

Evaluasi Efektivitas Fitur Layanan Candi Borobudur Berdasarkan Analisis Sentimen Ulasan *Google Maps* Menggunakan Teknik *Naive Bayes*

Meyer Mega Eklesia Silaban¹, Rosena Shintabella², Sahid Triambudhi³

^{1,2,3} Politeknik Negeri Indramayu, Jalan Raya Lohbener Lama Nomor 08, Kabupaten Indramayu Kode Pos 45252

e-mail: meyermega@polindra.ac.id¹, rosenasbella@polindra.ac.id², sahid@polindra.ac.id³

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi : 4 April 2026

Revisi Akhir :13 Mei 2026

Diterbitkan Online : 29 Mei 2026

Kata Kunci:

Analisis Sentimen, Pariwisata Digital, *Google Maps*, Candi Borobudur, *Naive Bayes*.

Korespondensi :

Telepon / Hp : +6282376681773

E-mail : meyermega@polindra.ac.id

ABSTRAK

Perkembangan pariwisata digital telah mentransformasi interaksi wisatawan dengan destinasi wisata, menjadikan platform ulasan daring seperti *Google Maps* sebagai sumber informasi penting sekaligus aset data yang bernilai untuk evaluasi kualitas layanan. Penelitian ini bertujuan untuk menganalisis persepsi wisatawan terhadap efektivitas fitur layanan Candi Borobudur berdasarkan sentimen ulasan *Google Maps* menggunakan algoritma *Naive Bayes*. Data penelitian berupa 1000 ulasan pengguna *Google Maps* yang dikumpulkan melalui observasi digital dan proses ekspor data ulasan pada periode tertentu. Tahapan penelitian meliputi pengumpulan data, preprocessing teks menggunakan Orange Data Mining, pelabelan sentimen otomatis berbasis lexicon multibahasa, ekstraksi fitur TF-IDF, serta klasifikasi sentimen menggunakan algoritma *Naive Bayes* ke dalam tiga kelas, yaitu positif, netral, dan negatif. Evaluasi model dilakukan menggunakan metode 5-fold *cross validation*. Hasil penelitian menunjukkan bahwa model *Naive Bayes* memperoleh performa yang baik dengan nilai *accuracy* sebesar 0,894, *precision* sebesar 0,896, *recall* sebesar 0,894, *F1-score* sebesar 0,894, dan *Area Under Curve (AUC)* sebesar 0,983. Hasil analisis menunjukkan bahwa sebagian besar ulasan pengunjung cenderung memiliki sentimen positif, khususnya terhadap aspek keindahan wisata, fasilitas, dan pengalaman berkunjung, meskipun masih ditemukan sentimen negatif terkait kepadatan pengunjung, harga tiket, dan pengelolaan fasilitas tertentu. Penelitian ini diharapkan dapat menjadi sumber evaluasi berbasis data bagi pengelola wisata dalam meningkatkan kualitas layanan dan pengalaman pengunjung secara berkelanjutan.

1. PENDAHULUAN

Pariwisata merupakan salah satu sektor strategis dalam pembangunan nasional yang memberikan kontribusi signifikan terhadap perekonomian, baik melalui peningkatan Produk Domestik Bruto (PDB) maupun penciptaan lapangan kerja di suatu negara [1]. Tempat wisata biasanya banyak dikunjungi oleh para wisatawan pada saat hari libur sebagai tempat untuk menghabiskan waktu bersama orang terdekat, mencari kesenangan, dan menyegarkan pikiran. Namun, potensi yang besar pada sektor pariwisata dihadapkan pada sebuah tantangan utama, yaitu tentang bagaimana mengelola destinasi wisata agar senantiasa dapat memberikan layanan yang berkualitas dan sesuai dengan harapan para wisatawan [2].

Candi Borobudur merupakan salah satu destinasi wisata prioritas nasional yang menerima jutaan kunjungan wisatawan setiap tahunnya. Tingginya aktivitas wisata tersebut turut menghasilkan ribuan ulasan digital pada platform *Google Maps* yang memuat opini pengunjung mengenai kualitas layanan, fasilitas, kebersihan, aksesibilitas, harga tiket, dan pengalaman wisata secara keseluruhan. Besarnya volume ulasan digital tersebut menjadikan pengelolaan opini publik secara manual menjadi kurang efektif, sehingga diperlukan pendekatan komputasional berbasis analisis sentimen untuk memperoleh evaluasi layanan secara lebih objektif dan terukur.

Seiring berjalannya waktu, perkembangan teknologi informasi dan komunikasi melalui platform *online* semakin memudahkan para wisatawan dalam menggali informasi mengenai sebuah destinasi wisata [3]. Salah satu platform *online* yang paling banyak digunakan adalah *Google Maps*, sebuah layanan di mana masyarakat dapat menuliskan ulasan untuk berbagai tempat yang pernah mereka kunjungi. Saat ini, dinamika perilaku wisatawan digital dituntut semakin bergantung pada ulasan-ulasan daring sebagai referensi utama untuk menentukan destinasi mereka [2]. Bagi pihak pengelola wisata, mengkaji ulasan pengunjung ini sangat penting karena dapat memberikan wawasan berharga untuk lebih memahami persepsi pelanggan serta mengoptimalkan layanan operasional yang ada [3]. Melalui ulasan masyarakat di *Google Maps* tersebut, pengelola juga dapat melihat secara spesifik apakah penilaian masyarakat cenderung mengarah ke sentimen positif atau negatif terhadap tempat wisata yang dikunjunginya [4].

Meski begitu, jumlah ulasan daring yang sangat masif kerap membuat proses analisis secara manual menjadi kaku, rumit, dan tidak efisien untuk mencapai akurasi yang tinggi. Oleh karena itu, diperlukan sebuah pendekatan komputasi otomatis, salah satunya melalui analisis sentimen [1]. Analisis sentimen merupakan proses penambangan opini (*opinion mining*) untuk memahami, mengekstrak, dan mengolah data tekstual

secara otomatis guna mendapatkan informasi mengenai kecenderungan pandangan pengguna dalam sebuah kalimat [5]. Sistem analisis sentimen ini sangat relevan diterapkan pada data ulasan dari platform seperti *Google Maps* untuk menilai kepuasan pengunjung.

Dalam praktiknya, pemanfaatan algoritma *machine learning* (pembelajaran mesin) pada analisis sentimen terbukti dapat memberikan hasil klasifikasi teks yang efektif. Salah satu metode algoritma yang paling sering digunakan untuk menganalisis sentimen berdasar pada probabilitas dan *Teorema Bayes* adalah algoritma *Naive Bayes Classifier* [6]. Algoritma *Naive Bayes* dinilai sangat cocok untuk digunakan ketika terdapat jumlah masukan data teks ulasan yang sangat besar. Di samping itu, metode klasifikasi ini lebih banyak disukai karena memiliki kecepatan komputasi yang baik, kesederhanaan dalam modelnya, serta terbukti handal dalam mengklasifikasikan ulasan tempat wisata [7], [4].

Beberapa penelitian sebelumnya telah menerapkan analisis sentimen pada sektor pariwisata menggunakan algoritma *Naive Bayes*, termasuk penelitian terhadap ulasan wisata Candi Borobudur berbasis *Tripadvisor*. Namun demikian, penelitian terdahulu masih berfokus pada analisis opini umum wisatawan dan belum secara spesifik mengevaluasi efektivitas fitur layanan destinasi wisata berbasis ulasan *Google Maps* yang bersifat lebih real-time, kontekstual, dan terhubung langsung dengan pengalaman lokasi wisata. Oleh karena itu, penelitian ini memiliki kontribusi dalam mengevaluasi kualitas layanan Candi Borobudur melalui klasifikasi sentimen ulasan *Google Maps* menggunakan algoritma *Naive Bayes*. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen masyarakat dari kumpulan ulasan di platform *Google Maps* demi mengukur efektivitas dan performa layanan wisata di kawasan Candi Borobudur menggunakan metode algoritma *Naive Bayes*. Dengan demikian, penelitian ini diangkat dengan judul "**Evaluasi Efektivitas Fitur Layanan Candi Borobudur Berdasarkan Analisis Sentimen Ulasan *Google Maps* Menggunakan Teknik *Naive Bayes***".

2. LANDASAN TEORI

2.1 Information Retrieval

Pengambilan Informasi (*Information Retrieval/IR*) adalah proses mencari dokumen atau data tidak terstruktur (seperti teks) dari koleksi yang besar untuk memenuhi kebutuhan pengguna. Tugas utamanya bukan sekadar menemukan data, melainkan menyaring informasi yang paling relevan berdasarkan kueri yang diberikan. Di tengah data yang masif dan tidak teratur, IR berfokus pada ketepatan hasil yang sesuai dengan konteks dan kebutuhan spesifik pengguna [8].

Pengambilan Informasi (*Information Retrieval/IR*) merupakan disiplin ilmu dan teknologi yang berfokus pada penemuan dokumen atau data tidak terstruktur—seperti teks, citra, maupun video—dalam koleksi berskala masif. Secara operasional, IR berfungsi sebagai jembatan antara gudang data yang kompleks dengan kebutuhan kognitif manusia. Tugas utamanya

melampaui sekadar pencocokan kata kunci (*keyword matching*); IR bertujuan untuk menyaring dan memeringkatkan (*ranking*) informasi yang paling relevan berdasarkan kueri atau representasi kebutuhan pengguna.

Di era ledakan data saat ini, tantangan IR semakin kompleks karena data seringkali hadir dalam format yang tidak terorganisir dan ambigu. Oleh karena itu, sistem IR modern kini mengintegrasikan teknik pemrosesan bahasa alami (NLP) dan pembelajaran mesin untuk memahami konteks semantik. Tujuannya adalah memastikan bahwa hasil yang disajikan tidak hanya akurat secara teknis, tetapi juga memiliki nilai guna yang tinggi dan sesuai dengan preferensi spesifik serta batasan konteks pengguna [9].

2.2 Analisis Sentimen

Sentiment Analysis merupakan suatu proses krusial dalam ranah *data mining* yang memiliki tujuan utama untuk mengidentifikasi, mengekstrak, dan memahami informasi terkait sentimen publik yang terkandung dalam kumpulan data tekstual. Lebih dari sekadar identifikasi, proses ini memungkinkan peneliti untuk menangkap nuansa sikap, pendapat, emosi, serta evaluasi yang diekspresikan oleh individu atau kelompok dalam teks tersebut, menjadikannya alat yang ampuh untuk memantau persepsi publik [10]. Secara fundamental, *Sentiment Analysis* seringkali difokuskan pada klasifikasi polaritas, yaitu menentukan apakah sebuah teks bersifat positif, negatif, atau netral. Namun, cakupannya dapat diperluas hingga ke klasifikasi subjektivitas, yang membedakan antara pernyataan faktual dan pernyataan opini. Berbagai pendekatan teoritis telah dikembangkan untuk mendukung tujuan ini, termasuk metode berbasis leksikon yang mengandalkan kamus kata-kata dengan skor sentimen tertentu, serta metode berbasis *machine learning* yang melatih model untuk mengenali pola sentimen dari data yang telah dilabeli.

2.3 Algoritma Naive Bayes

Sebagai metode probabilistik yang dibangun di atas *Teorema Bayes*, *Naive Bayes classifier* melakukan pengklasifikasian sederhana terhadap data secara efisien menggunakan training set yang tersedia. Inti dari metode ini terletak pada asumsi independensi fitur (*naive assumption*), yang menyatakan bahwa setiap fitur dalam himpunan data berkontribusi secara independen terhadap probabilitas suatu kelas, meskipun dalam kenyataannya asumsi ini sering kali dilanggar namun tetap memberikan hasil yang memuaskan dalam praktik [11].

Metode *Naive Bayes* menggunakan prediksi bersyarat dalam melakukan prediksi, yang secara fundamental dibangun di atas konsep probabilitas bersyarat. Sebelum menjelaskan lebih lanjut mengenai prediksi bersyarat, penting untuk memahami definisi prediksi itu sendiri.

2.4 Fitur Layanan Pariwisata

Dalam definisi luasnya, kepariwisataan mengartikan wisata sebagai sebuah tindakan atau gerakan perjalanan oleh individu atau kolektif (wisatawan) dengan motivasi rekreasi dan pelepas lelah dari rutinitas harian. Sementara itu, dalam konteks hukum, pariwisata kerap kali dikonseptualisasikan sebagai akumulasi kegiatan wisata yang didukung oleh penyediaan dan penatalayanan fasilitas serta jasa publik yang memadai [12]. Fasilitas dan layanan ini merupakan jejaring kompleks yang dapat dikelola dan disediakan oleh berbagai aktor, mulai dari masyarakat setempat yang berperan sebagai penyedia akomodasi dan jasa kuliner lokal, pengusaha di sektor pariwisata yang mengoperasikan berbagai jenis usaha seperti hotel dan biro perjalanan, hingga pemerintah daerah yang bertugas dalam pengembangan infrastruktur dan promosi destinasi.

Sektor pariwisata, dengan segala kompleksitasnya, merupakan salah satu sektor strategis yang secara konsisten diandalkan oleh pemerintah sebagai instrumen penting untuk mendatangkan devisa negara dan sebagai sumber pendapatan asli daerah (PAD) yang signifikan bagi provinsi maupun kota/kabupaten di mana destinasi wisata tersebut berlokasi [13]. Dalam konteks Candi Borobudur, identifikasi dan evaluasi fitur-fitur ini sangat penting untuk meningkatkan kualitas layanan dan kepuasan pengunjung.

2.1.5 Ulasan Daring (*Online Reviews*)

Ulasan daring adalah ulasan yang memiliki keunggulan jangkauan yang melampaui batas geografis, mampu menjangkau komunitas lokal yang jauh serta diakses oleh audiens global. Ulasan daring dapat didefinisikan sebagai komentar, penilaian, atau umpan balik yang dipublikasikan oleh konsumen pada platform digital seperti situs web, media sosial, dan aplikasi ulasan (misalnya, *Google Maps*, *TripAdvisor*). Komentar - komentar ini merefleksikan pengalaman, opini, dan rekomendasi konsumen terhadap produk, layanan, atau destinasi tertentu. Sebagai konsekuensinya, ulasan daring telah diakui sebagai sumber data yang sangat berharga untuk analisis sentimen karena sifatnya yang otentik dan kemampuannya untuk secara langsung mencerminkan persepsi pengguna [14]. Komentar, penilaian, atau umpan balik yang dipublikasikan oleh pengguna di platform digital seperti situs web, media sosial, dan aplikasi ulasan (e.g., *Google Maps*, *TripAdvisor*).

2.1.6 State of The Art

Berbagai studi terdahulu telah secara ekstensif mengeksplorasi potensi ulasan digital sebagai indikator kepuasan pengunjung di destinasi wisata prioritas, namun upaya mengevaluasi efektivitas fitur layanan secara spesifik di Candi Borobudur melalui pemetaan sentimen berbasis algoritma *Naive Bayes* masih menjadi

diskursus yang krusial untuk mengoptimalkan manajemen pengalaman wisatawan.

Tabel 1. Penelitian Terdahulu

Judul	Peneliti	Tahun	Hasil Utama
Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN	Prasetyo et al.	2020	Hasil penelitian menunjukkan bahwa metode KNN memberikan tingkat akurasi yang lebih tinggi dibandingkan dengan metode NB dalam mengklasifikasikan sentimen terkait pemindahan IKN. Metode KNN mencapai akurasi 88.12%, presisi 93.98%, dan recall 81.53%. Sementara itu, metode NB menghasilkan akurasi 82.27%, presisi 86.36%, dan recall 76.93% .[15]
Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naïve Bayes Classifier	Yerik Afrianto Singgalen.	2022	Mengidentifikasi sentimen positif dan negatif dari ulasan, serta fitur-fitur yang paling banyak dibicarakan.[16]
Penerapan Algoritma <i>Naive Bayes</i> untuk Analisis Sentimen Review Data Twitter BMKG Nasional.	Darwis et al.	2021	Hasil penelitian menunjukkan bahwa algoritma <i>Naive Bayes</i> dapat digunakan untuk mengklasifikasikan tweet berdasarkan sentimen. Hasil perhitungan probabilitas kata dan dokumen menunjukkan bahwa tweet dapat dikategorikan ke dalam dokumen tweet opini yang sesuai dengan isi teksnya. [17]
Identifikasi Potensi Pariwisata di Desa Kawasan Borobudur Kabupaten Magelang Provinsi Jawa Tengah; Studi Kasus Desa Candirejo	Yovan et al.	2025	Menganalisis umpan balik pengunjung untuk menilai efektivitas fitur digital, namun menggunakan metode analisis kualitatif dan kuantitatif yang berbeda.[18]

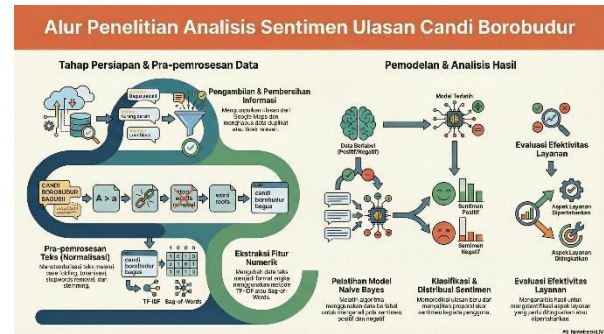
Judul	Peneliti	Tahun	Hasil Utama
Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes	Tanggraeni et al.	2022	Hasil pelabelan manual menunjukkan bahwa aplikasi "Sentuh Tanahku" menerima respon positif dari pengguna, dengan 407 ulasan positif dibandingkan dengan 235 ulasan negatif. [19]

3. METODOLOGI

Penelitian ini mengadopsi pendekatan kuantitatif sebagai landasan metodologisnya, dengan fokus utama pada penerapan metode analisis sentimen berbasis machine learning. Keputusan untuk menggunakan pendekatan kuantitatif didasarkan pada kebutuhan fundamental untuk mengukur persepsi publik secara objektif dan terukur, yang mana hanya dapat dicapai melalui akumulasi dan analisis data numerik yang dihasilkan dari klasifikasi sentimen [20]. Dalam ranah analisis sentimen, algoritma *Naïve Bayes* dipilih atas dasar efektivitasnya yang telah terbukti secara empiris dalam tugas klasifikasi teks, khususnya dalam domain ulasan daring yang kompleks dan kaya nuansa, serta kapabilitasnya dalam memproses volume data yang masif secara efisien, sebuah aspek krusial mengingat besarnya potensi data ulasan. Data ulasan yang dihimpun dari platform *Google Maps* dipilih sebagai sumber data primer karena menyajikan sampel yang luas dan beragam dari pengalaman aktual pengunjung, mencakup berbagai aspek layanan dan fasilitas yang ditawarkan. Pemilihan data *Google Maps* ini sejalan dengan studi terkini yang menekankan pentingnya memanfaatkan sumber data otentik dari pengguna untuk mendapatkan wawasan yang reliabel mengenai kepuasan pelanggan [21].

3.1. Alur Penelitian

Penelitian ini melaksanakan pendekatan kuantitatif dengan memanfaatkan analisis sentimen berbasis teks sebagai metode untuk mengukur pandangan pengunjung mengenai Candi Borobudur. Desain penelitian ini dirancang secara sekuensial, dimulai dari tahapan pengumpulan data hingga evaluasi akhir efektivitas fitur layanan. Proses dimulai dari pengumpulan sumber data primer berupa ulasan pengguna yang terpublikasi pada platform *Google Maps* Candi Borobudur. Data ulasan ini merupakan rekaman naratif yang berisi opini, pengalaman personal, serta penilaian kualitatif pengunjung terhadap berbagai aspek layanan dan fasilitas yang tersedia di destinasi wisata tersebut.



Gambar 1. Alur Penelitian

Tahapan awal penelitian adalah Pengambilan Informasi (*Information Retrieval*) untuk mengunduh data ulasan dari sumber yang ditentukan, dilanjutkan dengan Pembersihan Data. Tahapan ini memiliki sasaran untuk mengeliminasi data yang tidak relevan, duplikat, atau mengandung "noise" (karakter khusus, salah ketik) agar tercipta dataset yang bersih dan terstruktur. Keberhasilan analisis kuantitatif sangat bergantung pada kualitas data yang dihasilkan dari tahap krusial ini.

Selanjutnya, teks menjalani pra-pemrosesan untuk memastikan kesiapannya dalam analisis komputasional. Tahapan-tahapan yang dilakukan meliputi pengubahan semua huruf menjadi kecil (*case folding*), pemecahan teks menjadi token-token individual (*tokenisasi*), penghilangan kata-kata yang tidak berkontribusi pada makna (*stopwords removal*), dan pengurangan kata ke bentuk primernya (*stemming*) untuk mengelompokkan variasi kata yang sama.

Setelah normalisasi teks, dilakukanlah Ekstraksi Fitur guna mentransformasi data tekstual menjadi format numerik yang dapat diolah oleh algoritma machine learning. Pemilihan metode TF-IDF didasarkan pada kapasitasnya untuk menyeimbangkan frekuensi kemunculan kata dalam suatu dokumen dengan kekurangannya di seluruh kumpulan data (*korpus*), sehingga mampu menonjolkan kata-kata yang paling informatif dalam setiap ulasan. Dalam fase pelatihan model, algoritma *Naïve Bayes* diterapkan. Pemilihan algoritma ini didasarkan pada superioritasnya dalam klasifikasi teks, efektivitasnya pada kumpulan data yang terbatas, serta mudahnya interpretasi nilai probabilitas yang dihasilkan. Model dilatih dengan data ulasan yang telah diberi label sentimen untuk mengenali pola linguistik terkait persepsi positif, negatif, atau netral. Model yang telah melalui proses pelatihan selanjutnya difungsikan untuk mengkategorikan sentimen dari ulasan-ulasan baru yang belum diberi label, guna menentukan apakah ulasan tersebut mencerminkan pandangan positif, negatif, atau netral terhadap Candi Borobudur. Hasil klasifikasi sentimen akan disajikan secara kuantitatif, seperti ringkasan skor atau distribusi proporsi sentimen, untuk memberikan gambaran umum persepsi pengunjung. Visualisasi data dapat membantu menyajikan temuan ini secara efektif.

Tahap akhir metodologi adalah evaluasi efektivitas layanan dan pengelolaan Candi Borobudur berdasarkan analisis sentimen. Dengan mengaitkan sentimen dengan aspek spesifik seperti fasilitas atau kebersihan, penelitian

ini dapat mengidentifikasi kelebihan dan kekurangan dari perspektif pengunjung, yang dapat menjadi dasar pengembangan strategi peningkatan kualitas dan keberlanjutan destinasi wisata.

3.2. Pengumpulan Data

Data penelitian diperoleh dari ulasan pengguna Google Maps pada destinasi wisata Candi Borobudur melalui proses observasi digital dan ekspor data ulasan. Pengumpulan data dilakukan pada periode Januari 2025 hingga Februari 2025 dengan total 1000 ulasan pengguna berbahasa Indonesia. Proses seleksi data dilakukan dengan beberapa kriteria, yaitu ulasan harus relevan dengan pengalaman pengunjung terhadap layanan dan fasilitas wisata, tidak mengandung spam, tidak duplikat, serta memiliki isi teks yang dapat dianalisis. Ulasan yang hanya berisi emoji, simbol, atau teks yang tidak relevan dieliminasi pada tahap preprocessing.

Observasi dalam ranah digital merupakan metode fundamental dalam pengumpulan data untuk studi ini, yang memungkinkan peneliti untuk memantau fenomena persepsi publik secara langsung dalam situasi aslinya lewat platform daring. Proses observasi dilakukan tanpa partisipasi dengan mengumpulkan ulasan dari pengguna *Google Maps* terhadap destinasi wisata Candi Borobudur. Pendekatan ini memungkinkan peneliti untuk merekam opini, pengalaman, serta penilaian pengguna yang muncul secara alami tanpa intervensi, sehingga data yang didapatkan mencerminkan kondisi riil yang ada.

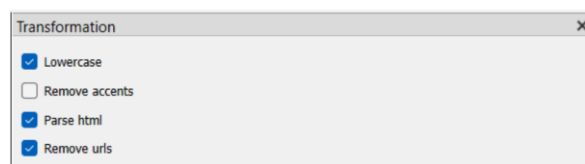
Sebagaimana observasi, data sekunder yang dikumpulkan juga mencakup ulasan pengguna yang telah tersedia secara publik dalam periode waktu tertentu. Data ini menjadi acuan utama dalam merepresentasikan persepsi masyarakat mengenai mutu layanan dan fasilitas. Ulasan-ulasan tersebut lantas dikompilasi dan diseleksi berdasarkan patokan yang telah ditetapkan, seperti relevansi, kelengkapan, dan rentang waktu publikasi, demi menghasilkan kumpulan data yang memadai untuk analisis lebih mendalam.

Untuk memperkuat dasar penelitian, studi literatur juga dimanfaatkan sebagai komponen integral. Aktivitas ini melibatkan pencarian dan analisis berbagai literatur ilmiah, termasuk artikel jurnal, makalah konferensi, dan buku, yang berkaitan dengan analisis sentimen, *text mining*, dan *Business Intelligence*. Tujuan utama studi literatur adalah untuk membangun landasan teoretis yang solid, mengerti metodologi riset terdahulu, dan mengenali area riset yang masih dapat dieksplorasi.

Dengan mengintegrasikan observasi data digital, data sekunder dari ulasan pengguna, dan studi literatur, penelitian ini berupaya menghasilkan data yang triangulatif dan komprehensif. Strategi ini dirancang untuk menguatkan validitas dan reliabilitas temuan penelitian, serta memberikan gambaran yang lebih akurat mengenai pandangan publik terhadap subjek yang diteliti.

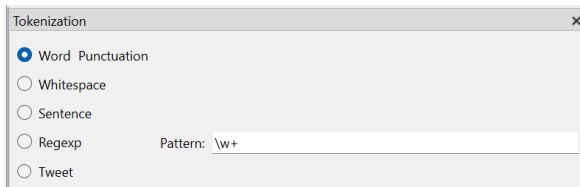
3.3. Persiapan Data

Proses awal persiapan data meliputi serangkaian kegiatan, yaitu seleksi, pembersihan, dan pra-pemrosesan. Tahap seleksi hanya mengikutsertakan kolom atau atribut teks yang relevan dari data yang tersedia, tanpa menggunakan keseluruhan data. Proses pembersihan data dilakukan setelah atribut teridentifikasi, yang bertujuan untuk membuang spam atau data yang tidak bersesuaian dengan tujuan riset, khususnya pendapat mengenai destinasi wisata. Tahap berikutnya adalah pra-pemrosesan teks yang mencakup transformasi, tokenisasi, dan penyaringan. Seluruh tahapan dalam proses penanganan data, yang mencakup tahap awal pembersihan teks hingga diperolehnya output akhir, telah diselesaikan dengan memanfaatkan Software Orange. Tujuan utama pra-pemrosesan teks adalah menata data agar terstruktur dan siap untuk tahap analisis lebih lanjut [19]. Sementara itu, transformasi meliputi konversi struktur kalimat menjadi huruf kecil, perubahan kode HTML, serta penghapusan URL dalam teks, sebagaimana ditunjukkan pada Gambar 2.



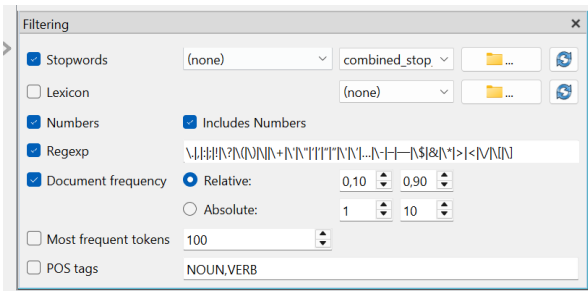
Gambar 2. Transformation pada tahap Pre-Processing

Untuk tahap penelitian selanjutnya, tokenisasi adalah proses awal pemrosesan data teks yang penting, berfungsi memecah teks menjadi unit-unit analisis. Proses ini, yang secara esensial mengubah urutan karakter menjadi entitas terpisah untuk analisis lebih lanjut [20], diadopsi dalam studi ini. Teknik yang digunakan adalah *word punctuation*, yang memilah kalimat berdasarkan kata dan tanda baca, seperti diilustrasikan pada Gambar 3. Pemilihan ini bertujuan memastikan kedalaman analisis yang memadai untuk data teks, guna memfasilitasi identifikasi dan penghitungan frekuensi kata yang akurat dalam proses ekstraksi informasi.



Gambar 3. Tokenization pada tahap Pre-Processing

Dalam tahapan pra-pemrosesan, penyaringan Gambar 4 bertugas memfilter token kata dari data korpus. Proses ini secara khusus mengeliminasi elemen yang dianggap tidak esensial, termasuk stopwords yang disesuaikan domainnya, angka, serta pola regex yang tidak dikehendaki. Pemilihan strategi penyaringan ini krusial agar analisis hanya melibatkan kata-kata bermakna, sehingga meningkatkan efektivitas pemodelan teks.

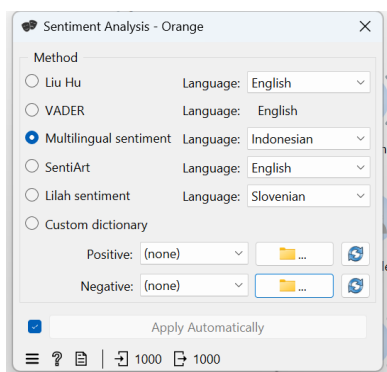


Gambar 4. Filtering pada tahap Pre-Processing

3.4. Pengolahan Data

Pasca data dipersiapkan, fokus beralih pada pengolahan data dengan menggunakan pendekatan Text Mining. Metode ini, yang mengadaptasi prinsip Data Mining untuk menemukan pola dan menambang informasi dari data tekstual demi memenuhi tujuan spesifik, menjadikan Sentiment Analysis sebagai salah satu teknik utamanya. Sentiment Analysis, yang merupakan salah satu turunan populer dari Text Mining, secara khusus ditujukan untuk menghimpun dan menganalisis pandangan masyarakat mengenai permasalahan yang ada.

Klasifikasi sentimen tweet yang siap diolah akan dikategorikan menjadi tiga kelas: positif, netral, dan negatif. Pada penelitian ini, Orange Data Mining digunakan sebagai tools pengolahan data dan implementasi workflow analisis teks. Metode lexicon multibahasa pada Orange digunakan pada tahap awal untuk melakukan pelabelan sentimen otomatis terhadap dataset ke dalam tiga kelas sentimen, yaitu positif, netral, dan negatif. Setelah proses pelabelan selesai dilakukan, algoritma *Naive Bayes* diterapkan sebagai model machine learning untuk proses klasifikasi sentimen dan evaluasi performa model menggunakan metode 5-fold cross validation.



Gambar 5. Tahap Pengolahan Data menggunakan *Sentiment Analysis* pada Orange

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Data Penelitian

Penelitian ini menggunakan seribu ulasan pengunjung mengenai Candi Borobudur yang diambil dari *Google Maps* melalui observasi digital. Data ulasan

yang terkumpul, baik yang singkat maupun deskriptif, mencerminkan pengalaman dan opini pengunjung tentang layanan dan fasilitas.

Berikut contoh ulasan berdasarkan masing-masing kelas sentimen:

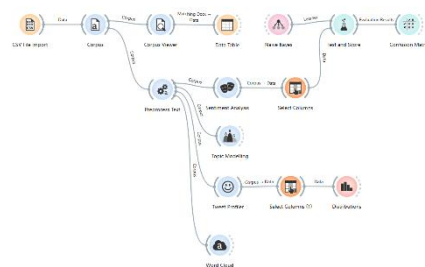
Tabel 2. Contoh Ulasan Sentimen

Kelas Sentimen	Contoh Ulasan
Positif	“Tempat wisata yang sangat indah dan megah, fasilitas cukup lengkap dan area cukup bersih.”
Netral	“Tempatnya ramai saat liburan, fasilitas cukup baik namun antrean cukup panjang.”
Negatif	“Harga tiket cukup mahal dan area wisata terlalu padat sehingga kurang nyaman.”

Karakteristik data meliputi teks tidak terstruktur berbahasa mayoritas Indonesia dengan sentimen beragam (positif, negatif, netral), serta mencakup topik seperti fasilitas, kebersihan, pelayanan, harga, aksesibilitas, dan keindahan wisata. Setelah melalui tahap seleksi dan pembersihan untuk menghilangkan data tidak relevan dan noise, ulasan ini menjadi dasar analisis sentimen menggunakan metode *Naive Bayes*. Contoh ulasan menunjukkan variasi opini, misalnya mengenai keindahan tempat, harga tiket, keramaian, kualitas pelayanan, serta kemegahan candi. Kumpulan 1000 ulasan ini dianggap representatif untuk memotret persepsi pengunjung dan selanjutnya digunakan untuk klasifikasi sentimen yang objektif.

4.2. Analisis Sentimen

Penelitian ini mengaplikasikan perangkat lunak *Orange Data Mining* untuk menjalankan analisis sentimen, guna mengelompokkan ulasan dari pengunjung Candi Borobudur ke dalam tingkatan sentimen positif, negatif, dan netral. Fokus analisis adalah mendalami persepsi pengunjung mengenai mutu layanan dan sarana yang disediakan.



Gambar 6. Alur Proses Analisis Menggunakan *Orange Data Mining*

Analisis sentimen dalam penelitian ini melibatkan beberapa tahapan metodologis yang dimulai dengan impor 1000 data ulasan pengguna *Google Maps* dalam format CSV ke perangkat lunak *Orange*. Untuk keperluan pemrosesan lanjutan, data ini ditransformasi menjadi sebuah korpus. Tahapan awal penanganan teks meliputi normalisasi huruf, pemecahan teks menjadi unit-unit linguistik, pembuangan kata-kata fungsional, dan pembersihan karakter asing demi kemurnian dan standarisasi data. Selanjutnya, analisis sentimen dilakukan menggunakan metode leksikon agar dapat menggolongkan setiap tinjauan sebagai positif, negatif, atau netral. Kolom teks ulasan dan label sentimen dipilih untuk digunakan dalam pelatihan model machine learning. Pelatihan model *Naive Bayes* didasarkan pada data yang telah melalui tahap pembersihan. Pengujian yang memanfaatkan validasi silang diterapkan untuk mengevaluasi kinerja model, yang menghasilkan metrik seperti akurasi, presisi, recall, dan F1-score. Analisis lebih lanjut diperdalam melalui visualisasi menggunakan confusion matrix. Analisis tambahan mencakup distribusi sentimen, *word cloud* kata-kata dominan, dan pemodelan topik.

4.3 Distribusi Data Sentimen



Gambar 7. Alur Proses Analisis Menggunakan *Orange Data Mining*

Pengolahan data melalui *Orange Data Mining* menghasilkan visualisasi yang menampilkan distribusi rating bersama dengan hasil klasifikasi sentimen dari ulasan pengguna. Tujuan visualisasi ini adalah untuk menggambarkan korelasi antara skala penilaian (rating) dan persepsi pengguna yang terbagi menjadi sentimen positif, netral, dan negatif. Analisis menunjukkan bahwa pada rating 1 dan 2, sentimen negatif mendominasi ulasan, mengindikasikan bahwa pengguna yang memberikan skor rendah umumnya memiliki pengalaman yang kurang memuaskan dengan layanan yang diterima. Fenomena ini konsisten dengan temuan riset yang menyebutkan sentimen negatif dalam ulasan digital mencerminkan adanya kelemahan dalam aspek layanan yang dirasakan langsung oleh konsumen.

Selanjutnya, rating 3 didominasi oleh distribusi sentimen netral, yang menandakan pengguna mengalami pengalaman yang moderat; layanan yang diterima tidak sepenuhnya memenuhi harapan namun juga tidak

menimbulkan ketidakpuasan berarti. Dalam studi sebelumnya, sentimen netral kerap dianggap sebagai indikator kualitas layanan yang masih standar dan memerlukan perbaikan. Sementara itu, rating 4 dan 5 didominasi oleh sentimen positif. Rating 4 menunjukkan frekuensi tertinggi, menandakan sebagian besar pengguna merasa puas dengan layanan yang ditawarkan. Rating 5 merepresentasikan tingkat kepuasan tertinggi, di mana seluruh ulasan bersifat positif. Hasil ini sejalan dengan penelitian yang menemukan mayoritas ulasan pengunjung cenderung positif, meski terdapat keseimbangan sentimen antara positif, netral, dan negatif.

Selain menggambarkan sentimen umum pengunjung, penelitian ini juga menemukan beberapa aspek layanan yang paling sering dibahas dalam ulasan pengguna *Google Maps*. Aspek layanan tersebut meliputi kebersihan area wisata, fasilitas umum, harga tiket, pelayanan petugas, aksesibilitas lokasi, kenyamanan lingkungan, dan kepadatan pengunjung. Mayoritas ulasan positif berkaitan dengan keindahan bangunan candi, pengalaman wisata budaya, dan fasilitas pendukung yang memadai. Sementara itu, ulasan negatif lebih banyak berkaitan dengan antrean panjang, tingginya harga tiket, dan kondisi area yang padat saat musim liburan.

4.4 Evaluasi Model

Penilaian terhadap kinerja algoritma *Naive Bayes* dalam pengelompokan sentimen dari ulasan pengguna telah dilaksanakan. Penggunaan *Confusion matrix* sebagai metode penilaian membantu menyajikan data kuantitatif mengenai akurasi prediksi pada setiap kategori sentimen yang teridentifikasi.

4.4.1 Hasil *Confusion matrix*

		Predicted			Σ
		negative	neutral	positive	
Actual	negative	340	0	0	340
	neutral	0	262	68	330
	positive	0	38	292	330
Σ		340	300	360	1000

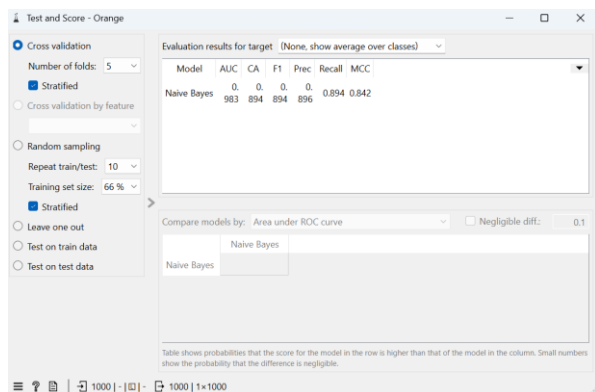
Gambar 8. *Confusion Matrix* pada *Orange*

Data yang tersaji dalam Gambar 8 merupakan produk dari pengumpulan dan kajian terhadap 1000 ulasan yang telah dikumpulkan. Berdasarkan pengujian menggunakan model *Naive Bayes* untuk klasifikasi sentimen menunjukkan performa yang sangat memuaskan dalam mengidentifikasi sentimen ulasan pengunjung. Dari jumlah 1000 data yang dievaluasi, seluruh 340 data berkarakteristik negatif berhasil diklasifikasikan tanpa satupun kesalahan. Untuk kategori netral, 262 data berhasil diklasifikasikan dengan benar,

namun 68 data lainnya keliru ditempatkan pada kategori positif. Kelas positif pun sebagian besar berhasil diidentifikasi secara akurat, meskipun sejumlah kecil data mengalami kesalahan klasifikasi. Hasil confusion matrix menunjukkan bahwa model *Naive Bayes* memiliki performa yang baik dalam mengklasifikasikan sentimen ulasan pengunjung. Dari total 1000 data ulasan, model mampu mengidentifikasi sebagian besar data secara tepat pada masing-masing kelas sentimen. Meskipun demikian, masih ditemukan beberapa kesalahan klasifikasi terutama pada kelas netral dan positif yang memiliki kemiripan pola linguistik dan konteks semantik. Namun demikian, model masih menemui hambatan dalam membedakan secara presisi antara sentimen netral dan positif, yang diduga karena kesamaan fitur linguistik dan kedekatan makna semantik antara kedua kelas tersebut.

Secara menyeluruh, algoritma *Naive Bayes* terbukti memiliki tingkat kepercayaan yang tinggi untuk tugas klasifikasi sentimen, meskipun upaya pengembangan lebih lanjut tetap esensial untuk meningkatkan performa pada sentimen yang memiliki ambiguitas.

4.4.2 Test and Score



Gambar 9. Test and Score Matrix pada Orange

Melalui penerapan metode 5-fold cross validation pada Orange Data Mining, algoritma *Naive Bayes* menunjukkan performa klasifikasi yang baik dalam menganalisis sentimen ulasan pengunjung Candi Borobudur. Model memperoleh nilai accuracy sebesar 0,894, precision sebesar 0,896, recall sebesar 0,894, F1-score sebesar 0,894, dan AUC sebesar 0,983. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang efektif dalam mengidentifikasi pola sentimen pada data ulasan digital.

Tabel 3. Penpaparan Hasil Analisis

Metrik Evaluasi	Nilai
Accuracy	0,894
Precision	0,896
Recall	0,894

F1-Score	0,894
AUC	0,983
MCC	0,842

5. KESIMPULAN

Penelitian ini berhasil menerapkan analisis sentimen terhadap 1000 ulasan Google Maps Candi Borobudur menggunakan algoritma *Naive Bayes* untuk mengevaluasi persepsi pengunjung terhadap layanan dan fasilitas wisata. Hasil klasifikasi menunjukkan bahwa mayoritas ulasan memiliki sentimen positif, terutama pada aspek keindahan wisata, fasilitas, dan pengalaman pengunjung. Namun demikian, masih ditemukan sentimen negatif terkait kepadatan pengunjung, antrean, dan harga tiket. Berdasarkan hasil evaluasi model menggunakan 5-fold cross validation, algoritma *Naive Bayes* menunjukkan performa yang baik dengan accuracy sebesar 0,894 dan AUC sebesar 0,983 sehingga efektif digunakan dalam klasifikasi sentimen ulasan wisata digital.

Penelitian ini memiliki beberapa keterbatasan, di antaranya jumlah dataset yang masih terbatas pada 1000 ulasan serta penggunaan satu algoritma klasifikasi tanpa perbandingan dengan metode machine learning lainnya. Selain itu, proses pelabelan sentimen masih menggunakan pendekatan lexicon otomatis sehingga memungkinkan adanya ambiguitas pada beberapa data ulasan.

Untuk penelitian selanjutnya, disarankan menggunakan jumlah dataset yang lebih besar, menerapkan perbandingan beberapa algoritma klasifikasi seperti Support Vector Machine atau Deep Learning, serta melakukan analisis aspek layanan secara lebih mendalam agar hasil evaluasi wisata menjadi lebih komprehensif dan akurat.

DAFTAR PUSTAKA

- [1] R. Hashim, B. Omar, N. Saeed, A. Ba-Anqud, and H. Al-Samarrhaie, "THE APPLICATION OF SENTIMENT ANALYSIS IN TOURISM RESEARCH: A BRIEF REVIEW."
- [2] Ni Wayan Anggreni, I Made Artayasa, and I Wayan Sukita, "Identifikasi Ulasan Wisatawan Daring untuk Meningkatkan Layanan Destinasi Wisata," *Jurnal Kajian dan Terapan*
- [3] Y. Lin, W. Yu, H. Li, Y. Zhang, W. Yang, and Q. Chen, "Satisfaction Analysis and Optimization of the Management of the Forbidden City in Beijing Based on Online Reviews," *Advances in Engineering Research Possibilities and Challenges*, p. 17, Mar. 2025, doi: 10.63313/aerpc.9002.
- [4] W. Khofifah, D. N. Rahayu, and A. M. Yusuf, "Analisis Sentimen Menggunakan *Naive Bayes*

- Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps,” *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 16, no. 4, pp. 28–38, Jan. 2022, doi: 10.35969/interkom.v16i4.192.
- [5] A. Ameer, S. Hamdi, and S. Ben Yahia, “Sentiment Analysis for Hotel Reviews: A Systematic Literature Review,” *ACM Comput. Surv.*, vol. 56, no. 2, Feb. 2024, doi: 10.1145/3605152.
- [6] D. T. Hermanto, M. Ziaurrahman, M. A. Bianto, and A. Setyanto, “Twitter Social Media Sentiment Analysis in Tourist Destinations Using Algorithms *Naive Bayes* Classifier,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Dec. 2018. doi: 10.1088/1742-6596/1140/1/012037.
- [7] Y. A. Setyawan, F. E. Nastiti, and A. A. Sari, “Analisis Sentimen Ulasan Tempat Wisata Umbul Sigidang Pada Google Maps Menggunakan Algoritma *Naive Bayes* Dan Support Vector Machine Sentiment Analysis of Umbul Sigidang Tourist Attraction Reviews on Google Maps Using *Naive Bayes* Algorithm and Support Vector Machine,” *Jurnalnya Orang Pintar Komputer*, vol. 14, no. 2, 2025, doi: 10.30591/smartcomp.v13i1.7467.
- [8] R. Blloshmi, T. Pasini, N. Campolungo, S. Banerjee, R. Navigli, and G. Pasi, “IR like a SIR Sense-enhanced Information Retrieval for Multiple Languages.” [Online]. Available: <https://github.com/SapienzaNL/P/sir>.
- [9] Q. Ai *et al.*, “Information Retrieval Meets Large Language Models: A Strategic Report from Chinese IR Community,” Jul. 2023, [Online]. Available: <http://arxiv.org/abs/2307.09751>
- [10] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode *Naive Bayes* Classifier,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.
- [11] Felicia Watratan, A. B. Puspita, D. Moeis, S. Informasi, and S. Profesional Makassar, “Implementasi Algoritma *Naive Bayes* Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia,” 2020. [Online]. Available: <http://journal.isas.or.id/index.php/JACOST>
- [12] J. Penerapan Teknologi Informasi dan Komunikasi, F. Bryan Prasetio, and T. Wellem, “IT-EXPLORE”.
- [13] M. Rifan, H. Inaya Fikriya, P. I. Lembaga Konsultasi dan Bantuan Hukum Rumah Keadilan Jalan Kembang Kertas No, K. Lowokwaru, K. Malang, and J. Timur, “DISHARMONISASI PERIODE PEMBERLAKUAN RENCANA INDUK PEMBANGUNAN KEPARIWISATAAN DAERAH DI INDONESIA (IMPLEMENTASI PERATURAN PEMERINTAH NOMOR 50 TAHUN 2011 TENTANG RENCANA INDUK PEMBANGUNAN KEPARIWISATAAN) Disharmonized the Impelementation Period of Regional Tourism Development Plan in Indonesia (The Implementation of Government Regulation Number 50 Year 2011 Concerning Tourism Development Master Plan),” 2020.
- [14] T. F. Hidayatullah and Y. Sutarso, “Peran Media Sosial, Ulasan Daring, dan Kepedulian Lingkungan pada Perilaku Pembelian Green Product,” *At-Tadbir : jurnal ilmiah manajemen*, vol. 7, no. 1, p. 23, Feb. 2023, doi: 10.31602/atd.v7i1.9217.
- [15] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma *Naive Bayes* dan KNN,” *Jurnal KomtekInfo*, 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [16] Y. A. Singgalen, “Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma *Naive Bayes* Classifier,” *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2486.
- [17] D. Darwis, N. Siskawati, and Z. Abidin, “PENERAPAN ALGORITMA *NAIVE BAYES* UNTUK ANALISIS SENTIMEN REVIEW DATA TWITTER BMKG NASIONAL,” *Jurnal Tekno Kompak*, vol. 15, no. 1, p. 131, Feb. 2021, doi: 10.33365/jtk.v15i1.744.
- [18] Y. Alvin Rivera, S. Mariati, and F. Asmaniaty Institut Pariwisata Trisakti, “Identifikasi Potensi Pariwisata di Desa Kawasan Borobudur Kabupaten Magelang Provinsi Jawa Tengah; Studi Kasus Desa Candirejo,” 2025.
- [19] A. I. Tanggraeni and M. N. N. Sitokdana, “Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma *Naive Bayes*,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 2, 2022, doi: 10.35957/jatisi.v9i2.1835.
- [20] Y. Mao, Q. Liu, and Y. Zhang, “Sentiment analysis methods, applications, and challenges: A systematic literature review,” Apr. 01, 2024, *King Saud bin Abdulaziz University*. doi:

- 10.1016/j.jksuci.2024.102048.
- [21] Alfandi Safira and F. N. Hasan, "ANALISIS SENTIMEN MASYARAKAT TERHADAP PAYLATER MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER*," *ZONAsi: Jurnal Sistem Informasi*, vol. 5, no. 1, 2023, doi: 10.31849/zn.v5i1.12856.