0.168914	bawa	0.194400	85rb	0.130850	toko	0.131335
sepeda	0.163174	oleh2	0.176435	pernik	0.129871	bilang	0.130151
tikar	0.163141	poll	0.166450	orang	0.129744	anak	0.116455
prambanan	0.162265	gedung	0.166450	area	0.129669	jasa	0.115675
pakai	0.149943	scan	0.154097	pernak	0.128046	antri	0.114069
hampar	0.149234	orang	0.150432	jenis	0.128046	loket	0.110812
kafe	0.146871	masuk	0.149332	plus	0.125586	kayak	0.110644

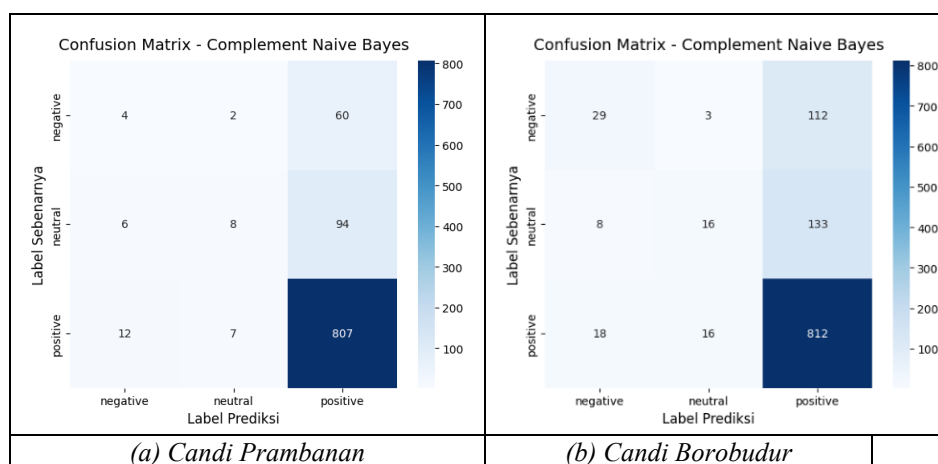
E. Klasifikasi dengan Naive Bayes

Model klasifikasi yang digunakan adalah *Complement Naive Bayes*, varian yang efektif untuk data teks dengan distribusi kelas tidak seimbang. Data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Model dilatih menggunakan matriks TF-IDF sebagai fitur dan pelabelan sebagai target.

Tabel 3. Hasil evaluasi Model Klasifikasi Sentimen pada Dataset

Prambanan	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>	Borobudur	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
<i>negative</i>	0.18	0.06	0.09	66	<i>negative</i>	0.53	0.20	0.29	144
<i>neutral</i>	0.47	0.07	0.13	108	<i>neutral</i>	0.46	0.10	0.17	157
<i>positive</i>	0.84	0.98	0.90	826	<i>positive</i>	0.77	0.96	0.85	846
<i>accuracy</i>			0.82	1000	<i>accuracy</i>			0.75	1147
<i>macro avg</i>	0.50	0.37	0.37	1000	<i>macro avg</i>	0.58	0.42	0.44	1147
<i>weighted avg</i>	0.76	0.82	0.77	1000	<i>weighted avg</i>	0.70	0.75	0.69	1147

Model memperoleh akurasi sebesar 81,9%, artinya sebagian besar data ulasan berhasil diklasifikasikan dengan benar. Kategori positif memiliki hasil terbaik dengan nilai *precision* 0,84 dan *recall* 0,98, sedangkan kategori netral dan negatif masih sering salah terdeteksi sebagai sentimen positif.



Gambar 4. Confusion Matrix Hasil Klasifikasi pada dataset

Berdasarkan hasil pengujian, model memperoleh akurasi sebesar 81,9%, artinya sebagian besar data ulasan berhasil diklasifikasikan dengan benar. Kategori positif memiliki hasil terbaik dengan nilai *precision* 0,84 dan *recall* 0,98, sedangkan kategori netral dan negatif masih sering salah terdeteksi sebagai

sentimen positif. Model memperoleh hasil akurasi sebesar 74,7%, artinya sebagian besar data ulasan berhasil diklasifikasikan dengan benar. Kategori positif memiliki hasil terbaik dengan nilai *precision* 0,77 dan *recall* 0,96, menunjukkan kemampuan model yang kuat dalam mengenali opini positif wisatawan. Namun demikian, kelas netral dan negatif masih cenderung salah diklasifikasikan sebagai positif, karena adanya kemiripan kata bernada umum dan ekspresi halus yang sulit dibedakan oleh model berbasis frekuensi kata.

V. KESIMPULAN DAN SARAN

Hasil analisis sentimen terhadap 10.000 ulasan Google Maps menunjukkan bahwa Candi Prambanan memperoleh proporsi sentimen positif yang lebih tinggi dibandingkan Candi Borobudur, meskipun jumlah data yang dianalisis sama, yaitu 5.000 ulasan masing-masing destinasi. Sebanyak 82,5% ulasan Prambanan bersentimen positif dan hanya 6,6% negatif, sedangkan Borobudur memiliki 73,7% sentimen positif dan 12,6% negatif. Temuan ini mengindikasikan bahwa wisatawan menilai Prambanan sebagai destinasi yang lebih nyaman dan terawat, dengan tingkat kepuasan yang lebih tinggi terhadap pengelolaan kawasan dan fasilitas wisata.

Analisis *wordcloud* memperkuat temuan tersebut, di mana kata kunci negatif pada Prambanan berfokus pada isu “panas”, “harga”, dan “parkir”, sedangkan ulasan positif pada Borobudur didominasi oleh kata “bagus”, “indah”, dan “megah” yang merepresentasikan apresiasi wisatawan terhadap keindahan arsitektur dan nilai historisnya. Model klasifikasi *Complement Naive Bayes* yang diterapkan menunjukkan kinerja yang baik, dengan akurasi sebesar 81,9% untuk Prambanan dan 74,7% untuk Borobudur, sehingga membuktikan bahwa pendekatan *machine learning* efektif dalam mengenali pola opini wisatawan secara otomatis.

Berdasarkan hasil tersebut, pengelola kedua destinasi disarankan untuk memanfaatkan temuan ini sebagai dasar dalam pengambilan keputusan berbasis data (*data-driven decision making*). Peningkatan kualitas layanan dapat dilakukan melalui penyediaan area teduh, perluasan lahan parkir, perbaikan sistem manajemen pengunjung, serta evaluasi kebijakan harga tiket guna meningkatkan kenyamanan dan kepuasan wisatawan.

Penelitian ini memberikan kontribusi praktis dalam penerapan sentiment analysis pada pariwisata budaya, sekaligus membuka peluang pengembangan riset lanjutan menggunakan model *deep learning* seperti LSTM atau BERT untuk pemahaman konteks linguistik yang lebih dalam. Studi mendatang juga disarankan untuk memperluas sumber data dari berbagai platform ulasan daring agar hasil analisis lebih representatif terhadap persepsi wisatawan secara keseluruhan.

DAFTAR RUJUKAN

- [1] D. Nuryadi, F. Metandi, N. A. Hadiwijaya, M. Z. Rohman, S. Hartanto, A. Syafrizal and E. Yadie, "Fine Tuning Indobert Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Tiket.Com Di Google Play Store," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, 2025.
- [2] D. Rifaldi, A. Fadlil and Herman, "Teknik Preprocessing pada Text Mining Menggunakan Data Tweet 'Mental Health,'" *Jurnal Pendidikan Teknologi Informasi*, vol. 2, pp. 131-136, 2023.
- [3] P. K. A. Duta, P. D. S. Pitanatri and C. P. Loanata, "Big Data-Based Sentiment Analysis on Tripadvisor Reviews Using Naïve Bayes Classification: A Case Study on Luxury Resort in Bali," *Media Wisata*, vol. 23, 2025.
- [4] P. G. Rahardandi, M. Fauzan and M. P. Wicaksana, "Systematic Literature Review: Analisis Sentimen Pada Situs Google Dan Penerapannya," *Innovative: Journal Of Social Science Research*, vol. 4, pp. 8300-8317, 2024.
- [5] R. Merdiansah, S. and A. A. Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 7, pp. 221-228, 2024.

- [6] S. N. Budiman, S. Lestanti and E. , "Analisis Sentimen Berdasarkan Hasil Review Lokasi Google Map Menggunakan Natural Language Toolkit TextBlob dan Naïve Bayes," *JAMI: Jurnal Ahli Muda Indonesia*, vol. 5, p. 114–126, 2024.
- [7] Y. Asri, W. N. Suliyanti, D. Kuswardani and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile," *PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika*, vol. 15, pp. 221-230, 2022.
- [8] M. T. Anwar and D. R. A. Permana, "Analisis Sentimen Masyarakat Indonesia Terhadap Produk Kendaraan Listrik Menggunakan VADER," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, p. 783–792, 2023.
- [9] M. Gultom, J. Marikros, W. Rusli and V. C. Mawardi, "Penerapan Vader Sentiment untuk Mendeteksi Sentimen Bahasa Inggris berbasis Website," *Seminar Nasional Penelitian dan Pengabdian Kepada Masyarakat CORISINDO* , 2024.
- [10] M. S. Hadi, J. Akbar and M. F. Zulkarnain, "Analisis Sentimen Wisata Air Terjun Di Kabupaten Lombok Tengah Menggunakan MetodeSupport Vector Machine (SVM)," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 8, pp. 318-329, 2025.
- [11] S. Ernawati and R. Wati, "Evaluasi Performa Kernel SVM dalam Analisis Sentimen Review Aplikasi ChatGPT Menggunakan Hyperparameter dan VADER Lexicon," *Jurnal Buana Informatika*, vol. 15, 2024.
- [12] M. P. Syah, A. P. Wardani, M. Idhom and T. , "Perbandingan Representasi Teks Tf-Idf Dan Bert Terhadap Akurasi Cosine Similarity Dalam Penilaian Otomatis Jawaban Berbasis Teks," *Data Sciences Indonesia*, 2025.
- [13] M. I. Maulana, E. Budianita, M. Fikry and F. Yanto, "Klasifikasi Sentiment Ulasan Aplikasi Sausage Man Menggunakan VADER Lexicon dan Naive Bayes Clasifier," *Jurnal Sistem Komputer dan Informatika (JSON)*, 2023.
- [14] D. Rifaldi, T. S. Famuji, B. O. S. Miranda, F. P. Ramadhan, I. P. Mulyadi, V. Saputra and G. P. I. Fanani, "Evaluasi Sentimen Pengguna ChatGPT Menggunakan Naive Bayes: Tinjauan dari Confusion Matrix dan Classification Report," *Jurnal Riset Sistem dan Teknologi Informasi (RESTIA)*, vol. 3, 2025.
- [15] M. Y. Pratama and C. Crysdian, ""Klasifikasi User Review pada Aplikasi Online Travel Booking Menggunakan Multinomial Naïve Bayes,"" *The Indonesian Journal of Computer Science (IJCS)*, vol. 14, 2025.