

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Di era *digital* yang ditandai dengan maraknya *platform* ulasan *online*, suara wisatawan telah menjadi kompas penting dalam pengembangan destinasi pariwisata kelas dunia. Candi Borobudur sebagai salah satu warisan budaya dunia yang diakui *UNESCO* telah menjadi ikon pariwisata Indonesia yang mendunia. Candi Borobudur merupakan salah satu destinasi wisata bersejarah yang memiliki daya tarik besar bagi wisatawan domestik maupun mancanegara. Sebagai warisan budaya dunia yang diakui oleh *UNESCO*, Candi Borobudur memiliki nilai historis, religius, dan estetika yang tinggi. Candi Borobudur juga merupakan salah satu monumen Buddha terbesar di dunia dan menjadi simbol kejayaan peradaban Jawa kuno. Menurut data *UNESCO* (1991), Candi Borobudur dibangun pada abad ke-9 oleh dinasti Syailendra dan dikenal sebagai situs arkeologi terbesar di Indonesia, dengan dimensi luas dan tinggi yang signifikan. Keberadaan candi Buddha terbesar di dunia ini tidak hanya memiliki nilai historis dan religius yang tinggi, tetapi juga menjadi motor penggerak ekonomi kreatif di wilayah Magelang dan sekitarnya [1]. Dalam beberapa tahun terakhir, jumlah kunjungan wisatawan ke Candi Borobudur menunjukkan peningkatan yang signifikan, mencapai lebih dari 3 juta pengunjung per tahun sebelum pandemi *COVID-19* [2]. Namun, peningkatan jumlah pengunjung ini diiringi dengan berbagai keluhan dan masukan dari wisatawan yang tersebar di berbagai *platform* ulasan *online* seperti *Tiket.com* dan *Tripadvisor*. Review yang ditinggalkan mencerminkan beragam aspek perjalanan, termasuk kualitas sarana, tarif masuk, tingkat keramaian, serta performa pemandu wisata. [3], dalam konteks ini, muncul sejumlah permasalahan yang perlu diidentifikasi lebih lanjut. Pertama, volume ulasan yang sangat besar menyimpan potensi informasi yang bernilai tinggi, namun sulit dianalisis secara manual karena keterbatasan waktu dan sumber daya manusia. Kedua, tidak semua pengunjung meninggalkan ulasan, sehingga menyebabkan ketidakseimbangan data (*data*

imbalance) yang dapat memengaruhi akurasi analisis. Ketiga, belum terdapat kejelasan mengenai metode atau algoritma *machine learning* mana yang paling efektif untuk menganalisis *sentiment* ulasan wisatawan secara otomatis.

Perkembangan teknologi *digital* telah mengubah cara wisatawan dalam memilih dan mengevaluasi destinasi wisata. Dalam era digital, ulasan daring telah menjadi rujukan utama bagi pengunjung dalam memilih destinasi liburan. [4]. Fenomena ini menjadikan data ulasan pengunjung sebagai sumber informasi yang sangat berharga bagi pengelola destinasi wisata, berikut dibawah ini adalah gambar data pengunjung pengunjung candi borobudur dari domestik maupun mancanegara.



Bulan	Jumlah Pengunjung Obyek Wisata Candi Borobudur Menurut Asal Wisatawan dan Bulan di Kabupaten Magelang	
	Domestik	Mancanegara
	2023	2023
Januari	121.138	6.278
Februari	87.706	6.241
Maret	67.520	8.520
April	93.561	9.510
Mei	126.410	14.438
Juni	158.430	16.696
Juli	127.271	29.166
Agustus	58.900	34.092
September	63.997	24.852
Oktober	77.455	18.180
November	86.941	12.291
Desember	211.897	12.789
Kabupaten Magelang	1.281.226	193.053

Gambar 1. 1 Data Pengunjung Wisata Candi Borobudur

Sumber : [5]

Di satu sisi, ulasan positif dapat meningkatkan daya tarik suatu destinasi, sementara ulasan negatif yang tidak ditangani dengan baik dapat menurunkan minat wisatawan [6]. Salah satu kendala bagi pengelola adalah banyaknya jumlah ulasan harian yang sulit ditangani secara manual karena keterbatasan waktu dan tenaga. Selain itu, karakteristik ulasan yang tidak terstruktur dan mengandung bahasa informal serta *slang* turut menyulitkan proses analisis konvensional [7].

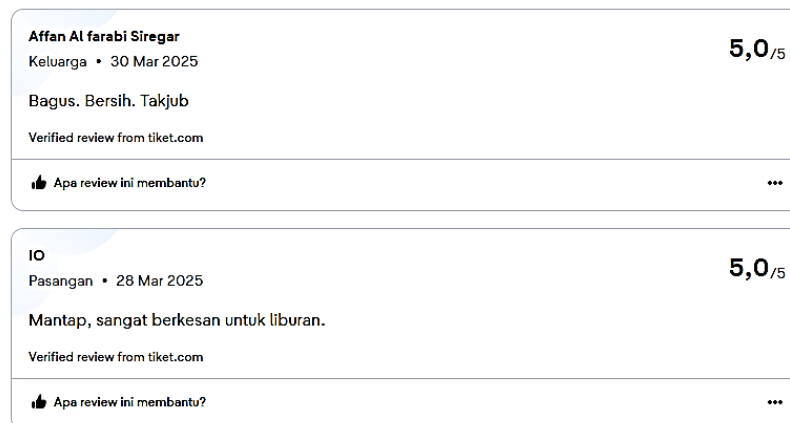
Berdasarkan permasalahan tersebut, penelitian ini menawarkan solusi berupa penerapan analisis sentimen otomatis guna mengelompokkan opini pengunjung tentang Candi Borobudur. Analisis sentimen merupakan cabang dari *text mining* yang mampu mengekstrak dan mengklasifikasikan opini dari data teks tidak terstruktur [8]. Dalam penelitian ini, dua algoritma *machine learning* yaitu *Naive Bayes* dan *Support Vector Machine (SVM)* akan dibandingkan performanya. *Naive Bayes* dipilih karena kesederhanaan dan efektivitasnya dalam menangani klasifikasi teks, terutama untuk data dalam jumlah besar [9]. Sementara itu, *SVM* dipilih karena kemampuannya dalam menangani data berdimensi tinggi dan *non-linear* dengan menggunakan *kernel trick* [10]. Kedua algoritma ini akan diimplementasikan pada *dataset* ulasan yang dikumpulkan dari *Tiket.com* dan *Tripadvisor*, dua *platform* ulasan wisata yang paling banyak digunakan di Indonesia [11].

Pada gambar 1.2 dan 1.3 adalah sample ulasan pengguna pada aplikasi *Tiket.com* dan *Tripadvisor*.



Gambar 1. 2 Ulasan Pengguna Tripadvisor

Sumber : [12]



Gambar 1. 3 Ulasan Pengguna Tiket.com

Sumber : [13]

Penelitian terdahulu tentang analisis sentimen *multibahasa* menunjukkan tantangan unik dalam *preprocessing*. Studi oleh Rianto et al. (2021) menemukan bahwa *stemming* berlebihan pada teks campuran Inggris-Indonesia justru menurunkan akurasi [14]. Temuan ini sejalan dengan analisis data awal kami yang didominasi *slang* internet (e.g., 'lucuuk', 'recommended') dan minim unsur bahasa daerah. Oleh karena itu, teknik *preprocessing* yang dikembangkan meliputi: (1) normalisasi *slang* Indonesia-Inggris (e.g., 'worth it' → 'bermanfaat'), (2) penghapusan *stopword* gabungan dua bahasa, serta (3) *stemming* selektif yang hanya diterapkan pada teks berbahasa Indonesia. Implementasi ini berbeda dengan metode R. V. Lucky et al. (2023) [15] yang berfokus pada bahasa Jawa, karena karakteristik *dataset* kami tidak mendukung pendekatan tersebut."

Hasil dari penelitian ini diharapkan dapat memberikan beberapa kontribusi penting. Pertama, bagi pengelola Candi Borobudur, hasil analisis dapat menjadi dasar pengambilan keputusan untuk perbaikan fasilitas dan pelayanan. Kedua, bagi akademisi, penelitian ini dapat menambah literatur tentang penerapan *text mining* dalam sektor pariwisata budaya. Ketiga, bagi pengembang sistem, penelitian ini memberikan gambaran tentang perbandingan performa dua algoritma klasifikasi teks pada data ulasan dalam bahasa

Indonesia. Dengan demikian, penelitian ini tidak hanya memiliki nilai akademis tetapi juga aplikasi praktis dalam pengembangan pariwisata berkelanjutan [16].

1.2 Identifikasi Masalah

Pada latar belakang “Perbandingan Algoritma *Support Vector Machine* Dan *Naïve Bayes* Pada Analisis Sentimen Ulasan Candi Borobudur.” beberapa masalah utama telah diidentifikasi sebagai berikut:

1. volume ulasan pengunjung yang besar di platform daring menyimpan informasi berharga tentang pengalaman wisata, namun sulit dianalisis secara manual karena keterbatasan waktu dan sumber daya;
2. data ulasan cenderung tidak seimbang akibat tidak semua pengunjung meninggalkan ulasan, sehingga dibutuhkan pendekatan analisis yang mampu mengakomodasi kondisi ketidakseimbangan data;
3. belum terdapat kejelasan mengenai algoritma *machine learning* yang paling efektif dalam menganalisis sentimen ulasan wisatawan, sehingga perlu dilakukan perbandingan performa antara algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes*.

1.3 Rumusan Masalah

Berdasarkan uraian dalam latar belakang, rumusan masalah penelitian ini adalah:

1. Bagaimana ulasan wisatawan pada platform *Tiket.com* dan *Tripadvisor* dapat dijadikan dasar evaluasi untuk meningkatkan fasilitas, manajemen keramaian, dan pelayanan pemandu wisata di Candi Borobudur?
2. Bagaimana performa algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen pada *dataset* ulasan Candi Borobudur yang tidak terstruktur?
3. Bagaimana melakukan analisis sentimen yang efektif terhadap ulasan Candi Borobudur yang mengandung karakteristik unik berupa Campuran bahasa Indonesia-Inggris (*code-mixing*).

1.4 Batasan Masalah

Agar penelitian ini fokus dan terukur, penulis membatasi ruang lingkup sebagai berikut:

1. data yang digunakan merupakan ulasan wisatawan dari dua platform, yaitu *Tiket.com* sebanyak 2.642 ulasan dan *Tripadvisor* sebanyak 1.817 ulasan, dengan total awal sebanyak 4.459 ulasan. Setelah dilakukan proses *data cleaning* awal, jumlah data yang digunakan dalam penelitian ini berkurang menjadi 4.341 ulasan, dengan periode pengambilan data mulai dari 1 September 2010 hingga 30 Maret 2025;
2. penelitian ini hanya menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)*;
3. klasifikasi sentimen dibatasi pada dua label yaitu positif dan negatif.

1.5 Tujuan Penelitian

Tujuan pada penelitian ini adalah sebagai berikut:

1. mengevaluasi ulasan wisatawan pada platform *Tiket.com* dan *Tripadvisor* sebagai bahan rekomendasi untuk meningkatkan: Fasilitas pendukung (toilet, area istirahat, transportasi internal), sistem pembayaran tiket (elektronik/konvensional), dan kualitas pelayanan pemandu wisata;
2. mengukur dan membandingkan performa algoritma *Naïve Bayes* dan *SVM* dalam klasifikasi sentimen positif dan negatif, dengan fokus pada evaluasi metrik akurasi, precision, recall, dan F1-score sebagai dasar pengambilan kesimpulan model terbaik;
3. melakukan analisis sentimen terhadap 4.341 data ulasan Candi Borobudur pada ulasan *Tiket.com* dan *Tripadvisor*.

1.6 Manfaat Penelitian

Berdasarkan tujuan penelitian, manfaat yang diharapkan adalah sebagai berikut:

1. mengetahui pola sentimen wisatawan terhadap Candi Borobudur melalui analisis ulasan di *Tiket.com* dan *Tripadvisor*, termasuk aspek dominan yang memengaruhi kepuasan (fasilitas, pelayanan, keramaian);
2. mengetahui tingkat akurasi algoritma *Naïve Bayes* dan *SVM* dalam mengklasifikasikan sentimen pada data kompleks (*slang*, *code-mixing*, istilah lokal), termasuk kelebihan dan kelemahan masing-masing;
3. menghasilkan *dataset* berlabel ulasan Candi Borobudur yang dapat digunakan untuk penelitian sejenis di masa depan.

1.7 Sistematika Penulisan

Untuk memudahkan serta mengarahkan dalam penulisan, maka dibuat sistematika penulisan dan dibagi menjadi lima bab disusun sebagai berikut:.

BAB I : PENDAHULUAN

Bab ini menjelaskan latar belakang, identifikasi, rumusan masalah, batasan, tujuan, manfaat penelitian, serta sistematika penulisan.

BAB II : TINJAUAN PUSTAKA

Bab ini memuat tinjauan penelitian terdahulu dan teori-teori yang mendukung topik penelitian.

BAB III : METODOLOGI PENELITIAN

Bab ini menjelaskan tentang objek penelitian, kerangka penelitian, metode pengumpulan data dan metode analisis data.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini memuat penjelasan hasil dan pemahaman mengenai hasil dari analisis yang dilakukan pada saat penelitian.

BAB V : PENUTUP

Bab ini merupakan kesimpulan terhadap penelitian yang telah dilakukan serta saran yang diberikan kepada pembaca untuk dilanjutkan atau dikembangkan pada penelitian selanjutnya.