

PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN PENGUNJUNG DI PANTAI KEJAWANAN

Rizal Dwi Rizki Apriliansyah^{1*}, Rini Astuti², Willy Prihartono³, Ryan Hamonangan⁴

^{1,3,4}STMIK IKMI Cirebon; Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135; Telp. (0231) 490480

²STMIK LIKMI Bandung; Jl. Ir. H. Juanda No.96, Lebakgede, Kecamatan Coblong, Kota Bandung, Jawa Barat 40132; Telp. (022) 2502121

Received: 18 Desember 2024

Accepted: 14 Januari 2025

Published: 20 Januari 2025

Keywords:

naïve bayes; objek wisata, data mining; analisis sentimen; pengunjung.

Corespondent Email:

rizaldwirizki194@gmail.com

Abstrak. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengunjung terhadap objek wisata Pantai Kejawanan dengan menggunakan algoritma Naïve Bayes. Sentimen diklasifikasikan menjadi dua kategori, yaitu positif dan negatif. Ulasan daring merupakan sumber data penting untuk memahami persepsi pengunjung terhadap destinasi wisata. Algoritma Naïve Bayes dipilih karena keunggulannya dalam menganalisis teks. Dataset terdiri dari 998 ulasan yang dikumpulkan melalui scraping menggunakan SerApi, dengan periode data dari 2020 hingga 2024. Analisis dilakukan melalui tahapan *Knowledge Discovery in Databases* (KDD), mencakup pemilihan data, pre-processing (pembersihan data, case folding, tokenisasi, penghapusan stopword, dan stemming), serta transformasi menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Dataset dibagi menjadi 70% data pelatihan dan 30% data pengujian sebelum penerapan algoritma. Hasil menunjukkan bahwa model mampu mencapai akurasi 78%, presisi 92%, recall 80%, dan F1-score 86%, yang mencerminkan performa baik dalam klasifikasi sentimen. Evaluasi menggunakan confusion matrix menunjukkan model konsisten dalam membedakan sentimen positif dan negatif. Analisis ini menunjukkan bahwa mayoritas ulasan pengunjung Pantai Kejawanan bersentimen positif, menggambarkan pengalaman yang memuaskan.

Abstract. This research aims to analyze the sentiment of visitor reviews of the Kejawanan Beach tourist attraction using the Naïve Bayes algorithm. Sentiment is classified into two categories, namely positive and negative. Online reviews are an important data source for understanding visitors' perceptions of tourist destinations. The Naïve Bayes algorithm was chosen because of its superiority in analyzing text. The dataset consists of 998 reviews collected through scraping using SerApi, with a data period from 2020 to 2024. Analysis was carried out through the Knowledge Discovery in Databases (KDD) stages, including data selection, pre-processing (data cleaning, case folding, tokenization, stopword removal , and stemming), as well as transformation using the Term Frequency-Inverse Document Frequency (TF-IDF) method. The dataset is divided into 70% training data and 30% testing data before applying the algorithm. The results show that the model is able to achieve 78% accuracy, 92% precision, 80% recall, and 86% F1-score, which reflects good performance in sentiment classification. Evaluation using the confusion matrix shows that the model is consistent in distinguishing positive and negative sentiment. This analysis shows that the majority of reviews from visitors to Kejawanan Beach are positive, describing a satisfactory experience.

1. PENDAHULUAN

Dalam beberapa tahun terakhir, Objek Wisata telah menjadi salah satu sektor penting dalam mendukung perkembangan ekonomi di berbagai daerah dengan potensi yang memiliki sumber daya alam yang melimpah [1]. Media sosial kini berperan signifikan sebagai platform penyebaran informasi tentang destinasi wisata, termasuk Kabupaten Cirebon, yang memiliki Pantai Kejawanan sebagai salah satu destinasi wisata unggul. Kehadiran media sosial memungkinkan wisatawan membagikan pengalaman dan opini mereka, sehingga memberikan peluang bagi pengelola destinasi untuk memahami persepsi wisatawan serta meningkatkan kualitas layanan. Namun demikian, pengelolaan data ulasan wisatawan yang ada di platfrom *Google Maps* dalam skala besar tetap menjadi tantangan signifikan karena sifatnya yang subjektif dan sangat beragam.

Dalam analisis sentimen, algoritma naïve bayes telah terbukti untuk mengidentifikasi dan mengklasifikasi ulasan pengunjung pada objek wisata pantai kejawanan. Penelitian terdahulu menunjukkan bahwa algoritma naïve bayes terbukti efektif dalam analisis sentimen, dengan tingkat akurasi 92% yang signifikan, sebagaimana ditunjukkan dalam studi terkait data pariwisata di Lombok[2]. Studi ini juga menunjukkan bahwa algoritma Naïve Bayes dapat dimanfaatkan sebagai pendekatan yang efisien dalam mengklasifikasikan sentimen positif dan negatif pada ulasan wisatawan terhadap candi borobudur. Namun, keterbatasan terkait variasi data sering kali menjadi tantangan dalam penerapannya[3]. Secara umum, Penelitian sebelumnya mengungkapkan bahwa algoritma ini menunjukkan kinerja yang baik dalam menangani data opini yang kompleks di sektor destinasi wisata.

Namun, penelitian sebelumnya memiliki keterbatasan yang perlu diperhatikan. Sebagian besar studi lebih berfokus pada destinasi berskala nasional atau internasional, sementara analisis sentimen pada destinasi lokal seperti Pantai Kejawanan masih belum banyak di teliti oleh literatur sebelumnya. dalam analisis terhadap destinasi wisata di Bali, menunjukkan potensi besar untuk memperluas penggunaan algoritma ini pada destinasi lokal lainnya[1]. Selain itu, tantangan terkait variasi data seperti bahasa dan format ulasan masih menjadi kendala dalam pengolahan data sentimen[4].

Walaupun algoritma Naïve Bayes telah menunjukkan efektivitasnya dalam analisis sentimen pariwisata, penelitian yang berfokus pada destinasi lokal seperti Pantai Kejawanan masih sangat terbatas. Mayoritas studi yang ada cenderung mengkaji destinasi berskala besar dengan volume data yang luas, sehingga pengelolaan data opini wisatawan lokal sering kali terabaikan. Selain itu, literatur yang menjelaskan bagaimana algoritma ini dapat diterapkan untuk mendukung pengambilan keputusan berbasis data oleh pengelola destinasi wisata lokai masih relatif minim. Penelitian ini bertujuan untuk menjawab kesenjangan tersebut dengan mengeksplorasi algoritma Naïve Bayes dalam menganalisis sentimen objek wisata terhadap Pantai Kejawanan.

Penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasikan sentimen pengunjung terhadap Pantai Kejawanan dengan menggunakan algoritma Naive Bayes. Proses ini mencakup analisis ulasan pengunjung untuk menentukan apakah sentimen tersebut bersifat positif, atau negatif. Selain itu, penelitian ini juga bertujuan untuk mengukur tingkat akurasi algoritma Naive Bayes dalam menganalisis sentimen tersebut, sehingga dapat diketahui sejauh mana algoritma ini efektif dalam memberikan hasil klasifikasi yang akurat terhadap data sentimen pengunjung Pantai Kejawanan.

Penelitian ini memanfaatkan pendekatan kuantitatif dengan metode *Knowledge Discovery in Databases* (KDD) untuk mendukung strategi pengelolaan destinasi wisata, khususnya di Pantai Kejawanan. Tujuan utama dari penelitian ini adalah memberikan wawasan yang lebih mendalam terkait persepsi objek wisata melalui analisis sentimen, sehingga dapat digunakan untuk meningkatkan kualitas layanan yang disediakan. Selain itu, hasil penelitian ini diharapkan dapat menjadi rujukan dalam pengembangan sistem informasi destinasi berbasis data dan mendukung proses pengambilan keputusan strategis oleh pemangku kepentingan di sektor Objek Wisata.

2. TINJAUAN PUSTAKA

Analisis sentimen telah menjadi salah satu bidang penelitian yang berkembang pesat, terutama dalam memahami persepsi ulasan pengunjung terhadap destinasi wisata.

Penelitian ini berfokus pada penggunaan algoritma Naive Bayes dalam menganalisis sentimen pengunjung terhadap objek wisata Pantai Kejawan.

2.1. Analisis Sentimen

Analisis sentimen merupakan proses untuk mengidentifikasi arah polaritas dalam sebuah teks, baik dalam dokumen maupun bentuk lain, dengan tujuan mengkategorikan sentimen menjadi positif, dan negatif[5]. Penelitian ini menunjukkan bahwa analisis sentimen terhadap persepsi pelanggan terkait produk dan layanan restoran di Labuan Bajo [6] menggunakan pendekatan text mining telah membuktikan manfaatnya dalam meningkatkan kualitas layanan. Selain itu, penelitian ini juga menganalisis ulasan pengunjung Pulau Komodo dan Pulau Rinca dengan menggunakan kerangka kerja *CRISP-DM* [7], menyoroti pentingnya pengelolaan data yang terorganisasi untuk menghasilkan wawasan yang bernalih.

Sementara itu, penelitian ini juga menggunakan algoritma K-Nearest Neighbor (K-NN) dalam mengkaji sentimen pengunjung terhadap objek wisata Dunia Fantasi[8]. Hasilnya menunjukkan bahwa algoritma ini mampu melakukan klasifikasi analisis sentimen secara efektif, meskipun menghadapi tantangan dalam pengolahan dataset yang besar.

2.2. Algoritma Naïve Bayes

Algoritma Naive Bayes adalah metode pembelajaran mesin berbasis probabilitas yang sering digunakan untuk klasifikasi sentimen. Penelitian ini membandingkan performa algoritma *Naïve Bayes Classifier* (NBC) dengan *Decision Tree* (DT) dan *Support Vector Machine* (SVM) dalam klasifikasi ulasan pengunjung Candi Borobudur[9]. Hasil penelitian menunjukkan bahwa Naive Bayes memiliki keunggulan dalam efisiensi serta memberikan akurasi yang baik, terutama untuk dataset berukuran kecil hingga menengah.

Selain itu, Penelitian lain juga mengungkapkan keberhasilan penerapan Naive Bayes dalam klasifikasi ulasan pengunjung Danau Toba dengan menggunakan pendekatan *CRISP-DM*[10]. Pendekatan ini menunjukkan hasil yang relevan dan efektif dalam proses klasifikasi data ulasan.

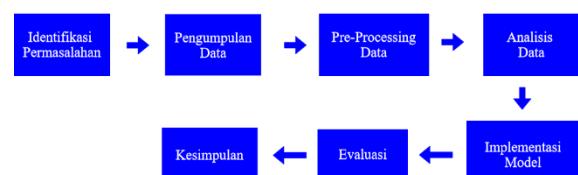
Sementara itu, analisis sentimen yang membandingkan Naive Bayes dan SVM pada dataset kurikulum pendidikan berbasis e-sport

menggaris bawahi pentingnya pemilihan algoritma yang sesuai dengan karakteristik dataset untuk menghasilkan performa yang optimal [11].

Penelitian ini menerapkan algoritma Naive Bayes untuk mengkategorikan sentimen yang terkandung dalam ulasan pengunjung Pantai Kejawan. Berdasarkan hasil studi terdahulu, algoritma ini dikenal karena kesederhanaannya sekaligus kemampuannya dalam memberikan hasil klasifikasi yang akurat.

3. METODE PENELITIAN

Metode penelitian yang diterapkan dalam penelitian ini menggunakan pendekatan kuantitatif. Pendekatan kuantitatif mengacu pada metode pengumpulan dan analisis data yang berbasis pada angka dan statistik. Dalam penelitian berjudul Analisis Sentimen Pengunjung Terhadap Objek Wisata Pantai Kejawan Menggunakan Algoritma Naive Bayes, data diolah menggunakan algoritma Naive Bayes. Algoritma ini berfungsi untuk mengklasifikasikan sentimen pengunjung menjadi kategori positif dan negatif berdasarkan dataset yang diperoleh dari ulasan pengunjung Pantai Kejawan yang diambil melalui platform *Google Maps*. Adapun tahapan penelitian dapat dilihat pada gambar 3.1.



Gambar 3. 1. Tahapan Penelitian

3.1. Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari ulasan pengunjung objek wisata Pantai Kejawan yang tersedia di Google Maps. Data tersebut mencakup beberapa variabel, seperti username, ulasan, dan rating. Fokus utama dari penelitian ini adalah pada variabel ulasan dan rating, yang diharapkan dapat memberikan wawasan signifikan mengenai sentimen pengunjung terhadap objek wisata Pantai Kejawan. Data yang telah dikumpulkan dalam bentuk dataset

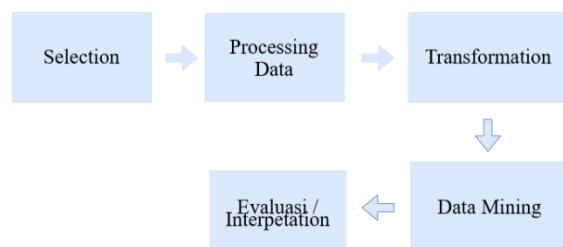
berformat .xlsx guna memfasilitasi proses analisis lebih lanjut.

3.2. Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan melalui metode web scraping. Metode ini merupakan proses pengambilan data semi-terstruktur dari situs web secara otomatis menggunakan program komputer. Dalam penelitian ini, data yang dikumpulkan berupa ulasan pengunjung yang diperoleh melalui Google Maps. Proses pengumpulan dimulai dengan menyusun kode program menggunakan bahasa pemrograman Python yang terintegrasi dengan layanan SrapApi. Layanan SrapApi digunakan untuk mengakses dan mengunduh data ulasan berdasarkan parameter tertentu, seperti nama pengguna, ulasan, rating, dan tanggal ulasan. Data yang diperoleh dari SrapApi dikembalikan dalam format JSON, yang kemudian diolah lebih lanjut. Tahapan pembersihan data dilakukan menggunakan pustaka Python, seperti pandas, untuk menghapus informasi yang tidak relevan, menghilangkan data duplikat atau kosong, serta menyusun data dalam format yang lebih terstruktur. Data yang telah melalui proses pembersihan ini kemudian disimpan dalam file Excel untuk keperluan analisis lebih lanjut.

3.3. Teknik Analisis Data

Penelitian ini menerapkan algoritma Naive Bayes untuk menganalisis sentimen pengunjung terhadap objek wisata Pantai Kejawanan. Proses analisis data mengikuti tahapan Knowledge Discovery in Databases (KDD), yang terdiri atas lima langkah utama: selection (seleksi), pre-processing (prapemrosesan), transformation (transformasi), data mining, dan interpretation/evaluation (interpretasi / evaluasi). Proses tahapan KDD ini ditunjukkan pada gambar.



Gambar 3. 2. Proses KDD

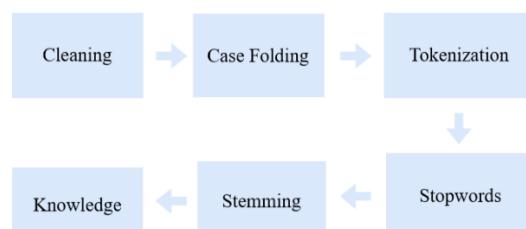
Berdasarkan gambar 3.2 Proses Tahapan KDD tersebut dapat dijelaskan setiap tahapan sebagai berikut :

3.3.1. Selection

Pada tahap ini, data dikumpulkan untuk analisis sentimen dari ulasan pengunjung objek wisata yang ada di Google Maps. Dataset yang diambil meliputi Username, Ulasan, dan Rating yang diberikan oleh pengunjung terhadap objek wisata Pantai Kejawanan dari Tahun 2020 - 2024.

3.3.2. Processing Data

Processing merupakan proses pengolahan data sebelum data tersebut masuk ke tahap pemrosesan lebih lanjut. Pada tahap ini, masih banyak dataset yang tidak bersih, seperti adanya kesalahan sistem saat pencatatan data yang menyebabkan duplikasi. Tahap Preprocessing ini melibatkan beberapa proses untuk mempersiapkan data agar siap dianalisis, yaitu :



Gambar 3. 3. Processing Data

3.3.2.1. Cleaning adalah tahap untuk membersihkan data dalam dataset dan menyaring kata-kata yang tidak relevan, tidak memiliki makna, atau mengandung makna yang dapat mempengaruhi sentimen.

3.3.2.2. Case Folding adalah proses mengubah semua huruf dalam teks menjadi huruf kecil untuk menjaga menyamakan format dan mengurangi perbedaan yang disebabkan oleh huruf kapital, sehingga memudahkan proses analisis.

3.3.2.3. Tokenisasi adalah proses memecah data set menjadi token atau potongan kata untuk memudahkan tahap selanjutnya.

3.3.2.4. Stopword adalah tahap menghilangkan kata-kata yang sering muncul namun tidak memiliki arti penting dan tidak mempengaruhi sistem [12].

3.3.2.5. Stemming adalah proses menghapus imbuhan dari sebuah kata untuk mengubahnya menjadi bentuk dasar.

3.3.2.6. Labeling adalah proses menambahkan tanda atau label pada data untuk mengklasifikasikan atau mengidentikasinya berdasarkan kategori tertentu, seperti sentimen "Positif" dan "Negatif".

3.3.3. Transformation

Data ulasan yang sudah diproses diubah menjadi format numerik dengan menggunakan teknik *TF-IDF*. **TF-IDF** adalah teknik yang digunakan untuk mengukur pentingnya suatu kata dalam dokumen berdasarkan frekuensinya (TF) dan seberapa langka kata tersebut di seluruh dokumen (IDF). Representasi ini memungkinkan kata-kata dalam ulasan dapat dihitung frekuensinya dan dinilai sejauh mana pengaruhnya dalam menentukan sentimen pengunjung.

3.3.4. Data Mining

Dalam penelitian ini, algoritma Naive Bayes diterapkan untuk mengklasifikasikan sentimen pengunjung terkait objek Pantai Kejawanan. Naive Bayes merupakan algoritma yang sederhana namun sangat efektif dalam klasifikasi teks, termasuk analisis sentimen.

3.3.5. Evaluasi / Interpretation

Evaluasi, model akan diukur berdasarkan metriks evaluasi diantaranya akurasi (*Accuracy*), *precision*, *recall*, *f1-score* dan *confusion matrix* untuk mengetahui seberapa baik model dalam mengklasifikasikan analisis sentimen.

4. HASIL DAN PEMBAHASAN

4.1. Selection

Tabel 4.1 menunjukkan Hasil scraping, dataset tersebut terdiri beberapa variabel, seperti *username*, *ulasan*, *rating*, dan *date*. Namun, dataset ini ada 2 variabel utama yaitu Ulasan dan Rating digunakan dalam analisis sentimen. Oleh karena itu, diperlukan tahapan processing data agar dapat digunakan dalam proses analisis mengenai sentimen pengunjung terhadap objek wisata Pantai Kejawanan.

Tabel 4. 1. Hasil Scraping

Ulasan	Rating
Pantainya bersih, terkelola dengan baik Cuma karena musim panas air lautnya agak menyusut Masuk pertama bayar 2.000 parkir motor 2.000 tiket masuk objek wisata 5.500/org	5

Salah satu pantai utara Jawa terbaik yang pernah saya kunjungi, mulai dari penataan, kebersihan, fasilitas umum seperti tempat duduk, kios makan, semuanya tersusun dengan baik dan cantik. Total biaya yang diperlukan adalah masuk kawasan 2k, parkir motor 2k, dan masuk ke pantainya 5.5k per orang. Terbilang bersih dan banyak satuan pengaman yang memantau dan aktif berkeliling. Untuk yang sedang galau juga cocok karena tersedia ayunan tepi pantai serta karaokean dangdut dengan voltase ampli yang yahud, membuat jiwa sedih menjadi ingin berdendang.	5
Sebagai orang asli Cirebon pasti gak asing lagi sama tempat ini,tapi pas kesini baru"" ini loh kok glow up ya,serasa ke tempat yang berbeda,udah nyaman, bersih, tersedia tempat duduk santai, tempat bermain anak,dll Mungkin yang masih kurang nyaman kamar mandi nya kali ya,,hhe KLO pengen sepi Dateng ny Jum'at ajah jangan Minggu rame , Oh satu lagi, tiketnya juga GK mahal kok sangat Murmer,di bawah 10k kok perorang nya	5

4.2. Processing Data

Tahapan ini dimulai dengan mengambil data mentah yang dihasilkan dari proses seleksi awal, kemudian dilanjutkan dengan serangkaian langkah pre-processing yang meliputi cleaning, case folding, stopword removal, tokenisasi, stemming, dan pemberian label (labeling).

4.2.1. Cleaning Data

Pada tabel 4.2 menampilkan hasil claning yang telah dibersih kan menghasilkan data yang lebih terstruktur dan siap digunakan untuk analisis lebih lanjut.

Tabel 4. 2. Hasil Cleaning Data

Ulasan	Cleaning
Pantainya bersih, terkelola dengan baik Cuma karena musim panas air lautnya agak menyusut	Pantainya bersih, terkelola dengan baik Cuma karena musim panas air lautnya agak menyusut

musim panas air lautnya agak menyusut Masuk pertama bayar 2.000 parkir motor 2.000 tiket masuk objek wisata 5.500/org	panas air lautnya agak menyusut Masuk pertama bayar parkir motor tiket masuk objek wisata org	masuknya Rp.5.000/org + asuransi Rp.500 jadi totalnya Rp.5.500/org. Lingkungannya bersih dan juga ada tersedia Pos Keamanan kalau perlu bantuan karena di pantainya suka ada ubur Jadi bagi orangtua yang main dipantai dengan anaknya agar lebih memperhatikan kondisi lingkungannya.
Salah satu pantai utara Jawa terbaik yang pernah saya kunjungi, mulai dari penataan, kebersihan, fasilitas umum seperti tempat duduk, kios makan, semuanya tersusun dengan baik dan cantik. Total biaya yang diperlukan adalah masuk kawasan 2k, parkir motor 2k, dan masuk ke pantainya 5.5k per orang. Terbilang bersih dan banyak satuan pengaman yang memantau dan aktif berkeliling. Untuk yang sedang galau juga cocok karena tersedia ayunan tepi pantai serta karaokean dangdut dengan voltase ampli yang yahud, membuat jiwa sedih menjadi ingin berdendang.	Sebagai orang asli Cirebon pasti gak asing lagi sama tempat initapi pas kesini baru ini loh kok glow up yaserasa ke tempat yang berbedaudah nyaman bersih tersedia tempat duduk santai tempat bermain anakdll Mungkin yang masih kurang nyaman kamar mandi nya kali yahhe KLO pengen sepi Dateng ny Jumat ajah jangan Minggu rame Oh satu lagi tiketnya juga GK mahal kok sangat Murmerdi bawah k kok perorang nya	Berdasarkan Hasil Tabel 4.2 menujukan perbedaan yang jelas antara Ulasan dan Cleaning. Pada kolom Ulasan, teks masih mengandung elemen-elemen yang tidak relevan, seperti angka ("2.000", "5.500", "10"), simbol non-alfabet (seperti tanda baca), spasi berlebih, serta format informal. Teks ini juga belum terstruktur dengan baik, sehingga kurang ideal untuk dianalisis. Setelah dilakukan proses cleaning, hasilnya ditampilkan dalam kolom Cleaning, di mana angka, simbol non-alfabet, dan spasi berlebih telah dihapus. Teks menjadi lebih terstruktur dan bersih, dengan elemen-elemen teknis yang mengganggu telah dihilangkan. Meski demikian, istilah informal seperti "Murmer", "hhe", dan "GK" tetap dipertahankan karena proses cleaning hanya berfokus pada aspek teknis. Hasilnya, teks pada kolom Ulasan Sesudah Cleaning lebih sederhana, rapi, dan siap untuk dianalisis lebih lanjut yaitu tahapan Case Folding.
Tempat liburan dengan anak yang ramah dikantong. Untuk tiket masuk kendaraan roda 4 Rp.6.000 dan tarif parkirnya Rp.4.000. Saat keluar parkir ada bapak parkir yang bantuin tetap saya kasih seikhlasnya aja Gak apaapa ya Pak Untuk tiket masuknya Rporg asuransi Rp	Tempat liburan dengan anak yang ramah dikantong Untuk tiket masuk kendaraan roda Rp dan tarif parkirnya Rp Saat keluar parkir ada bapak parkir yang bantuin tetap saya kasih seikhlasnya aja Gak apaapa ya Pak Untuk tiket masuknya Rporg asuransi Rp	4.2.2. Case Folding Pada tabel 4.3 menampikan hasil case folding yang telah mengubah semua huruf dalam teks menjadi huruf kecil.

Tabel 4. 3. Case Folding

Claning	Case Folding
Pantainya bersih terkelola dengan baik Cuma karena musim panas air lautnya agak menyusut Masuk	pantainya bersih terkelola dengan baik cuma karena musim panas air lautnya agak menyusut masuk pertama bayar

pertama bayar parkir motor tiket masuk objek wisata org	parkir motor tiket masuk objek wisata org
---	---

Berdasarkan Tabel 4.10 Hasil proses case folding , pada kolom 'Cleaning' menunjukkan bahwa seluruh teks berhasil diubah menjadi huruf kecil pada kolom baru yang dinamakan 'CaseFolding'. Setelah selesai tahapan tersebut hasil ini dianalisis lebih lanjut ke tahapan yaitu Tokenization.

4.2.3. Tokenization

Pada tabel 4.4 dapat dilihat menampilkan hasil tokenization yang telah memecah teks menjadi token atau potongan kata.

Tabel 4. 4. Tokenization

Case Folding	Tokenization
pantainya bersih terkelola dengan baik cuma karena musim panas air lautnya agak menyusut masuk pertama bayar parkir motor tiket masuk objek wisata org	'pantainya', 'bersih', 'terkelola', 'dengan', 'baik', 'cuma', 'karena', 'musim', 'panas', 'air', 'lautnya', 'agak', 'menyusut', 'masuk', 'pertama', 'bayar', 'parkir', 'motor', 'tiket', 'masuk', 'objek', 'wisata', 'org'

Berdasarkan tabel 4.4 di atas menunjukkan hasil tokenisasi dari beberapa teks yang telah diproses. Setiap kalimat dalam kolom "Case Folding" diubah menjadi serangkaian token (kata-kata atau frasa kecil) yang terpisah, yang tercermin dalam kolom "Tokenization". Proses tokenisasi ini memecah teks menjadi bagian-bagian yang lebih kecil, seperti kata-kata, dan menampilkan setiap token dengan tanda petik tunggal untuk mempermudah pembacaan. Setiap baris mewakili teks yang telah diubah menjadi token, yang akan mempermudah analisis lanjutan seperti analisis sentimen atau identifikasi pola dalam teks. Tokenisasi ini menjadikan teks lebih terstruktur dan siap untuk diproses lebih lanjut, yaitu tahapan stopword.

4.2.4. StopWords

Pada tabel 4.5 menujukan yang hasil stopword pada kolom tokenization dan kolom stopword.

Tabel 4. 5. StopWords

Tokenization	Stopword
['pantainya', 'bersih', 'terkelola', 'dengan', 'baik', 'cuma', 'karena', 'musim', 'panas', 'air', 'lautnya', 'agak', 'menyusut', 'masuk', 'pertama', 'bayar', 'parkir', 'motor', 'tiket', 'masuk', 'objek', 'wisata', 'org']	['pantainya', 'bersih', 'terkelola', 'musim', 'panas', 'air', 'lautnya', 'menyusut', 'masuk', 'bayar', 'parkir', 'motor', 'tiket', 'masuk', 'objek', 'wisata']

Berdasarkan Tabel 4.5 di atas menggambarkan proses penghapusan stopwords dari teks, sehingga hanya kata-kata yang lebih relevan dan bermakna yang tersisa. Sebagai contoh, pada baris pertama, kata-kata seperti "dengan" dan "cuma" dihilangkan, sedangkan kata-kata kunci seperti "pantainya", "bersih", dan "parkir" tetap dipertahankan. Proses ini diterapkan pada seluruh teks untuk mengurangi kata-kata yang tidak memberikan nilai tambah bagi analisis.

4.2.5. Stemming

Pada tabel 4.6 di bawah menampilkan hasil dari stemming.

Tabel 4. 6. Stemming

StopWord	Stemming
'pantainya', 'bersih', 'terkelola', 'musim', 'panas', 'air', 'lautnya', 'menyusut', 'masuk', 'bayar', 'parkir', 'motor', 'tiket', 'masuk', 'objek', 'wisata'	pantai bersih kelola musim panas air laut susut masuk bayar parkir motor tiket masuk objek wisata

Berdasarkan Tabel 4.6 , menunjukan hasil dari proses stemming yang diterapkan pada teks. Kolom "**Stopword**" menampilkan teks yang telah dibersihkan dari kata-kata tidak penting, sedangkan kolom "**Stemming**" menunjukkan teks yang telah diubah ke bentuk dasarnya. Sebagai contoh, pada baris pertama, kata-kata seperti "pantainya", "bersih", dan "terkelola" di kolom "**Stopword**" diubah menjadi bentuk dasar seperti "pantai", "bersih", dan "kelola" di kolom "**Stemming**". Proses ini bertujuan untuk menyederhanakan teks sehingga lebih fokus pada kata-kata utama yang

relevan, menjadikannya lebih mudah dianalisis untuk keperluan lebih lanjut.

4.2.6. Labeling

Pada tabel 4.7 ini menunjukkan hasil proses penglabelan sebagai berikut.

Tabel 4. 7. Labeling

Stemming	Status
pantai bersih kelola musim panas air laut susut masuk bayar parkir motor tiket masuk objek wisata	Positif
salah pantai utara jawa baik kunjung tata bersih fasilitas duduk kios makan susun cantik total biaya masuk kawasan parkir motor masuk pantai orang bilang bersih satu kam pantau aktif keliling galau cocok sedia ayun tepi pantai karaoke dangdut voltase ampli yahud jiwa sedih dendang	Positif
orang asli cirebon gak asing initapi pas kesini loh glow up yaserasa berbedaudah nyaman bersih sedia duduk santai main anakdll nyaman kamar mandi kali yahhe klo ken sepi dateng jumat ajah minggu rame tiket gk mahal murmerdi orang	Positif

Berdasarkan Tabel 4.7 di atas menampilkan hasil penglabelan pada data ulasan yang telah melalui proses *Pre-processing* untuk meningkatkan kualitas data sebelum dilakukan analisis menggunakan algoritma Naive Bayes pada sentimen pengunjung objek wisata Pantai Kejawanan. Proses ini menghasilkan label sentimen berupa "*Positif*" dan "*Negatif*". Kemudian hasil tersebut dapat diliat jumlah total sentimen positif dan negatif, seperti yang ditunjukkan pada tabel 4.8 menunjukan hasil penjumlahan sentimen positif dan negatif.

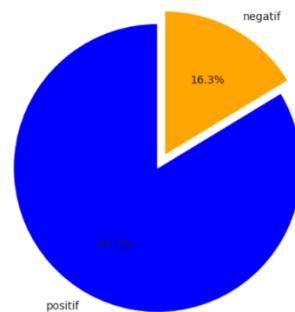
Tabel 4. 8. Perjumlahan Sentimen

Status	Jumlah Sentimen
Positif	835
Negatif	165

Berdasarkan Tabel 4.8 menunjukan hasil jumlah sentimen dari total 998 ulasan pengunjung, terdapat 835 ulasan (83,7%) yang dikategorikan sebagai positif, dan 163 ulasan (16,3%) yang dikategorikan sebagai negatif. Untuk memberikan pemahaman yang lebih mendalam mengenai distribusi ulasan tersebut,

perbandingan antara ulasan positif dan negatif disajikan dalam Gambar 4.1 Visualisasi Penglabelan.

Diagram Hasil Sentimen Positif Negatif



Gambar 4. 1. Visualisasi Penglabelan

4.3. Tranformasi

Pada gambar 4.2 menampilkan hasil tf – idf di bawah ini.

	air	ases	amen	aneka	angin	area	asus	bebas	banting	...	sidah	utara	utk	view	varga	wong	wisata	yr	ye
0	0.26357	0.0	0.00000	0.00000	0.0	0.0	0.0	0.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00000	0.00000
1	0.00000	0.0	0.00000	0.00000	0.0	0.0	0.0	0.246971	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00000	0.00000
2	0.00000	0.0	0.00000	0.00000	0.0	0.0	0.0	0.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00000	0.00000
3	0.00000	0.0	0.24314	0.367983	0.0	0.0	0.0	0.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.204687	0.00000
4	0.20575	0.0	0.00000	0.481983	0.0	0.0	0.0	0.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00000	0.361674

Gambar 4. 2. Hasil TF-IDF

Berdasarkan gambar 4.2 Matriks TF-IDF menggambarkan relevansi kata dalam suatu dokumen dibandingkan dengan keseluruhan dataset. Kata yang tidak muncul diberi nilai 0, sementara kata yang dianggap signifikan, seperti "air" dengan skor 0.263570 pada dokumen pertama, memiliki nilai tertentu. Metode ini secara otomatis menyaring kata-kata umum, sehingga hanya kata-kata yang relevan yang digunakan untuk analisis. Transformasi ini mengubah data teks tidak terstruktur menjadi format numerik, mendukung analisis lebih lanjut seperti klasifikasi sentimen atau pengelompokan dokumen.

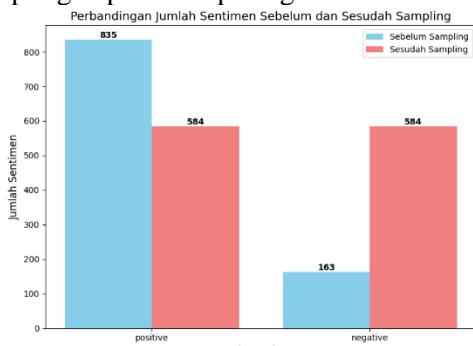
Setelah proses TF-IDF selesai, dataset dibagi menjadi dua bagian, yaitu data *training* sebesar 70% dan data *testing* sebesar 30%. Pembagian ini dilakukan untuk mendukung proses pelatihan dan evaluasi model secara terstruktur, sekaligus meminimalkan risiko overfitting dan underfitting. Pendekatan ini dirancang untuk memastikan bahwa model Naive Bayes yang dikembangkan mampu beroperasi secara optimal dan konsisten ketika diterapkan pada data baru, sehingga meningkatkan tingkat keandalan dan akurasi model.

Hasil pembagian data ini dapat dilihat pada tabel 4.9.

Tabel 4. 9.

Jumlah Data Latih	696
Jumlah Data Uji	299

Setelah data dibagi menjadi data pelatihan dan pengujian, langkah selanjutnya adalah menerapkan teknik random sampling untuk memastikan keseimbangan data dalam dataset. Pendekatan ini bertujuan untuk menghindari dominasi salah satu kelas, meningkatkan keandalan model Naive Bayes, serta mendukung akurasi prediksi pada data pengujian. Berikut hasil visualisasi random sampling dapat dilihat pada gambar 4.3.

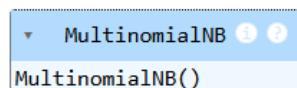


Gambar 4. 3

Gambar 4.3 menggambarkan penerapan teknik SMOTE (Synthetic Minority Over-sampling Technique) untuk menangani ketidakseimbangan kelas dalam dataset. Teknik ini menghasilkan sampel sintetik pada kelas minoritas guna mencapai keseimbangan dengan kelas mayoritas. Dataset yang telah diseimbangkan ini selanjutnya digunakan untuk melatih model Naive Bayes, sehingga meningkatkan kinerja model pada dataset dengan distribusi kelas yang tidak merata.

4.4. Data Mining

Setelah melalui tahapan transformasi yang melalui TF-IDF, pembagian data, dan random sampling, tahap selanjutnya adalah Data Mining, yang bertujuan untuk menemukan pola atau informasi penting dalam data. Pada analisis sentimen pengunjung, tahapan ini menggunakan algoritma Naive Bayes. Hasil proses model Naive Bayes dapat dilihat pada gambar 4.4.



Gambar 4. 4. Naïve Bayes

Gambar 4.4 menggambarkan penggunaan algoritma Multinomial Naive Bayes, yang sering digunakan untuk klasifikasi data diskret, seperti frekuensi kemunculan kata. Algoritma ini didasarkan pada Bayes Theorem dengan asumsi independensi antar fitur, sehingga efektif untuk berbagai tugas, termasuk klasifikasi dokumen, deteksi spam, dan analisis sentimen. Parameter bawaan, seperti alpha yang digunakan untuk proses smoothing, dapat disesuaikan untuk meningkatkan kinerja model sesuai karakteristik data.

4.5. Evaluasi / Interpretation

Hasil evaluasi yang dihasilkan oleh proses tersebut dapat dilihat pada gambar 4.5 berikut.

Classification Report:					
	precision	recall	f1-score	support	
negatif	0.40	0.65	0.50	51	
positif	0.92	0.80	0.86	248	
accuracy			0.78	299	
macro avg	0.66	0.72	0.68	299	
weighted avg	0.83	0.78	0.79	299	

Confusion Matrix:	
[33 18]	
[49 199]	

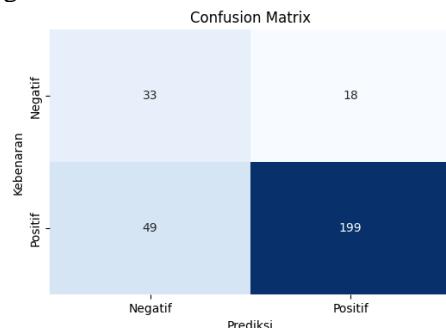
Gambar 4. 5. Hasil Evaluasi Model

Berdasarkan Gambar 4.5 diatas , evaluasi menunjukkan bahwa model memiliki performa yang lebih baik dalam memprediksi kelas positif dibandingkan kelas negatif, sebagaimana tercermin dari nilai precision, recall, dan F1-score yang lebih tinggi pada kelas positif. Akurasi keseluruhan sebesar 78%, Precision 92%, Recall 80%, F1-Score 86% mengindikasikan kinerja yang cukup baik, meskipun akurasi terhadap kelas negatif masih memerlukan perbaikan.

4.6. Confusion Matrix

Confusion Matrix dalam perhitungan ini menggambarkan tabel yang menunjukkan hasil klasifikasi yang benar dan salah oleh model Naive Bayes, serta distribusi klasifikasi untuk setiap kategori. Tabel ini penting untuk memahami bagaimana model Naive Bayes membedakan ulasan positif dan negatif. Dengan menganalisis nilai True Positives (TP), True Negatives (TN), False Positives (FP), dan False Negatives (FN), kita dapat mengevaluasi sejauh mana model berhasil mengklasifikasikan ulasan dengan benar dan mengidentifikasi potensi area yang perlu perbaikan. Confusion Matrix juga menjadi dasar untuk menghitung metrik

evaluasi seperti Akurasi, Precision, Recall, dan F1-Score, yang memberikan gambaran lebih lengkap mengenai kinerja model dalam mengklasifikasikan data.



Gambar 4, 6. Confusion Matrix

Berdasarkan Gambar 4.6 menunjukkan jumlah True Positives (TP) 199 positif yang diklasifikasi benar, True Negatives (TN) 33 negatif yang diklasifikasi benar, False Positives (FP) 18 negatif yang diklasifikasi sebagai positif dan False Negatives (FN) 49 positif yang diklasifikasi sebagai negatif.

4.7. Word Cloud

Word Cloud merupakan representasi visualisasi yang menggambarkan frekuensi kata yang muncul dalam suatu teks, di mana kata yang lebih sering muncul ditampilkan dengan ukuran yang lebih besar sedangkan kata yang jarang muncul ditampilkan dengan ukuran yang lebih kecil. Dalam analisis sentimen pengunjung terhadap objek wisata Pantai Kejawanan, word cloud digunakan untuk mengidentifikasi kata-kata yang dominan terkait dengan sentimen positif dan negatif yang dirasakan pengunjung.



Gambar 4.7. Ulasan Positif

Gambar 4.7 menunjukkan word cloud yang menggambarkan kata-kata dengan frekuensi tinggi dalam ulasan positif mengenai Pantai Kejawanan. Kata-kata seperti "Pantai," "Bagus," "Bersih," dan "Lumayan" ditampilkan

dengan ukuran lebih besar, mencerminkan tingkat kemunculannya yang signifikan. Visualisasi ini memberikan wawasan tentang aspek yang diapresiasi oleh pengunjung, meliputi kondisi pantai, kebersihan, dan kesan umum terhadap kualitas destinasi. Dengan demikian, word cloud ini berfungsi sebagai alat analisis yang efektif untuk memahami persepsi pengunjung terhadap elemen utama pengalaman wisata di Pantai Kejawanan.



Gambar 4. 8. Ulasan Negatif

Gambar 4.8 menunjukkan word cloud yang merepresentasikan ulasan negatif mengenai Pantai Kejawanan. Kata-kata seperti "Kotor," "Mahal," "Parkir," dan "Sampah" ditampilkan dengan ukuran lebih besar, menunjukkan tingginya frekuensi kemunculannya dalam ulasan. Hal ini mencerminkan keluhan pengunjung terkait kurangnya kebersihan, tarif yang dianggap tidak sebanding dengan fasilitas, permasalahan parkir, serta keberadaan sampah di area wisata. Visualisasi ini memberikan gambaran tentang isu-isu utama yang dapat menjadi prioritas untuk meningkatkan kualitas pengalaman pengunjung.

5. KESIMPULAN

Berdasarkan hasil penelitian mengenai analisis sentimen pengunjung terhadap objek wisata pantai kejawanan menggunakan algoritma Naive bayes, maka dapat disimpulkan sebagai berikut :

- a. Penelitian ini berhasil mengidentifikasi dan mengklasifikasikan sentimen pengunjung terhadap Pantai Kejawanan menggunakan algoritma Naive Bayes. Proses analisis dimulai dengan mengumpulkan data ulasan pengunjung yang di peroleh dari Google Maps , kemudian dilakukan pra- processing data untuk menghilangkan kata-kata yang tidak relevan. Selanjutnya, algoritma Naive Bayes diterapkan untuk mengklasifikasikan

- sentimen menjadi 2 (dua) kategori: Positif, dan Negatif. Hasil analisis menunjukkan bahwa algoritma Naive Bayes dapat dengan baik mengklasifikasikan sentimen dengan akurasi yang cukup tinggi.
- b. Algoritma Naive Bayes menunjukkan tingkat akurasi yang cukup tinggi sebesar 78% dalam menganalisis sentimen pengunjung terhadap Pantai Kejawanan. Berdasarkan pengujian yang dilakukan, model ini memiliki akurasi yang cukup baik dalam memprediksi sentimen, meskipun ada beberapa data yang tidak terkласifikasi dengan sempurna. Hal ini menunjukkan bahwa Naive Bayes efektif, namun masih terdapat ruang untuk peningkatan akurasi, seperti melalui penggunaan metode grid search dan cross validation untuk meningkatkan performa model.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] D. S. Utami and A. Erfina, “Analisis Sentimen Objek Wisata Bali Di Google Maps Menggunakan Algoritma Naive Bayes,” *J. Sains Komput. Inform.*, vol. 6, no. 1, pp. 418–427, 2022.
- [2] N. L. P. Merawati, A. Z. Amrullah, and Ismarmiyati, “Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 123–131, 2021, doi: 10.29207/resti.v5i1.2587.
- [3] Y. A. Singgalen, “Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naïve Bayes Classifier,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, p. 1343–1352, 2022, doi: 10.47065/bits.v4i3.2486.
- [4] G. K. Pati and E. Umar, “Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Danau Weekuri Menggunakan Metode Naive Bayes Classifier Dan K- Nearest Neighbor,” *J. Media Inform. Budidarma*, vol. 6, no. 4, pp. 2309–2315, 2022, doi: 10.30865/mib.v6i4.4635.
- [5] M. F. Y. Herjanto and Carudin, “Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 1204–1210, 2024, doi: 10.23960/jitet.v12i2.4192.
- [6] Y. A. Singgalen, “Sentiment Analysis on Customer Perception towards Products and Services of Restaurant in Labuan Bajo,” *J. Inf. Syst. Informatics*, vol. 4, no. 3, pp. 511–523, 2022, doi: 10.51519/journalisi.v4i4.294.
- [7] Y. A. Singgalen, “Analisis Sentimen Pengunjung Pulau Komodo dan Pulau Rinca di Website Tripadvisor Berbasis CRISP-DM,” *J. Inf. Syst. Res.*, vol. 4, no. 2, pp. 614–625, 2023, doi: 10.47065/josh.v4i2.2999.
- [8] R. Sari, “Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi menggunakan Algoritma K-Nearest Neighbor (K-NN),” *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, pp. 10–17, 2020, doi: 10.31294/evolusi.v8i1.7371.
- [9] M. Azhari, Z. Situmorang, and R. Rosnelly, “Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes,” *J. Media Inform. Budidarma*, vol. 5, no. 2, pp. 640–651, 2021, doi: 10.30865/mib.v5i2.2937.
- [10] Y. A. Singgalen, “Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT),” *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1551–1562, 2023, doi: 10.30865/mib.v7i3.6461.
- [11] R. Ardianto, T. Rivanie, Y. Alkhali, F. Nugraha, and W. Gata, “SENTIMENT ANALYSIS ON E-SPORTS FOR EDUCATION CURRICULUM USING NAIVE BAYES AND SUPPORT VECTOR MACHINE,” *J. Ilmu Komput. dan Inf. (Journal Comput. Sci. Information)*, vol. 2, no. 13, pp. 109–122, 2020, doi: 10.1109/ICECOS47637.2019.8984530.
- [12] F. Taufiqurrahman, S. Al Faraby, and M. D. Purbolaksono, “Klasifikasi Teks Multi Label pada Hadis Terjemahan Bahasa Indonesia Menggunakan Chi Square dan SVM,” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 10650–10659, 2021.